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

Graham E. Budd[1] and Richard P. Mann[2]

[1] Dept of Earth Sciences, Palaeobiology, Uppsala University, Villavägen 16, Uppsala, Sweden, SE 752 36. graham.budd@pal.uu.se
[2]Department of Statistics, School of Mathematics, University of Leeds, Leeds LS2 9JT, United Kingdom. r.p.mann@leeds.ac.uk

Abstract

Phylogenies may be modelled using "birth-death" models for speciation and extinction, but even when a homogeneous rate of diversification is used, survivorship biases can generate remarkable rate heterogeneities through time. One such bias has been termed the "push of the past", by which the length of time a clade has survived is conditioned on the rate of diversification that happened to pertain at its origin. This creates the illusion of a secular rate slow-down through time that is, rather, a reversion to the mean. Here we model the controls on the push of the past, and the effect it has on clade origination times, and show that it largely depends on underlying extinction rates. Crown group origins tend to become later and more dispersed as the size of the push of the past increases. An extra effect increasing early rates in lineages is also seen in large clades. The push of the past is an important but relatively neglected bias that affects many aspects of diversification patterns, such as diversification spikes after mass extinctions and at the origins of clades; it also influence rates of fossilisation, changes in rates of phenotypic evolution and even molecular clocks.

Keywords: Crown groups, diversification rates, mass extinctions, the push of the past, Cambrian Explosion, molecular clocks.

1 Introduction

The patterns of diversity through time have been of continuous interest ever since they were broadly recognized in the 19th century (e.g. [1]). In particular, both major radiations (such as the origin of animals [2]) or angiosperms [3]) and the great mass extinctions (e.g. the end-Permian [4] or end-Cretaceous [5] [6]) have attracted much attention, with an emphasis on trying to understand the causal mechanisms behind these very striking patterns. However, in the midst of this search, the effect of survival biases on creating the patterns under consideration has not been considered as fully as it could have been. The last decades have also seen a great deal of interest and work on mathematical approaches to diversification and extinction (e.g. [7] [8] [9] [10] [11]), including some that touch on the topics considered by this paper (for example, see especially [12] [13] [14] and [15]), but there is relatively *little* literature on the dynamics of clade origins from the perspective of survival biases and their effect on the fossil record. In this paper, then, we wish to explore the basis for such bias and then consider how it is exported to various important aspects of the observed large scale patterns of evolution, with particular focus on the sort of data that can be extracted from the fossil record. In the following analyses, we calculate over an interval of 0.1myrs and plot graphs with an interval of 2myrs.

2 The "push of the past"

Nee and colleagues [16] [17]) summarised the general mathematics of birth-death models as applied to phylogenetic diversification (see also especially [14]). In such models, each lineage has a certain chance of either disappearing ("death") or splitting into two ("birth") per unit time. Most models that consider diversity in this way have used a constant birth and death rate which have revealed much of interest about phylogenetic processes [[18] [19]]. More sophisticated models have incorporated varying rates of birth and death or allowed speciation to be spread over time [20] or used a diversity-dependent extinction rate in an attempt to produce more realistic models that can be tested against the fossil record and reconstructed molecular phylogenies [21] [11].

The existence of rapid diversification events in the fossil record has been documented or inferred on many occasions and in many clades (e.g. [22] [5] [23] [24] [25] [26] [27]). These patterns have been cited as evidence for qualitatively differing regimes of diversification during special intervals of earth history (e.g. the Cambrian explosion [28] or radiations after mass extinctions [29]). This in turn has lead to proposals of various models to explain these patterns (either intrinsic to the evolutionary process (e.g. [30]) or extrinsic, such as changes in the environment (e.g. [31]). Moreover, associated macroevolutionary phenomena such as the early appearance of the body plans of the extant phyla or other groups and their subsequent relative stasis afterwards have also attracted much interest and attempts at explanation (e.g. [32] [33] [34]). However, even though constant-rate birth-death models look fairly simple on the surface, they can generate surprising statistical patterns caused by often overlooked biases.

[16] identified one of these biases that they nicknamed "the push of the past", a nod towards another bias called the "pull of the recent" which is based on the very high "preservation rate" of the Recent compared to the fossil record [35]. Clades with a recent origin have had less time to go extinct and are thus more diverse than one would expect.

Controversy over the size and detection of the pull of the present in molecular phylogenies has led to various modifications of the simple birth-death model (see e.g. discussion in [11]; but see also [36] [37] [38] [39]). To avoid confusion between the two, [20] renamed the phylogenetic effect "the pull of the present". Here we use the acronyms PoTPa and PoTPr to refer to these two effects.). In this paper we are primarily concerned with what happens *before* the present is reached, but would note that the problems of comparing modern day lineages with fossil species may be part of these problems (c.f. [40] [14]).

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 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. ([41]) put it, such clades "get off to a flying start". Once clades become established, they are less vulnerable to random changes in the actual diversification rate, and this value therefore tends to revert to the mean through time. Long-lived clades thus tend to show high effective rates of diversification at their origin, which then decrease to their long-term average as the Recent is approached (c.f. [42]). A similar effect should apply to now *extinct* clades 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, i.e. 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.

Conversely, the "pull of the present" is 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 must increase faster than the overall 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).

Although the POTPa is known about (e.g. [43] [14]), emerges naturally from conditioning clade survival to the present in diversification simulations and has even been examined in some detail in a few cases [44] [42], it has had surprisingly little penetration into especially the palaeontological literature. In this paper we therefore wish to show how important an effect it is, and discuss its implications for general discussions about the reasons behind typical patterns of diversification seen in the fossil record.

A typical sort of question one might address to the fossil record, for example, is when modern (i.e. crown) groups are likely to emerge [45]. Raup [46] showed many years ago that under a birth-death model, the last common ancestor of any two living organisms is likely to lie deep in the past. In the context of modern systematics, we would like to ask the related question; given the emergence of a total group, what conditions govern the emergence of the extant crown-group from it? A previous attempt at tackling this problem by simulations showed that as the rate of diversification increases, crown-groups tend to emerge earlier and earlier, so that stem-groups become shorter and shorter [45]. Given the definitional requirement for crown groups to survive until the present day, this

pattern becomes entangled with the POTPa. Here we build on this previous effort [45], and capture these patterns analytically, as follows.

3 Mathematical analysis

In this section we extend on the approach of Nee *et al.* by explicitly conditioning the birth-death model on the number of extant species in the crown group. We also extend the work of Raup, by developing a fully probabilistic calculation of the distribution of crown group origins, that allows for both stochastic variation in the number of speciations and extinctions in any unit of time, and incorporating the selection-bias of the 'pull of the present' as a result of examining surviving clades. Raup as it stands is a fully deterministic model that does not allow any effective rate changes through time and thus cannot allow for survivorship biases.

3.1 Estimating the number of extant and surviving clades through time

As noted by previous studies [16] [47], 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 origin at time zero, obeys the following zero-truncated geometric probability distribution

$$\begin{aligned} P(n_t) &= \mathrm{G}(n_t - 1; 1 - a_t) \\ &\equiv (1 - a_t) a_t^{n_t - 1} \end{aligned} \tag{1}$$

with

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

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,

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

For the limiting case where $\lambda - \mu \to 0$, see [47]. Nee *et. al* proceeded by conditioning the distribution of n_t on the tree surviving until some future time, T. Here instead we move towards the theory of Raup by conditioning on there being n_T extant species at time T. By Bayes' rule:

$$P(n_t \mid n_T) = \frac{P(n_T \mid n_t) P(n_t)}{P(n_T)} \tag{4}$$

Two terms in this equation are given immediately from equation 1. We can evaluate the remaining term, $P(n_T \mid n_t)$, by recognising 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 $P(n_T \mid m_t)$

follows a truncated negative binomial distribution (with n_T taking a minimum value of m_t):

$$\begin{aligned} P(n_T \mid m_t) &= \mathrm{NB}(N - m_t; m_t; 1 - a_{T-t}) \\ &\equiv \binom{N-1}{N-m_t}(1 - a_{T-t})^{N-m_t} a_{T-t}^{m_t} \end{aligned} \tag{5}$$

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:

$$\begin{aligned} p(m_t|n_t) &= \mathrm{B}(m_t; n_t, s_{T-t}) \\ &\equiv \binom{n_t}{m_t} s_{T-t}^{m_t}(1 - s_{T-t})^{n_t - m_t} \end{aligned} \tag{6}$$

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:

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

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 which will have at least one descendant at time T. Here we integrate over possible values of n_t:

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

By looking at a small 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 effective rate of diversification for surviving clades:

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

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:

$$R_0 = (2 - s_T)\lambda. \tag{10}$$

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,

$$R_0 \simeq 2\lambda, \tag{11}$$

(we note that similar results concerning the interior branch lengths of reconstructed trees have been derived by [13]). Since every new surviving lineage begins a new tree with one original species, this logic implies that along any surviving lineage, the rate of speciation is always similarly determined by the probability of survival to the present and is always greater than λ. A notable feature is that in the Recent, as the survival probability s_T tends to one, the rate of speciation declines back towards λ. As the survival rate remains low for long periods of a radiation, it follows that the constant renewal of the surviving stem lineages follows a quasi-fractal pattern of repetition.

At what rate do these speciations give rise to additional surviving lineages? Consider the probability that, after a speciation from an initial state of $n_t = 1$ to $n_{t+\Delta t} = 2$, both lines go on to survive, conditioned on knowing that at least one does. Since survival of both implies survival of at least one, we have:

$$\begin{aligned} P(\text{both survive} \mid \text{at least one survives}) &= \frac{P(\text{both survive})}{P(\text{at least one survives})} \\ &= \frac{s_{T-t}}{2 - s_{T-t}} \end{aligned} \tag{12}$$

This implies that along each stem group, i.e., between the creation of each new lineage, there will on average be $[(2 - s_{T-t})/s_{T-t}] - 1$ plesions (i.e. extinct lineages: See Fig. 3 below). Since new speciations happen at a rate $(2 - s_{T-t})\lambda$, the total time taken along the stem will be, on average, $1/(\lambda s_{T-t})$. Equivalently the rate at which new lineages are created is λs_{T-t}, which for large values of $T - t$ is approximately $\lambda - \mu$, i.e. the baseline rate of diversification. In the Recent, as the survival probability tends to one, the rate of new lineage creation will increase to λ, the baseline speciation rate, since all new species will be expected to survive to the present.

An example plot displaying the various parameters that govern this analysis is given in Fig. 1A. (c.f. [16]). 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 effective 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. 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* 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. Fig 1B shows the implied large spike in the initial effective diversification rate, owing to the POTPa, with the initial effective diversification rate (~ 1) being 100 times the underlying average. Note also that this effect generates a (non-causal) correlation between diversity and diversification rate (Fig. 1C): as diversity increases, (average) rate of diversification decreases.

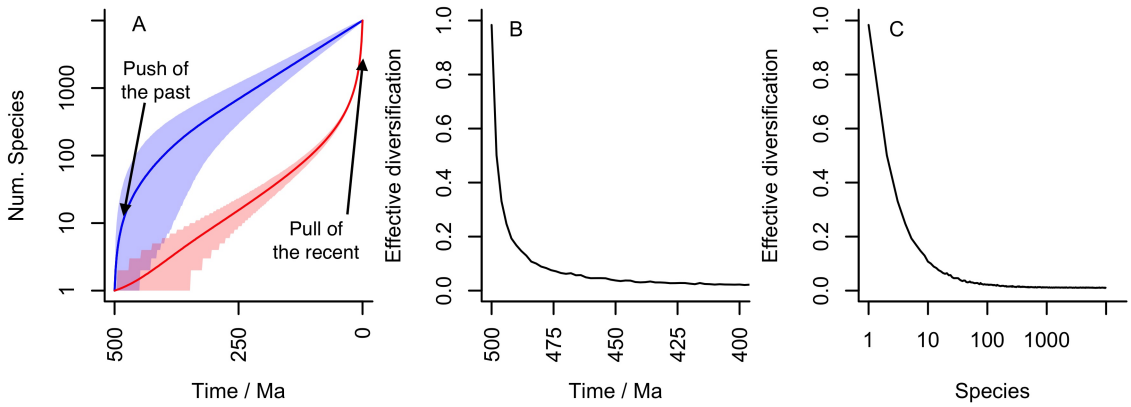

Figure 1: An example diversification with 10,000 living species, an extinction rate of 0.5 per species per million years and a diversification time of 500 Myrs. The implied speciation rate is 0.5107 per species per million years and thus the underlying diversification rate is 0.0107 per species per million 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** Effective diversification rate at beginning of diversification (note scale of 100 Myrs). C Implied diversification rate correlation with diversity generated by this distribution

.

3.2 Crown group origins

If the number of species at time T is known (n_T), and if the number of clades that will survive at time t is also known (m_t), we can calculate the probability, $W(t)$, that two randomly chosen species in the Recent will have a common ancestor at time t. This is the definition of a "randomly selected" crown group used by Raup [46]. We first need to pick the first species at random (with probability one), then we need to pick a second species that has the same ancestor at time t - this must first be one of the $n_T - m_t$ remaining species that do not inevitably have to join up with the other $m_t - 1$ ancestor species, and then a randomly selected species from this remaining set will have a $1/m_t$ probability of sharing an ancestor with the first selected, thus:

$$W(t|m_t) = \frac{n_T - m_t}{n_T m_t}. \tag{13}$$

Accounting for the uncertainty in our knowledge of m_t, our estimate of $W(t)$ requires a posterior-weighted summation over the possible values of m_t:

$$W(t) = \sum_{m_t=1}^{N} \frac{N - m_t}{N m_t} P(m_t \mid n_T), \tag{14}$$

where the posterior distribution on m_t is calculated as above. $W(t)$ 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, $w(t)$, is given by differentiation of $W(t)$:

$$w(t) = -\frac{dW}{dt}. \tag{15}$$

The depth of origin of random pairs of taxa is what Raup was attempting to model, but our model smears out the timing of origin of crown groups as defined in this way because of the effect of the POTPr of allowing a longer period of lower early lineage diversification rates, which Raup neglected (i.e. Raup's model is more akin to the Yule process). Nevertheless, it remains true that under "normal" conditions, 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 & Jackson [45] simulated the origin of the *first* crown groups in clades conditioned on survival (c.f. [42]). It is also possible to capture this result analytically. 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 by calculated by considering the rate of change in the probability that $m_t = 1$:

$$u(t) = -\frac{dP(m_t = 1 \mid n_T)}{dt}. \tag{16}$$

As summarised in [16] and Fig. 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 the expected length of time it takes for the first crown-group to emerge is given by:

$$\log(m_t) \simeq (\lambda - \mu)t. \tag{17}$$

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

$$t_{cg} \simeq T - \frac{\log 2}{\lambda - \mu}, \tag{18}$$

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:

$$\lambda - \mu \simeq \frac{\log 2}{T - t_{cg}}. \tag{19}$$

This relation broadly holds for positive diversification rates until approaching the neutral model ($\lambda = \mu$), where it breaks down, as when observing extant clades, the crown groups must have emerged before the present time. Budd and Jackson [45] found by simulation that neutral models tended to have crown-group emergence at about 0.5 of total group time, indicating that the broadening of time of emergence of crown groups seen in Fig. 3 eventually leads to a very uniform distribution.

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 Fig. 2B.

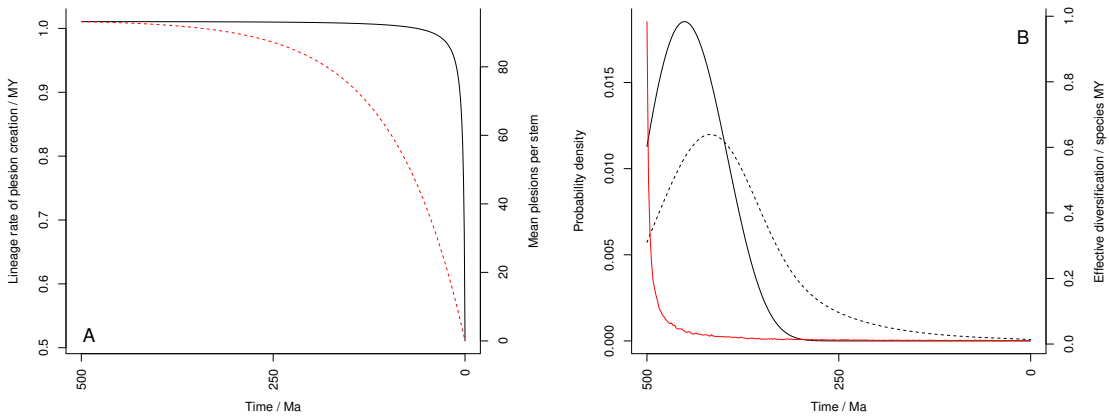

Figure 2: **A** Illustration of the rate of plesion creation along the surviving lineages (black solid line) and the mean number of plesions created along each stem group (red dashed line) through time. As rate of plesion creation is almost flat for most of the time, it follows that the decline in number of plesions per stem group depends on the stem groups decreasing in size temporally. **B** Effective 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 500myrs 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.

4 The push of the past and the fossil record

4.1 Overview

Within a particular total group, then, stem groups are characterised by high effective diversification rates and low diversity; and crown groups by low diversification rates and increasing diversity (c.f. 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. Clades in this view are analogous to K and r strategists. A large clade with many lineages has a inherent high rate of survival (i.e. like a K-strategist) as extinction requires loss of all of them. Conversely, small clades (like r-strategists) have inherently low rates of survival as they are vulnerable to small fluctuations. Hence, the survival of at least one of them is (in retrospect) guaranteed by having very many of them. The emergence of the crown group thus corresponds to a switch from the r to the K modes of survival.

It is important to note that the high rates of diversification in stem groups are not *general* features, as we are applying a homogeneous model of diversification. Rather, unusually high diversification rates are concentrated in the stem *lineage* that leads to the crown group(s) (c.f. [13]. Stem groups should thus generate a high number of plesions (i.e. extinct sister groups to the crown group [48] [49]) which themselves will diversify and go extinct at 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).

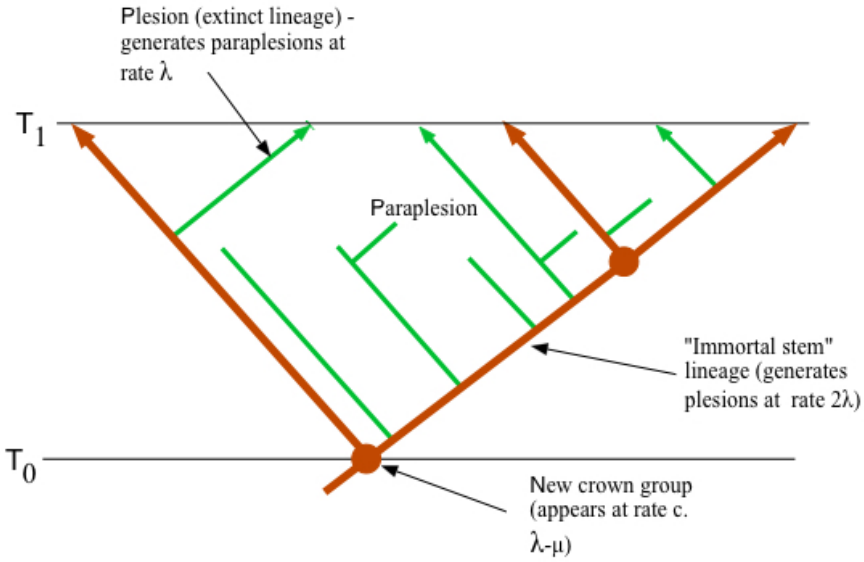

Figure 3: A small section of a tree at a time distant from the present. Red branches represent lineages that survived until the Present, and where they diverge represents the birth of a new crown group. Green branches represent plesions that do not survive until the present. As per the results presented herein and in [13], 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 $\lambda - \mu$.

This analysis gives us a remarkable perspective on the fossil record (Fig. 3), which is after all made up entirely of plesions (Budd 2003). 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 (as taxa in cladograms are only ever terminals) are stem lineages that speciate at twice the normal rate. It is only briefly, at the beginnings of radiations and 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.

4.2 Diversification scenarios

Armed with the mathematical analysis and example above, we are now in a position to analyse 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.

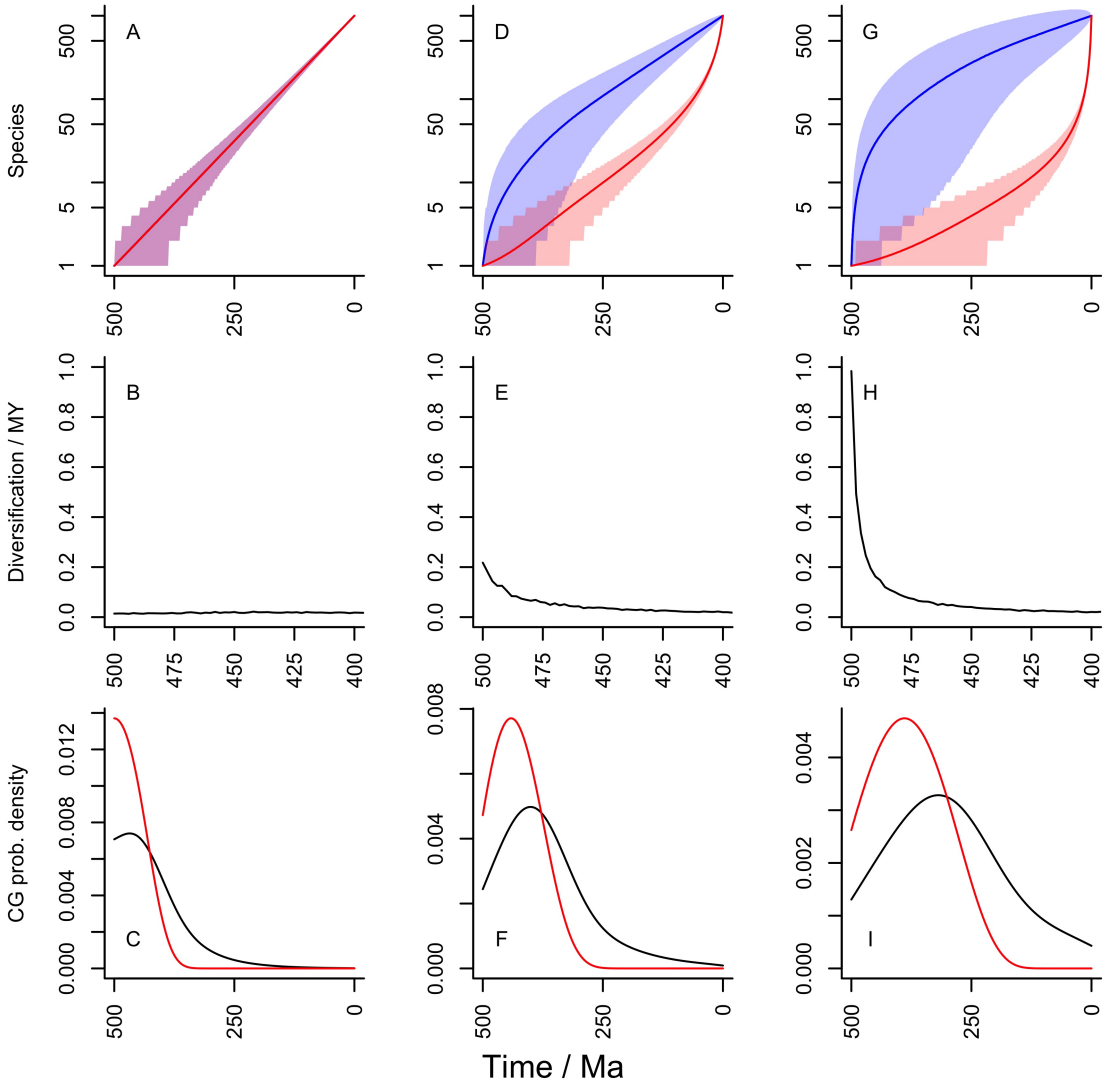

Figure 4: Patterns of diversification and stem- and crown-group formation for different (constant) diversification parameters. For each column T= 500Myrs and $n_T = 1000$. All rates given per species per million years. Shading gives 95% confidence areas. Row one gives plots of diversity and diversity that gives rise to extant species through time; row two gives effective average diversification rates through time over the first 100Myrs (ie. 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 $\lambda = 0.505$. Note different time scale on second row.

4.3 The Yule process

The Yule process [50] governs diversification processes with the assumption that no extinction takes place, i.e. that $\mu = 0$. Of course this is not realistic, especially 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 Fig. 1 are coincident (Fig. 4A) irrespective of the error on each. There is neither a pull of the recent 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. In the example, the rate of diversification is completely constant along the mean, since the diversification rate has been selected to generate n_T (ie 1000 species in this case). Nonetheless, as the confidence region shows, especially early fluctuations in this process are possible, which we consider further later. A deterministic Yule model is effectively the model used by Raup [46]. Another obvious feature of the no-extinction model is that total and crown groups are effectively coincident for any particular clade, as stem-groups grow by extinction [51]. A short lag at the beginning is possible though, before the first speciation event takes place.

4.4 Models with net diversification and extinction

Fig. 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 longevities (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 ([52]).

Fig. 4G-I is very close to the pattern that would be seen with the neutral model, which was of great historical importance in modelling patterns of diversity in the fossil record: it was the basis of the famous MBL model [53] [54]. In the neutral model, λ and μ have the same value, but can vary together if the model is conditioned on n_T and T. In other words, it is possible to have neutral models with high turnover (λ and μ are both high) or low turnover (λ and μ are both low).

In non-Yule processes such as these, because there is now extinction to be considered, and thus a POTPa, it follows that there must be "unlucky" clades that had a neutral or even negative effective diversification rate close to their origin, and these clades will not on balance have survived. For example, if one looks over a 250 million year interval with an average species survivorship of 2 million years (ie $\mu = 0.5$), then only about 1 percent of all clades actually survive; when μ is 0.1, about 4 percent do (see equation 3). Thus, the neutral model tends to be *highly* selective towards clades that happened to have high diversification rates early in their history [14]. Indeed, the POTPa generally has the paradoxical effect of making high extinction rates *increase* effective rates of diversification and numbers of living species - in clades that managed to survive.

4.5 The "Copernican" nature of 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 4H) 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. [14])- much lower than seems to be typical for at least Phanerozoic taxa, which typically have a life time of a few million years ([55]).

We have summarised the various effects of the POTPa in Fig. 5, by showing the effective rate of diversification plotted against extinction rate for different times in a diversification taking 500Myrs. Note that almost all of the POTPa is confined to about the first 25Myrs - after this early interval, all the curves flatten out as diversification rates revert to their long-term average (note the log scale on the y-axis).

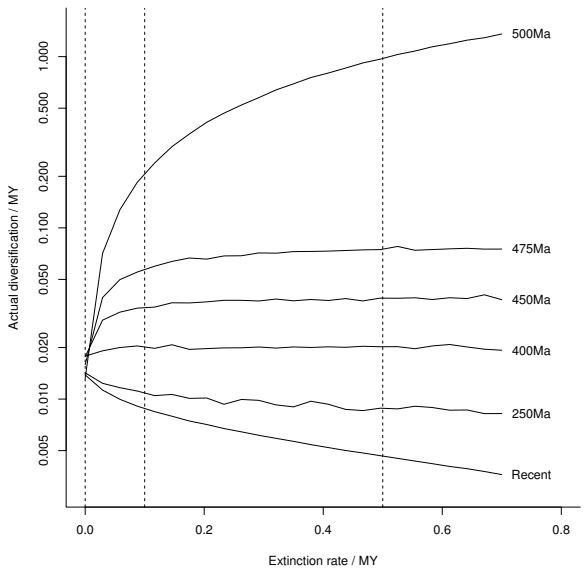

Figure 5: Effective diversification rate versus background extinction rate (μ) at different times during 500Myrs radiations to the present. Different radiations can be summarised by taking vertical sections at different background μ. The three vertical lines shown indicate the three radiations shown in Fig. 4.

Our modelling of different scenarios shows that different initial conditions generate effects that pull in two different directions. Clades or scenarios that have low extinction rates very successfully generate survivors, but not very many of them. Conversely, high-extinction rate clades generate large numbers of survivors, but they do not survive very often. This differential survivorship can be seen in a plot of expected initial diversification rate as a function of survival time for different rates of extinction. A high POTPa is required in order for clades with high rates of extinction to survive the first few million years (Fig. 6). Conversely, once a clade has escaped this danger zone and diversified sufficiently, it becomes very difficult to eliminate it (c.f. [56]), which is why the expected survival times flatten in this figure.

Because of the nature of the homogeneous model we are using, we wish to stress its 'Copernican' aspect, i.e. that diversification is on average the same at all times. Each stem lineage will be characterised 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.

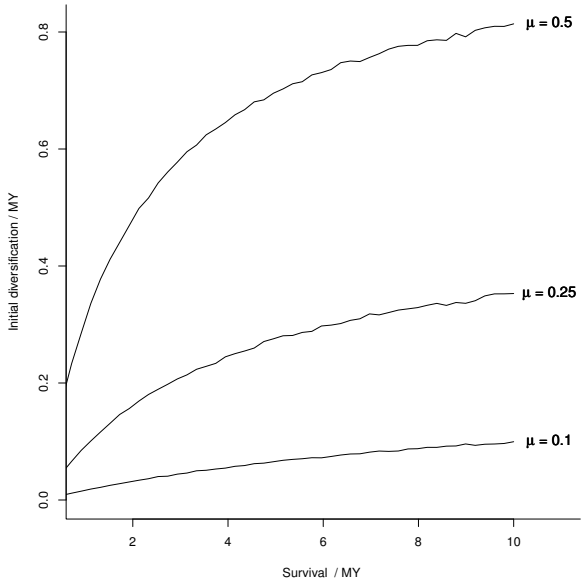

Figure 6: Expected effective initial diversification rate as a function of clade survival time for different values of underlying extinction rate, μ in a neutral model. As μ increases, a bigger and bigger POTPa is required to ensure the clade survives the first few million years.

5 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 (c.f. [57]). Nevertheless, the handful of mass extinctions through time have had a large impact on diversification patterns. The most important are perhaps the end-Permian (with c. 96% of all species going extinct [4] and the end-Cretaceous (c. 75% loss [58]). Such events could be considered, as simply "resetting the clock" - i.e., 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 [55] - see below, which divides Phanerozoic time more or less into two, with large POTPa, but largely uncommented, 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 500Ma, 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 250Ma down to only one species and the subsequent POTPa and re-radiation 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 re-radiation would be expected to generate only 3 living species.

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 effective diversification rate post-extinction (Fig. 6) for different rates of background extinction for a radiation that took 250Myrs to generate 1000 species. As can be seen, extinctions can indeed generate large POTPa, but the number of remaining species for the clade needs to be reduced to a few percent of their original numbers. Thus, large POTPa effects after a mass extinctions are likely to be contingent on large extinctions preceding them in the clade in question.

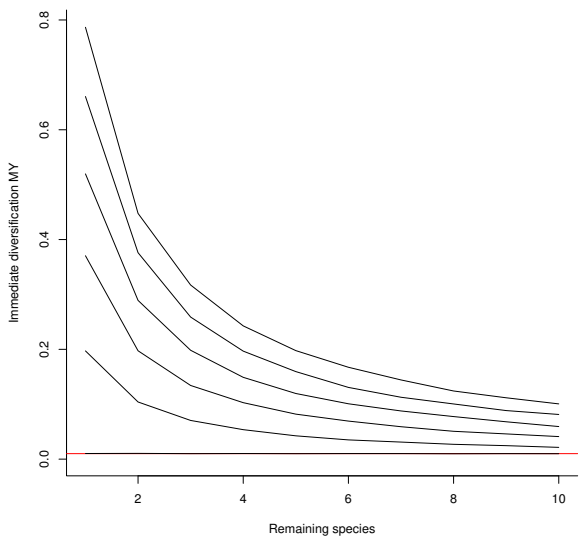

Figure 7: The impact of number of remaining species on a post-extinction POTPa on a clade that had diversified for 250Myrs to generate 1000 species, assuming clade survival to the present, with baseline diversification $\lambda - \mu = 0.01$. The curves from bottom to top represent background extinction rates of $\mu = 0$ (Yule process), 0.1, 0.2, 0.3, 0.4 and 0.5 per species per million years.

6 Timing of major clade origins

It should be stressed as a result of the above that although it is possible in *principle* to model any sort of crown- or stem-group length, in practice the most extreme examples

are rare. For the three cases presented in Fig. 4, the probability of survival to the Recent is 1 (Yule process); 0.082 (low μ net diversification) and 0.010 (high μ net diversification). Similarly, in contrast to the idea that major radiations can have long and low-diversity cryptic "fuses" (e.g. [59]), our modelling of crown-group origins suggest they emerge when they do because this is typically soon after the total group did. To take the Cambrian explosion as an example, the apparent high rates of diversification of at least arthropods in the initial phases [60] imply that the total group of animals only emerged a short time before, and similar arguments can be applied to the animals as a whole. One way of escaping this view would be to claim that the bilaterian animals, allegedly like the Ediacarans that preceded them, underwent a mass extinction at the end of the Ediacaran period before rebounding. The strength of the apparent POTPa in the Cambrian [61] [60] however suggests this would have to be of enormous effect, potentially indistinguishable from a true origin close to this time. At the very least, the idea of the bilaterians waiting in the shadows of the Ediacarans before emerging after the latter went extinct seems unlikely.

7 Diversification spikes in the fossil record: macroevolutionary consequences

An initial burst of diversification at the base of major (ie long lived) clades appears to be an inevitable effect of the selection bias that is required in order to allow clades to survive until the present day (or any other significant period of time). A corollary of this little-appreciated fact is that such diversity spikes do not require a special macroevolutionary explanation, for example, that the bases of clades *per se* have special ecological or developmental features (for example) that drive them. This observation is however, not to make the claim that diversification takes place without a cause. If one could investigate the origins of a clade, one might be able to understand the reasons why those *particular* members of a clade happened to diversify.

However, the converse pattern of the POTPa is that there should be members of clades that did *not* experience unusually high initial levels of diversification, and it was precisely these members that did not go on to found living clades. Given that (from our calculations) the POTPa is not instantaneous but is rather spread over some millions of years, it may be possible to examine differential diversification rates amongst those members that ultimately survived and those that did not [12]. Thus, the view of Gould that there is no difference between such taxa in, say, the Cambrian [62] is open to empirical examination. If apparent diversification bursts are simply down to the POTPa, then one would expect that if one could observe all diversification and extinction going on during that time, one would not see anything particularly "special" about the rates overall. Conversely, whilst we would predict that the ultimate non-survivors would indeed show lower diversification rates, it might also be the case that they too show increased diversification rates compared to the implied long-term average, implying a true macroevolutionary effect. Thus, a clear view of the effects of the POTPa is necessary to distinguish true macroevolutionary effects (such as diversity dependent diversification [42]) from survivorship bias.

8 Mass extinctions revisited: the POTPa and preservation potential

As noted above, another clear effect of the POTPa is to generate diversification spikes after mass extinctions (see e.g. [61] [63], [55]). Like original radiations, these two are not (necessarily) generated by special conditions that generally pertain after mass extinctions (such as them emptying of niches), but a necessary condition for initially low-diversity surviving clades to survive until the present. Nevertheless, the POPTa definitionally only applies to clades that survive, to the present or at least some time. Conversely, diversification spikes after mass extinctions can be recognised on the basis of raw counts in the fossil record [64] [61]. Some of this pattern may be based on interest bias towards extant clades (i.e. particular interest in the evolutionary history of living groups such as the bivalves might conceivably bias collection effort towards such clades).

However, 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 effective extinction rate ([65]; [66]). 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. Fig. 4 shows there is a strong relationship between survivorship on a million or sub-million 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 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 [67]). 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 [55]), 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.

Despite our calculations showing that living clades *should* have started off with a POTPa, not all clades appear to exhibit such a pattern when examined in the fossil record. For example, Hopkins and Smith ([68]) studied patterns of diversification in post-Palaeozoic echinoids and concluded that they showed a *series* of diversification peaks throughout their history, which were not concentrated at the origin of the clade; the patterns observed, moreover, differed considerably at different scales. Nevertheless, such a pattern, which shows at least large peaks at the opening of the Jurassic and Caenozoic, may be conditioned on the large mass extinctions just before each. It is worth noting, further, that the POTPa model has as a background hypothesis that extinction and origination rates are not *actually* significantly different at these times. Given that the POTPa affects clades with reasonable survival, it may thus be that the entire fossil record of a well-preservable clade like the echinoids might be capturing some of this true rate of diversification.

9 Lineage effects: an analogy to the POTPa in large clades

Our exploration of the POTPa shows that, when conditioned on survival, it remains nearly constant at along the surviving lineages at close to 2λ until close to the Recent. Hence, it largely cannot account for long-term declines in phenotypic, lineage diversification or molecular rates (see below). However, the model can also be conditioned on the number of survivors, n_T, and thus it is possible to examine the outliers in homogeneous birth-death models too [14] , which tend to have long tails (being exponential). For example, we recalculated Fig. 1, and conditioned it to generate 100,000 instead of 10,000 species (Fig. 8). Under such rare conditions, more species 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.

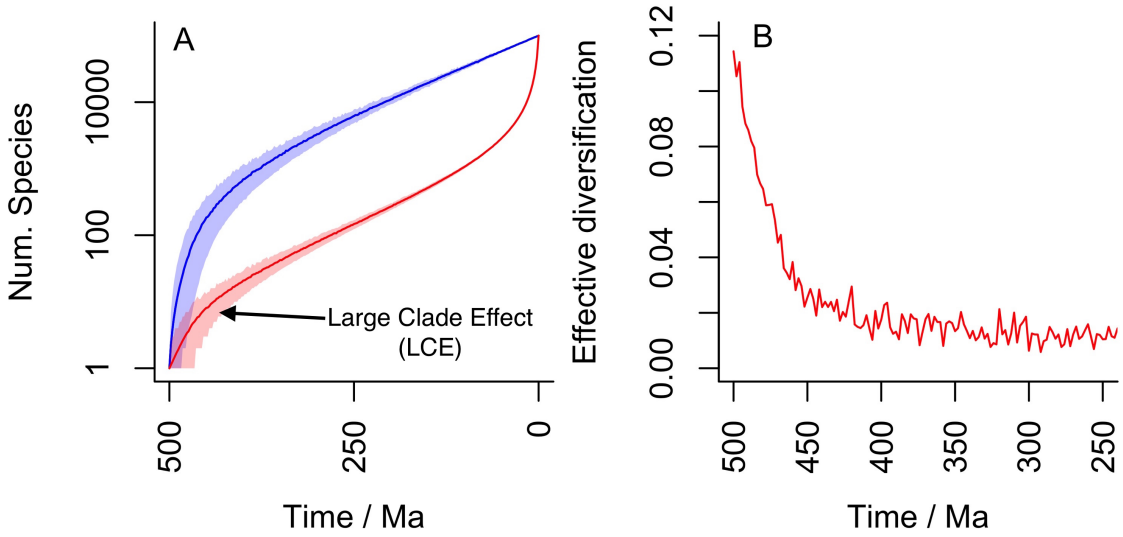

Figure 8: **A** Diversification when an exceptionally large clade is generated with the parameters of Fig. 1 (here, 10x larger than expected). An early lineage effect is introduced as this is where fluctuation in actual rates is most likely. **B** Calculated decline of lineage rate (thus in red) through time. Note that it lasts considerably longer than the POTPa although is of smaller effect (here initially c. 10x the background rate of $\lambda - \mu$). The variations are owing to Monte Carlo numerical integration of equation 8, because of low overlap between the prior distribution and likelihood function for m_t.

Under such circumstances, a *lineage* effect is produced, which could be called the large clade effect (LCE). Although it is smaller than the classical POTPa, it has the effect of speeding up the appearance of new living lineages near the beginning, and thus makes crown groups emerge (even) earlier. Like the POTPa, it has a quasi-fractal organisation. The lineage curve takes on a characteristic inverted "S" shape that is often seen in plots of molecular evolution (e.g. [11] [69]). 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 [14].

Although the effect is small unless the clade is large, it should be noted that a randomly chosen species is likely to be in a larger-than-average clade, and thus there is a consistent, if small, bias towards this effect appearing. In our example, an effect of about $3(\lambda - \mu)$ is present with a clade size of 30,000.

9.1 Rates of phenotypic change

The rate of phenotypic change through time is another pattern that has seen a great deal of interest (e.g. [70]; [33]; [60] [71] [72] [73]; but see also [74]). 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. [32]). 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. Whilst 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. [68]; [75]; but see also [76] who review the topic in general).

In our model, the fossil record consists of plesions that are generated by a rapid rate of speciation in an underlying but unseen stem lineage. In principal at least, each of these speciation events (at least as recognised 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. Averaging over all lineages (i.e. the rapid stems and the average plesions) should give a characteristic decline through time similar to that of diversification rates (see e.g. fig 3D of [60]). The notable study of lungfish evolution through time by Lloyd *et al.* [77] reconstructed rates of phenotypic evolution through time and indeed noted such a decline (see their fig. 4; note that it also shows a characteristic post-extinction spike). 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, whilst the plesions appear to show no such pattern (the authors do not differentiate between the two in their analysis of the decline in rates). 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 of the POTPa (compare their fig. 1 with our Fig. 3). Clearly, it would be possible to test this pattern in other groups too.

As well as the POTPa, the LCE also generates an effect on rates of phenotypic evolution, because it imposes an extra high rate on the lineages at the beginning which would then decline. This may be noticeable in the important recent study on rates of evolution in the arthropods [60] that shows a lineage-rate decline. Thus in large clades, one would expect to see both effects, which would have both more change across the entire clade and faster rates at the beginning.

9.2 The POTPa and molecular clocks

Finally, and perhaps most controversially, these influences may also affect molecular clock estimates of clade divergence times. Various studies have shown that rates of molecular change are at least loosely correlated with diversification rates (e.g. [78]; [79]; [80]; [81]; [82]). Whilst apparently not universal (e.g. [83]), such a linkage, for whatever reason,

would also create a POTPa effect in molecular clocks: early branches would tend to accumulate change faster than the long-term average because they would also tend to have net higher speciation rates. Such a pattern has indeed been noted ([84]; [60], once again with an overall pattern strikingly similar to a POTPa effect ([60], their fig 3A - but see also [85], who failed to find such an effect in a study of explosive radiations on islands). Whilst such correlations are typically weak, the large size of the POTPa may indeed generate enough of an effect to be noticeable.

Rates of molecular evolution are strictly a lineage effect, and in our model, as these all keep broadly similar (but high) rates until near the present, there should be no POTPa slow-down visible (averaging over existing lineages will not create any change through time, unlike the case with rates of phenotypic change which can have extinct branches). However, the LCE preferentially influences the early parts of diversification, and this *would* create a slow-down through time, with the effect of making molecular clocks over-estimate 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 [60] (but only in large clades, such as the arthropods [60] - and, of course, the animals themselves).

10 The end of the POTPa

As the Recent is closely approached, plesions that would normally have little chance of surviving until today start rather suddenly to do so. As a result, the average rate of speciation in living lineages drops quickly towards the background level of λ until as the present is reached, all species become living lineages. One way of thinking about this is to consider that the ordering effect of the POTPa relies on having information what the future holds. As the present approaches however, our knowledge about the future becomes shrouded in uncertainty, and our more complete knowledge about the past is of no help in predicting what will happen in the future: as a result, the POTPa effects dissipate.

11 Summary

In this paper we have explored the patterns of diversification that can be generated by a retrospective view of a purely homogeneous process of diversification. These patterns can be substantial and highly non-homogeneous, and it is essential to understand these "null" hypotheses before considering causal explanations for any residuals (c.f. [13]. Patterns of diversification throughout time have been much discussed in the literature (e.g. [68]), with a common pattern being seen that diversification rates are high at the beginning of major evolutionary radiations. Various mechanisms for such an effect 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 purely down to the "push of the past"? 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 ex-

tinctions. 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. [86] 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) [17], 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 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. [87] [60]), our analysis shows this to be particularly fraught with difficulties: the history of life was written by the victors.

References

[1] John Phillips. Palaeozoic series. *Penny Cyclopedia*, 17:153–154, 1840.

[2] Graham E Budd and Sören Jensen. A critical reappraisal of the fossil record of the bilaterian phyla. *Biological Reviews*, 75(2):253–295, 2000.

[3] Michael J Sanderson and Michael J Donoghue. Shifts in diversification rate with the origin of angiosperms. *Science*, pages 1590–1590, 1994.

[4] Douglas H Erwin. *The great Paleozoic crisis: life and death in the Permian*. Columbia University Press, 1993.

[5] Matt Friedman. Explosive morphological diversification of spiny-finned teleost fishes in the aftermath of the end-Cretaceous extinction. *Proceedings of the Royal Society of London B: Biological Sciences*, page rspb20092177, 2010.

[6] Pincelli M Hull, Richard D Norris, Timothy J Bralower, and Jonathan D Schueth. A role for chance in marine recovery from the end-Cretaceous extinction. *Nature Geoscience*, 4(12):856, 2011.

[7] Tanja Stadler, Daniel L Rabosky, Robert E Ricklefs, and Folmer Bokma. On age and species richness of higher taxa. *The American Naturalist*, 184(4):447–455, 2014.

[8] Shaopeng Wang, Anping Chen, Jingyun Fang, and Stephen W Pacala. Speciation rates decline through time in individual-based models of speciation and extinction. *The American Naturalist*, 182(3):E83–E93, 2013.

[9] Bruce S Lieberman. Analyzing speciation rates in macroevolutionary studies. *Fossils, phylogeny, and form: An analytical approach*, pages 323–339, 2001.

[10] Thomas HG Ezard, Paul N Pearson, Tracy Aze, and Andy Purvis. The meaning of birth and death (in macroevolutionary birth–death models). *Biology letters*, 8(1):139–142, 2012.

[11] Rampal S Etienne, Bart Haegeman, Tanja Stadler, Tracy Aze, Paul N Pearson, Andy Purvis, and Albert B Phillimore. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. In *Proc. R. Soc. B*, page rspb20111439. The Royal Society, 2012.

[12] Arne Mooers, Olivier Gascuel, Tanja Stadler, Heyang Li, and Mike Steel. Branch lengths on birth–death trees and the expected loss of phylogenetic diversity. *Systematic biology*, 61(2):195–203, 2011.

[13] Tanja Stadler and Mike Steel. Distribution of branch lengths and phylogenetic diversity under homogeneous speciation models. *Journal of theoretical biology*, 297:33–40, 2012.

[14] Robert E Ricklefs. Estimating diversification rates from phylogenetic information. *Trends in Ecology & Evolution*, 22(11):601–610, 2007.

[15] T Stadler. Recovering speciation and extinction dynamics based on phylogenies. *Journal of evolutionary biology*, 26(6):1203–1219, 2013.

[16] Sean Nee, Robert M May, and Paul H Harvey. The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 344(1309):305–311, 1994.

[17] Paul H Harvey, Robert M May, and Sean Nee. Phylogenies without fossils. *Evolution*, 48(3):523–529, 1994.

[18] Sean Nee. Birth-death models in macroevolution. *Annu. Rev. Ecol. Evol. Syst.*, 37:1–17, 2006.

[19] Sean Nee. Inferring speciation rates from phylogenies. *Evolution*, 55(4):661–668, 2001.

[20] Rampal S Etienne and James Rosindell. Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. *Systematic Biology*, 61(2):204–213, 2012.

[21] Daniel L Rabosky and Irby J Lovette. Density-dependent diversification in north american wood warblers. *Proceedings of the Royal Society of London B: Biological Sciences*, 275(1649):2363–2371, 2008.

[22] Bruce J MacFadden and Richard C Hulbert. Explosive speciation at the base of the adaptive radiation of Miocene grazing horses. *Nature*, 336(6198):466–468, 1988.

[23] Maureen A O'leary, Jonathan I Bloch, John J Flynn, Timothy J Gaudin, Andres Giallombardo, Norberto P Giannini, Suzann L Goldberg, Brian P Kraatz, Zhe-Xi Luo, and Jin Meng. The placental mammal ancestor and the post–K-Pg radiation of placentals. *Science*, 339(6120):662–667, 2013.

[24] Johannes Krause, Tina Unger, Aline Noçon, Anna-Sapfo Malaspinas, Sergios-Orestis Kolokotronis, Mathias Stiller, Leopoldo Soibelzon, Helen Spriggs, Paul H Dear, and Adrian W Briggs. Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evolutionary Biology*, 8(1):220, 2008.

[25] Alan Feduccia. Explosive evolution in Tertiary birds and mammals. *Science*, 267(5198):637, 1995.

[26] John Alroy. The fossil record of North American mammals: evidence for a Paleocene evolutionary radiation. *Systematic Biology*, 48(1):107–118, 1999.

[27] Graham E Budd. At the origin of animals: the revolutionary cambrian fossil record. *Current genomics*, 14(6):344–354, 2013.

[28] M Paul Smith and David AT Harper. Causes of the Cambrian explosion. *Science*, 341(6152):1355–1356, 2013.

[29] Douglas H Erwin. The end and the beginning: recoveries from mass extinctions. *Trends in Ecology & Evolution*, 13(9):344–349, 1998.

[30] Eric H Davidson and Douglas H Erwin. Gene regulatory networks and the evolution of animal body plans. *Science*, 311(5762):796–800, 2006.

[31] S Maruyama, Y Sawaki, T Ebisuzaki, M Ikoma, S Omori, and T Komabayashi. Initiation of leaking Earth: An ultimate trigger of the Cambrian explosion. *Gondwana Research*, 25(3):910–944, 2014.

[32] Douglas H Erwin. Early introduction of major morphological innovations. *Acta Palaeontologica Polonica*, 38(3-4), 1993.

[33] Graeme T Lloyd, Steve C Wang, and Stephen L Brusatte. Identifying heterogeneity in rates of morphological evolution: discrete character change in the evolution of lungfish (Sarcopterygii; Dipnoi) . *Evolution*, 66(2):330–348, 2012.

[34] David B Wake, Gerhard Roth, and Marvalee H Wake. On the problem of stasis in organismal evolution. *Journal of theoretical Biology*, 101(2):211–224, 1983.

[35] David Jablonski, Kaustuv Roy, James W Valentine, Rebecca M Price, and Philip S Anderson. The impact of the pull of the recent on the history of marine diversity. *Science*, 300(5622):1133–1135, 2003.

[36] S Blair Hedges, Julie Marin, Michael Suleski, Madeline Paymer, and Sudhir Kumar. Tree of life reveals clock-like speciation and diversification. *Molecular biology and evolution*, 32(4):835–845, 2015.

[37] Charles R Marshall and Tiago B Quental. The uncertain role of diversity dependence in species diversification and the need to incorporate time-varying carrying capacities. *Phil. Trans. R. Soc. B*, 371(1691):20150217, 2016.

[38] Rampal S Etienne, Alex L Pigot, and Albert B Phillimore. How reliably can we infer diversity-dependent diversification from phylogenies? *Methods in Ecology and Evolution*, 7(9):1092–1099, 2016.

[39] Daniel L Rabosky. Extinction rates should not be estimated from molecular phylogenies. *Evolution*, 64(6):1816–1824, 2010.

[40] Andy Purvis. Phylogenetic approaches to the study of extinction. *Annual Review of Ecology, Evolution, and Systematics*, 39, 2008.

[41] Sean Nee, Edward C Holmes, Robert M May, and Paul H Harvey. Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions: Biological Sciences*, pages 77–82, 1994.

[42] Albert B Phillimore and Trevor D Price. Density-dependent cladogenesis in birds. *PLoS biology*, 6(3):e71, 2008.

[43] Niklas Wahlberg, Julien Leneveu, Ullasa Kodandaramaiah, Carlos Peña, Sören Nylin, André VL Freitas, and Andrew VZ Brower. Nymphalid butterflies diversify following near demise at the Cretaceous/Tertiary boundary. *Proceedings of the Royal Society of London B: Biological Sciences*, 276(1677):4295–4302, 2009.

[44] Veronika Boskova, Sebastian Bonhoeffer, and Tanja Stadler. Inference of epidemiological dynamics based on simulated phylogenies using birth-death and coalescent models. *PLoS computational biology*, 10(11):e1003913, 2014.

[45] Graham E Budd and Illiam SC Jackson. Ecological innovations in the Cambrian and the origins of the crown group phyla. *Phil. Trans. R. Soc. B*, 371(1685):20150287, 2016.

[46] David M Raup. On the early origins of major biologic groups. *Paleobiology*, 9(2):107–115, 1983.

[47] Richard R Strathmann and Montgomery Slatkin. The improbability of animal phyla with few species. *Paleobiology*, 9(2):97–106, 1983.

[48] Graham E Budd. Tardigrades as 'stem-group arthropods': the evidence from the Cambrian fauna. *Zoologischer Anzeiger-A Journal of Comparative Zoology*, 240(3):265–279, 2001.

[49] AJ Craske and RPS Jefferies. A new mitrate from the Upper Ordovician of Norway, and a new approach to subdividing a plesion. *Palaeontology*, 32(1):69–99, 1989.

[50] GU Yule. A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS Philosophical Transactions of the Royal Society of London B213: 21-87. *Philosophical Transactions of the Royal Society of London B*, 1924, 1924.

[51] Graham E Budd. The Cambrian fossil record and the origin of the phyla. *Integrative and Comparative Biology*, 43(1):157–165, 2003.

[52] Nicholas J Butterfield. Macroevolution and macroecology through deep time. *Palaeontology*, 50(1):41–55, 2007.

[53] Stephen Jay Gould, David M Raup, J John Sepkoski, Thomas JM Schopf, and Daniel S Simberloff. The shape of evolution: a comparison of real and random clades. *Paleobiology*, 3(1):23–40, 1977.

[54] John Huss. The shape of evolution: the MBL model and clade shape, journal = The paleobiological revolution: Essays on the growth of modern paleontology. page 568, 2009.

[55] J John Sepkoski. Rates of speciation in the fossil record. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 353(1366):315–326, 1998.

[56] Steven M Stanley, Philip W Signor, Scott Lidgard, and Alan F Karr. Natural clades differ from "random" clades: simulations and analyses. *Paleobiology*, 7(1):115–127, 1981.

[57] Leigh Van Valen. A new evolutionary law. *Evol Theory*, 1:1–30, 1973.

[58] David Jablonski. Extinctions in the fossil record. *Philosophical Transactions: Biological Sciences*, pages 11–17, 1994.

[59] Richard A Fortey, Jennifer Jackson, and Jan Strugnell. *Phylogenetic fuses and evolutionary 'explosions': conflicting evidence and critical tests*, pages 41–65. CRC Press, 2003.

[60] Michael SY Lee, Julien Soubrier, and Gregory D Edgecombe. Rates of phenotypic and genomic evolution during the Cambrian explosion. *Current Biology*, 23(19):1889–1895, 2013.

[61] John Alroy. Dynamics of origination and extinction in the marine fossil record. *Proceedings of the National Academy of Sciences*, 105(Supplement 1):11536–11542, 2008.

[62] Stephen Jay Gould. *Wonderful life: the Burgess Shale and the nature of history*. WW Norton and Company, 1990.

[63] John Alroy, Martin Aberhan, David J Bottjer, Michael Foote, Franz T Fürsich, Peter J Harries, Austin JW Hendy, Steven M Holland, Linda C Ivany, and Wolfgang Kiessling. Phanerozoic trends in the global diversity of marine invertebrates. *Science*, 321(5885):97–100, 2008.

[64] John Alroy. Accurate and precise estimates of origination and extinction rates. *Paleobiology*, 40(3):374–397, 2014.

[65] Mike Foote. Estimating taxonomic durations and preservation probability. *Paleobiology*, 23(3):278–300, 1997.

[66] Mike Foote and David M Raup. Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology*, 22(2):121–140, 1996.

[67] Richard J Twitchett. Incompleteness of the Permian-Triassic fossil record: a consequence of productivity decline? . *Geological Journal*, 36(3-4):341–353, 2001.

[68] Melanie J Hopkins and Andrew B Smith. Dynamic evolutionary change in post-Paleozoic echinoids and the importance of scale when interpreting changes in rates of evolution. *Proceedings of the National Academy of Sciences*, 112(12):3758–3763, 2015.

[69] Luke J Harmon, James A Schulte, Allan Larson, and Jonathan B Losos. Tempo and mode of evolutionary radiation in iguanian lizards. *Science*, 301(5635):961–964, 2003.

[70] T Stanley Westoll. On the evolution of the Dipnoi. *Genetics, paleontology and evolution*, pages 121–184, 1949.

[71] Marcello Ruta, Peter J Wagner, and Michael I Coates. Evolutionary patterns in early tetrapods. I. Rapid initial diversification followed by decrease in rates of character change. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1598):2107–2111, 2006.

[72] Stephen L Brusatte, Michael J Benton, Graeme T Lloyd, Marcello Ruta, and Steve C Wang. Macroevolutionary patterns in the evolutionary radiation of archosaurs (Tetrapoda: Diapsida). *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 101(3-4):367–382, 2010.

[73] Mario Bronzati, Felipe C Montefeltro, and Max C Langer. Diversification events and the effects of mass extinctions on Crocodyliformes evolutionary history. *Royal Society open science*, 2(5):140385, 2015.

[74] Luke J Harmon, Jonathan B Losos, T Jonathan Davies, Rosemary G Gillespie, John L Gittleman, W Bryan Jennings, Kenneth H Kozak, Mark A McPeek, Franck Moreno Roark, and Thomas J Near. Early bursts of body size and shape evolution are rare in comparative data. *Evolution*, 64(8):2385–2396, 2010.

[75] Dean C Adams, Chelsea M Berns, Kenneth H Kozak, and John J Wiens. Are rates of species diversification correlated with rates of morphological evolution? *Proceedings of the Royal Society of London B: Biological Sciences*, page rspb. 2009.0543, 2009.

[76] Daniel L Rabosky and Dean C Adams. Rates of morphological evolution are correlated with species richness in salamanders. *Evolution*, 66(6):1807–1818, 2012.

[77] Graeme T Lloyd. Estimating morphological diversity and tempo with discrete character-taxon matrices: implementation, challenges, progress, and future directions. *Biological Journal of the Linnean Society*, 118(1):131–151, 2016.

[78] Robert Lanfear, Simon YW Ho, Dominic Love, and Lindell Bromham. Mutation rate is linked to diversification in birds. *Proceedings of the National Academy of Sciences*, 107(47):20423–20428, 2010.

[79] T Jonathan Davies, Vincent Savolainen, Mark W Chase, Justin Moat, and Timothy G Barraclough. Environmental energy and evolutionary rates in flowering plants. *Proceedings of the Royal Society of London B: Biological Sciences*, 271(1553):2195–2200, 2004.

[80] Timothy G Barraclough and Vincent Savolainen. Evolutionary rates and species diversity in flowering plants. *Evolution*, 55(4):677–683, 2001.

[81] Mark Pagel, Chris Venditti, and Andrew Meade. Large punctuational contribution of speciation to evolutionary divergence at the molecular level. *Science*, 314(5796):119–121, 2006.

[82] Lindell Bromham. Molecular clocks and explosive radiations. *Journal of molecular evolution*, 57(1):S13–S20, 2003.

[83] Xavier Goldie, Robert Lanfear, and Lindell Bromham. Diversification and the rate of molecular evolution: no evidence of a link in mammals. *BMC evolutionary biology*, 11(1):286, 2011.

[84] Stéphane Aris-Brosou and Ziheng Yang. Bayesian models of episodic evolution support a late Precambrian explosive diversification of the Metazoa. *Molecular biology and evolution*, 20(12):1947–1954, 2003.

[85] Lindell Bromham and Megan Woolfit. Explosive radiations and the reliability of molecular clocks: island endemic radiations as a test case. *Systematic Biology*, 53(5):758–766, 2004.

[86] Toke Reichstein and Michael S Dahl. Are firm growth rates random? analysing patterns and dependencies. *International Review of Applied Economics*, 18(2):225–246, 2004.

[87] Derek EG Briggs, Richard A Fortey, and Matthew A Wills. Morphological disparity in the cambrian. *Science*, 256(5064):1670–1673, 1992.