Detecting shifts in diversity limits from molecular phylogenies: what can we know?

Lynsey McInnes1,2,*, C. David L. Orme1,2 and A. Purvis1
1Division of Biology, Imperial College London, Silwood Park, Ascot, Berkshire SL5 7PY, UK
2Grantham Institute for Climate Change, Imperial College, London SW7 2AZ, UK

Large complete species-level molecular phylogenies can provide the most direct information about the macroevolutionary history of clades having poor fossil records. However, extinction will ultimately erode evidence of pulses of rapid speciation in the deep past. Assessment of how well, and for how long, phylogenies retain the signature of such pulses has hitherto been based on a—probably untenable—model of ongoing diversity-independent diversification. Here, we develop two new tests for changes in diversification 'rules' and evaluate their power to detect sudden increases in equilibrium diversity in clades simulated with diversity-dependent speciation and extinction rates. Pulses of diversification are only detected easily if they occurred recently and if the rate of species turnover at equilibrium is low; rates reported for fossil mammals suggest that the power to detect a doubling of species diversity falls to 50 per cent after less than 50 Myr even with a perfect phylogeny of extant species. Extinction does eventually draw a veil over past dynamics, suggesting that some questions are beyond the limits of inference, but sudden clade-wide pulses of speciation can be detected after many millions of years, even when overall diversity is constrained. Applying our methods to existing phylogenies of mammals and angiosperms identifies intervals of elevated diversification in each.

Keywords: diversification pulses; diversity dependence; extinction; mammals; angiosperms

1. INTRODUCTION

Clades are unlikely to have diversified at a constant per-lineage rate over deep time [1]. The fossil record has long been used to study tempo of evolution [2,3], but its incompleteness and temporal biases have often limited the strength of inference about macroevolutionary dynamics [4,5]. Time-calibrated phylogenies of extant taxa provide another window on diversification (reviewed in [6]). Analyses of such phylogenies have recently been used to identify pulses of speciation wherein clade richness rapidly increases [7], and to test whether such pulses coincide with the times of climatic shifts (e.g. [8]), tectonic movement [9] or mountain uplift [10]. Because these phylogenies do not contain extinct lineages, the signal of pulses will tend to be eroded by subsequent extinction [11,12].

Under what circumstances might we expect a diversification pulse to still be detectable? There has been some investigation of the power of phylogenies to reveal temporal changes in diversification [13–15], but many studies have relied on two simplifying assumptions. First, they have assumed that extinction played a negligible role in producing current patterns of diversity (reviewed in [6]), despite evidence from the fossil record that extinction is important (e.g. [16,17]). The lack of a record of extinction, therefore, results in a biased account of diversification [12,18]. Second, many studies have implicitly assumed that per-lineage rates of speciation and extinction have been constant, implying that there is no upper limit to clade diversification (e.g. [8,19]).

More recent studies have shown that patterns of clade accumulation through time are often consistent with models incorporating diversity limits [20–23]. Diversification slows as available geographical or niche space becomes saturated. Beyond this point, turnover continues but clade size stabilizes or increases more slowly (e.g. [1,20–25]). Controversy persists as to whether declining diversification rates are driven by decreasing speciation, increasing extinction rates or a combination [11,26–29], although declines driven by increasing extinction rates may not be detectable [11,27]. Additionally, the ability to detect declines depends on a high ratio of initial speciation rate to equilibrium extinction rate and is greatest when clades first reach equilibrium diversity [12,27]. As yet, however, there has been little exploration of how improved models using equilibrial diversity affect our ability to detect changes in diversification (but see [30]).

Equilibrium diversity emerges from the attributes of a clade and extrinsic factors including climate and the nature and size of the area available for diversification [1,21,25,31]. Extrinsic changes that add or remove suitable habitat, such as major climatic change, can therefore affect diversity limits [24,32,33], even without intrinsic changes such as key innovations. The fossil record indicates many biotic turnovers induced by climate change (e.g. mammalian responses reviewed in [33]). In particular, the Palaeocene–Eocene Thermal Maximum (PETM; 55–55.5 Myr ago) is associated with the dispersal of the modern mammalian orders Primates, Artiodactyla and Perissodactyla into new continents with subsequent rapid diversification [34]. A time-calibrated phylogeny of extant mammals shows an increase in diversification rate around the same time [7], but for how long should phylogenies retain a signal of past events?

* Author for correspondence (lynsey.mcinnes01@imperial.ac.uk).

Here, we simulate diversity-dependent cladogenesis to assess when a major change in maximum equilibrium diversity can be detected from a perfect molecular phylogeny of extant taxa. In the light of recent evidence for the importance of diversity limits to clade diversification [21,22] and the impact of discrete events on clade diversification [8–10], we focus on detecting transient pulses of diversification associated with changing equilibria rather than declines in diversification (e.g. [12,20,27]) or shifts from one constant rate to another (e.g. [8,15]). We propose and use two new statistical tests to explore how the timing and size of changes in diversity, and the background turnover rate, affect our ability to recover rule changes from a time-calibrated phylogeny of extant species. Finally, we apply these tests to a species-level supertree of mammals [35] and a family-level supertree of angiosperms [36].

2. METHODS

Evidence for diversity-dependent cladogenesis is mounting (e.g. [21,22,24,31]), but uncertainty remains over the form of diversity-dependence [15,37] and whether it acts through speciation rates, extinction rates or both [26,28,29,38,39]. Here, we use an equal-rates logistic model of diversity-dependent cladogenesis [37] and vary instantaneous per-lineage rates of both speciation (λ) and extinction (μ) as a function of initial speciation rate (b) and extinction rate at equilibrium (d), as the number of extant lineages (N) approaches the maximum possible diversity of the clade (M) [1]:

\[ \lambda = b^* \left(1 - \frac{N}{M}\right) \]

and

\[ \mu = d^* \frac{N}{M}. \]

As N rises, extinction increases and speciation decreases until N fluctuates stochastically around an equilibrium diversity, \( K_e \) defined by \( bM/(b + d) \), with a turnover rate of \( d/(b + d) \). Thus, when \( d > 0 \), M will not be reached [29]. This model differs from the critical birth–death model (where \( \lambda = \mu \)) which has no diversity limit, and from a Moran process, where \( N \) is deterministically held constant with every extinction met by immediate speciation (both reviewed in [6]). We use stochastic simulation, drawing each species’ waiting time to the next speciation or extinction using the current \( N \), and re-drawing waiting times after each event.

To this basic model, we introduce an increase in maximum diversity—a rule change—from \( M_1 \) to \( M_2 \) after a set time \( T_1 \) from the start of the simulation \( (T_0) \), resulting in an immediate increase in speciation and decrease in extinction. The simulation continues under this higher \( M \) for a further time period \( T_2 \). In our first set of simulations, we examine how variation in the size of the shift in \( M \) and the rate of turnover affect detection of the change in diversification. Simulations start with a single species at \( T_0 \) and we fix \( M_1 = 500 \), \( b = 1 \), and vary \( M_2 \) from 600 to 1000 in steps of 100 and \( d \) from 0.1 to 0.5 in steps of 0.1. These values of \( d \) correspond to turnover rates at equilibrium between 0.091 and 0.333, bracketing estimates from the mammalian fossil record of approximately 0.24 species per-lineage per million years [24]. We run the simulation over 110 Myr with the shift from \( M_1 \) to \( M_2 \) occurring at the PETM (55 Ma); both \( T_1 \) and \( T_2 \) are therefore 55 Myr in length. Equilibrium diversity levels are reached within 10 Myr of \( T_0 \) and \( T_1 \) (figure 1). We also investigate decreases in maximum diversity, using \( b = 1 \), \( T_1 = 55 \) Myr, \( T_2 = 55 \) Myr, \( M_1 = 1000 \), \( M_2 = 500 \) and varying \( d \) from 0.1 to 0.5 in steps of 0.1. These ‘downshift’ simulations represent a permanent reduction in diversity limits.

We consider two further scenarios, looking at how the signal deteriorates over time since the shift and at the signal left by non-selective mass extinctions. Both these simulations used \( b = 1 \) with \( d \) varying between 0.1 and 0.5. The first scenario additionally varies \( T_2 \) between 10 and 100 Myr with \( T_1 = 55 \) Myr, \( M_1 = 500 \) and \( M_2 = 1000 \). For mass extinction, we randomly remove 50 per cent of extant species from equilibrium diversity \( (M_1 = 1000) \) at \( T_1 = 55 \) Myr, and then allow diversity to recover from around 500 to then remain at the original equilibrium \( (M_1 = M_2) \) for a further 55 Myr \( (T_2) \). We compare this scenario to previous simulations where, after an initial 55 Myr \( (T_1) \), equilibrium diversity simply increases from 500 \( (M_2) \) to 1000 \( (M_2) \) until the end of the simulation 55 Myr later \( (T_2) \). We consider only random extinction for simplicity ([30,40] discuss the signal left by non-selective mass extinction). Finally, we investigate how detectability changes when only a subset of lineages is affected by the rule change with the remaining lineages continuing to diversify under \( M_1 \) (details in the electronic supplementary material, S1).

To assess the type I error rates and power of our detection methods, we also simulate a set of null trees with no diversification shift. All trees have \( b = 1 \), \( M = 500 \) and we vary \( d \) between 0.1 and 0.5 and tree age between 65 and 155 Myr, as above. In all scenarios, we simulate 100 replicate trees under each set of parameters; in total 14 000 trees are analysed.

(a) Detecting changes in diversification

Our simulated phylogenies contain all extant and extinct lineages, and plots of the number of then-extant lineages

Figure 1. Example lineage-through-time plots showing the loss of information when going from a complete phylogeny with all extinct and extant taxa included (solid line) to the best reconstructed phylogeny possible (all extinct taxa pruned from the tree, dashed line). The tree was grown under the following parameters: \( b = 1 \), \( d = 0.1 \), \( M_1 = 500 \), \( M_2 = 1000 \), \( T_1 = T_2 = 55 \) Myr.
against time clearly show the effects of diversification shifts (figure 1, solid line). To represent the best data that molecular phylogenies could provide, we only analyse phylogenies from which all extinct lineages have been removed (figure 1, dashed line). The resulting complete, perfectly time-calibrated molecular trees are, of course, far better than data currently available [41]. However, our primary interest is in identifying situations in which there is no prospect of ever confidently making inferences, rather than those where potentially recoverable signal is erased by incomplete sampling or imperfect dates (e.g. [14,42]).

We developed two methods to detect the signal of shifts in diversification within phylogenies, both of which use temporal windows sliding across evolutionary history to identify anomalous 1 Myr intervals. In both cases, we discard estimates for the first and last 20 Myr because they are biased: estimates in the first period are biased upwards because clades that happened to diversify slowly initially are likely to have gone extinct before the present, while the second period includes new lineages that will go extinct but have not yet done so [6]. Analyses are therefore based on a 70 Myr sequence centred on the change in diversification at 55 Myr. When analysing simulations where \( T_3 \) varies, we include the final 35 Myr of \( T_1 \) and only truncate \( T_2 \) where it is longer than 35 Myr; some sequences are therefore shorter than 70 Myr.

The first method extends Pybus & Harvey’s \( \gamma \) statistic [14]. Given a sequence of internode distances, \( \gamma \) measures whether those nodes are clustered towards the start (\( \gamma < 0 \)) or end (\( \gamma > 0 \)) of the sequence, compared with the expectations of a pure-birth process. Although it is more typically used to measure changes in diversification across a whole tree, it also applies to a sequence of nodes drawn from a time window. A diversification pulse within a 1 Myr slice will lead to a positive \( \gamma \) for a longer window that ends with that 1 Myr, and to a negative \( \gamma \) for a longer window that starts with the pulse (figure 2). The difference (\( \Delta \gamma \)) in \( \gamma \) between the earlier and later window therefore reflects changes in the per-lineage rates, with a negative \( \Delta \gamma \) indicating a diversification pulse in the 1 Myr where the windows overlap (figure 2). We simulated the distribution of \( \Delta \gamma \) and found it follows a normal distribution under both (a) a constant-rates pure-birth process (\( \Delta = 0.06, T = 110 \) Myr, 100 replicates: \( \text{mean} = -0.0425, \text{s.d.} = 1.65 \)) and (b) single logistic decline (\( b = 1, d = 0.1, T = 110 \) Myr, \( M = 500 \) lineages, 100 replicates: \( \text{mean} = 0.0183, \text{s.d.} = 1.69 \)). The \( \Delta \gamma \) statistic seems to avoid some of the biases exhibited by \( \gamma \), for instance, correlations with clade size [20] and number of branching times in the overlap window (results not shown), presumably because we sample only in narrow intervals and discard the intervals at the beginning and end of each tree. From our simulations, we calculate the significance of \( \Delta \gamma \) at 1 Myr intervals, using local windows of 5 Myr.

Our second method uses the maximum-likelihood (ML) diversification rate estimate for each 1 Myr time slice, \( (n - m)/s \), where \( n \) and \( m \) equal the number of lineages at the end and beginning of the slice, respectively, and \( s \) equals the total branch length within the slice [13]. Because we use phylogenies of currently extant species, lineage number can only increase with time (\( n \geq m \)) and the ML estimates are therefore bounded at zero. When \( M_2 > M_1 \), the expectation is that the slice including the rule change will have a higher rate than its neighbours. To detect elevated rates statistically, we fit generalized additive models (GAMs [43]) to the rate estimates through time, weighting by the number of lineages present at the end of each interval. The smoothed term from the GAM allows local rate heterogeneity in the tree to be modelled (rather than assuming rate constancy apart from the rule change) and hence the significance of outliers can be assessed using Studentized residuals relative to neighbouring time intervals within the tree, rather than relative to the predictions of null models parameterized from the tree. This method removes both the need to estimate background turnover rate and the assumption that there even is a background turnover rate: we are trying to recover major perturbations against a background of rate constancy or rate heterogeneity. However, the smoothing parameter \( k \) in GAMs always needs to be chosen with care [43]: too high and the model will trace all fluctuations, including large outliers; too low and the background rate will not be adequately characterized and, following preliminary tests, we have used \( k = 5 \). For those simulations where \( T_2 < 35 \) Myr (i.e. where the rule change occurs close to the present), we retain all time intervals up to the present.
(see above); this may impede our ability to detect a shift, as any increase in \( M \) will be conflated with the signal from lineages present that are committed to extinction in the future.

We assessed each method's type I error rate as the proportion of null simulations in which a significant shift in diversification rate was detected at the end of \( T_1 \). The power of each method for each scenario was calculated as the number of simulations showing a significant change at the end of \( T_1 \), minus the corresponding type I error rate from null simulations. We also recorded the number of other intervals identified as significant for each parameter combination and used a binomial test to compare this proportion to 0.05. Finally, we applied our methods to a suite of models of exponential diversification to determine whether the type I error rates are still reasonable under these scenarios (and thus that our results were not contingent on there being zero net-diversification apart from at the rule change); see the electronic supplementary material, S1.

(b) Testing our methods with empirical data

We applied both methods to a species-level supertree of mammals [35] and a family-level supertree of angiosperms [36]. It is unlikely that either clade evolved under a single or even two-phase homogeneous diversification process [36,44]; however, we were interested in whether our methods detect anomalous time intervals despite the heterogeneity of process expected within trees of this size. Similarly, it is not problematic that the angiosperm tree has families as tips, as we are testing for diversification pulses deeper in the tree where sampling is effectively complete. We also use one of the most parsimonious supertrees rather than the strict consensus (as in [36]) in order to exploit its completeness and full resolution. Diversification within older families is not reflected in the branching pattern of the supertree, so our methods will not detect diversification bursts exclusively occurring within these families. In mammals, we also applied the method to four sub-clades (marsupials and the three placental groups: Atlantogenata (Afrotheria þ Xenathra), Laurasiatheria and Euarchontoglires) to identify shared or unique pulses of diversification in these groups. The mammal tree has a number of extrapolated dates for taxa lacking sequence data; these tend to be distributed uniformly through time and will bias against detecting diversification pulses. Because of this and other imprecisions in dating phylogenies, we also assessed wider focal intervals than in our simulations, testing both 2 and 5 Myr intervals (for the \( \Delta y \) method we also adjusted the time windows to 10 and 25 Myr, respectively). The results of the three interval sizes are quantitatively similar and we present results only from the 2 Myr window analyses.

We initially tested whether diversification bursts are associated with the Cretaceous–Tertiary boundary (65.5 Ma), the PETM (55–55.5 Myr ago) and the Eocene–Oligocene boundary (33.9 Ma), given their association with responses in the mammalian and angiosperm fossil record [24,33,45]. In addition to testing these specific hypotheses, we also assessed the use of both methods for data exploration by identifying all significant time slices. This requires conducting multiple tests simultaneously. Numerous corrections for multiple testing have been proposed, most adjusting the level at which a result is considered significant to make it more stringent (reviewed in [46]); however, some methods have been criticized as being too conservative [46]. In particular, there is debate on whether it is necessary to control the probability of erroneously rejecting even one of the true null hypotheses (the family-wise error rate, FWE) or whether controlling the expected proportion of falsely rejected hypotheses (false discovery rate, FDR) is adequate [47]. An additional problem with our data is the expected autocorrelation among consecutive time intervals. To address these issues, we report significant intervals: (i) unadjusted for multiple testing, (ii) using the sequential Bonferroni correction (an FWE method), (iii) using the Benjamini and Yekutieli method [47] that controls the FDR when tests are not independent, and (iv) the sequential Bonferroni correction using the effective sample size after autocorrelation is accounted for (see the electronic supplementary material, S2).

3. RESULTS

The power analyses shown in figure 3 demonstrate that statistical tests are more likely to detect a diversification pulse when it is large, when it occurs against a background of low turnover and when it is recent. The two methods also do not differ substantially in the regions in which they demonstrate reasonable power (greater than 0.8). For instance, when turnover rates correspond to those estimated for the mammalian fossil record (\( d \sim 0.24; [24] \)), both methods' power to detect a diversification pulse falls to 50 per cent after 50 Myr (figure 3). Validity testing reveals a similar picture (electronic supplementary material, table S3): the proportion of false positives is lowest where the diversification pulse is most pronounced. Validity of the GAM method decreases as \( T_2 \) and turnover rate increases, whereas validity of the \( \Delta y \) method is less variable through parameter space.

The rebound from a 50 per cent random mass extinction is detected with similar power to a doubling in equilibrium diversity, although significant signal is maintained with higher turnover under the mass extinction scenario (table 1). Conversely, decreases in equilibrium diversity are much less likely to be detected than increases: plots of the ML rate estimates per million year interval (electronic supplementary material, figure S4) indicated that there is no unambiguous signal of a rule change across the five values of \( d \) tested. Rate estimates for the specified interval are zero but this zero rate also occurs in adjacent intervals and—for trees with high \( d \)—is the modal rate until near the present.

Applying \( \Delta y \) to the mammal and angiosperm phylogenies revealed significant increases in diversification for intervals close or coincident with the Eocene–Oligocene boundary (approx. 33.9 Ma) in the Laurasiatherian, marsupial and complete mammal tree. A signal was also found for the first two groups using the GAM method, but these results were not retained in marsupials once we had adjusted for multiple testing. Neither the Cretaceous–Tertiary boundary (65.5 Ma) nor the PETM (55–55.5 Myr ago) were associated with significant changes in diversification once we had adjusted for multiple testing. Looking throughout the phylogenies, significant diversification bursts were recovered (figure 4) in the angiosperm tree and the mammal tree both when analysed as a whole and by superorder, although adjusting for multiple tests removed some intervals. This correction was notably severe for two groups, Atlantogenata and marsupials, where significant intervals
were identified only using the \( \Delta\gamma \) method (see also the electronic supplementary material, table S5).

Finally, our methods are robust in the face of alternative models of diversification. We tested a variety of null and non-null diversity-independent models of cladogenesis and found the type I error rate was reasonable in all cases (see the electronic supplementary material, S1).

4. DISCUSSION

Our results indicate that some large rule changes long ago may still, in principle, be detectable from information gleaned exclusively from extant diversity. It therefore will be possible to combine information on known past events with phylogenies of extant species to ask whether the event had an impact on diversification (see also [8,42]). Our analyses of mammalian and angiosperm supertrees (figure 4 and electronic supplementary material, table S5) also demonstrate that our methods are useful for data exploration. However, our results indicate unambiguously that, under diversity-dependent cladogenesis with rapid turnover, even large increases in \( M \) will often not be apparent from reconstructed phylogenies (figure 3, see also [11,27]). In these cases, strong inferences will require fossil evidence [12,48], though this remains a difficult endeavor because of the paucity of both suitable data and tractable methodologies (but see [49]).

(a) Signals of alternative rule changes

Mass extinctions followed by rebounds produced equivalent signals to increases in \( M \) (table 1). Diversification pulses not associated with a change in \( M \) (e.g. [30]) are also expected to produce similar signals. Additional methods may therefore be required to distinguish between these events, for example, by investigating the time intervals preceding the identified rule change [40,50]. For example, random extinction culls a higher proportion of young lineages and so clades rebounding from extinction events retain a surplus of early lineages in comparison with increases in equilibrium diversity (electronic supplementary material, figure S6). Once an event has been recognized using a time-slice method, differences in the lineage-through-time plots may distinguish different scenarios, perhaps in conjunction with simulations (see also [50]). Any such approach, of course, assumes that, other than at the proposed event, the clade has been diversifying according to some diagnosable model of cladogenesis.

Some events, such as extensive habitat loss or the disappearance of suitable climatic regimes, may lead to

---

Table 1. Comparing power to detect rule changes—doubling-in-\( M \) versus mass extinction and rebound to initial \( M \) (see text for details).

<table>
<thead>
<tr>
<th></th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>doubling-in-( M )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAM</td>
<td>0.97</td>
<td>0.71</td>
<td>0.33</td>
<td>0.21</td>
<td>0.12</td>
</tr>
<tr>
<td>( \Delta\gamma )</td>
<td>0.95</td>
<td>0.62</td>
<td>0.35</td>
<td>0.18</td>
<td>0.16</td>
</tr>
<tr>
<td>mass extinction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAM</td>
<td>0.97</td>
<td>0.79</td>
<td>0.47</td>
<td>0.42</td>
<td>0.28</td>
</tr>
<tr>
<td>( \Delta\gamma )</td>
<td>0.96</td>
<td>0.65</td>
<td>0.44</td>
<td>0.27</td>
<td>0.22</td>
</tr>
</tbody>
</table>

were identified only using the \( \Delta\gamma \) method (see also the electronic supplementary material, table S5).
permanently reduced equilibrium diversity. Such decreases ('downshifts') are difficult to detect using time-slice methods alone (electronic supplementary material, figure S4), with intervals at and around the shift characterised by zero net diversification. This is unsurprising: until extinction forces clade size to the new equilibrium, diversity dependence will constrain further speciation. Higher \( d \) facilitates a more rapid approach to the new equilibrium, but is also associated with larger fluctuations (and more zero-rate intervals) across the rest of tree, eroding any signal of the downshift. Downshifts may be more easily detected when specific traits or conditions have led to a diversity loss in one part of a tree and stasis or gain in another part [44,51]. While periods of low diversification across the tree can be inferred, unambiguously pinpointing clade-wide reductions in \( M \) on a reconstructed phylogeny will be difficult.

**(b) Empirical results**

Application of our methods to the angiosperm family [36] and mammal [35] supertrees highlighted a number of time intervals associated with diversification bursts (figure 4 and electronic supplementary material, table S5), each also associated with fossil evidence for diversification shifts [52–55]. The Eocene–Oligocene boundary was significant in two groups (figure 4): in Laurasiatheria, as artiodactyl radiations replace the previously dominant perissodactyls [52]; and in marsupials corresponding to the origin of crown-group Macropodiformes [54]. The Oligocene (34–23 Ma) is commonly considered the epoch linking the archaic faunas of the hothouse Eocene to the modern faunas that had become well-established by the mid-Miocene. Our data exploration approach identified additional diversification bursts during this epoch for Laurasiatheria and Euarchontoglires, coincident with major clades of rodents and primates dispersing and diversifying into South America (figure 4; [53]). In angiosperms, significant intervals are clustered in the Cretaceous associated with the origin of the major orders [55]. The GAM method also highlights two Cretaceous intervals (100 and 90 Ma) associated with the origin of the mammalian superorders [7]; however, these results must be interpreted conservatively, as deep polytomies will also be

---

**Figure 4.** Diversification bursts in the angiosperm family, mammal and mammalian superorder trees analysed using 2 Myr intervals. Solid triangles denote significant intervals after correcting for multiple tests according to our modified sequential Bonferroni method. Open triangles are the additional intervals, significant only before correction. (a) GAM method: the grey dashed line is the fitted curve. (b) \( \Delta Y \) method. Dotted lines on both plots delimit the boundaries of the Cenozoic epochs tested (Cretaceous–Tertiary boundary, 65 Ma; Palaeocene–Eocene Thermal Maximum, 55 Ma; Eocene–Oligocene boundary, 34 Ma). Dashed lines mark the cut-offs for unanalysed intervals owing to biases (see text). Results for 1 Myr and 5 Myr intervals in the electronic supplementary material, table S5.
recovered as diversification bursts. The $\Delta y$ method identified bursts at 48 and 34 Ma, intervals also associated with low resolution in the tree. Although these polytomies might be hard and indicative of true diversification bursts, further tests are required to identify the underlying process. At the least, our methods underline these unusual intervals in the tree.

Our empirical results attest to the heterogeneity of processes occurring within large trees: while we identify pulses coincident with intervals identified by Bininda-Emonds et al. [7] and Davies et al. [36] using different methods, there are also discrepancies. For instance, the pulse at the PETM for mammals was not recovered, nor were the more recent diversification shifts found by Davies et al. Recent analyses highlight that diversification in both mammals and angiosperms is heavily influenced by available area (in agreement with our model of cladogenesis), but also trait variation, innovations and abiotic conditions [25,44,56]. Although the intervals we identify do correspond to events in the fossil record, a robust understanding of the diversification of large clades will entail incorporating fossil evidence and the effects of intrinsic traits and the extrinsic environment.

(c) Limitations

We have modelled diversity-dependent cladogenesis using a single $M$. Although an improvement on density-independent alternatives, our model is a caricature of the complex interplay between clades and their environments [25,44,56]. A more realistic model may comprise distinct adaptive zones [2], each able to sustain some equilibrium species diversity, and sub-clades diversifying in a diversity-dependent manner within them. While each adaptive zone persists, sub-clades will maintain deep nodes in the reconstructed phylogeny, retaining a signal of their initial diversification into their zones. Even if only certain sub-clades respond to regime change, our methods should then detect the diversification pulses within these sub-clades. Our supplementary analyses showed that, when only a subset of lineages responds to the change in $M$, the pulse is still detectable (electronic supplementary material, S1). Indeed, some of the pulses detected in our empirical trees are probably caused by only a subset of lineages [44,56]. Pinpointing the lineages responsible for diversification pulses requires additional tests incorporating tree topology [44,51]. Our methods are, however, unlikely to perform well if changes are weak or affect different sub-clades in opposite ways.

The above scenario suggests a route by which downshifts could be retained indefinitely in a phylogeny [44]: if incumbency effects [57] make it difficult to remove all members of an established group, extinction will be over-dispersed at the deepest levels of phylogeny and relict taxa will survive to be detected. This speculation does not contradict the evidence that extinction risk is phylogenetically clumped (e.g. [58]): it may be difficult to fully extirpate a group occupying extensive geographical or niche space, whether extinction is random or clumped at low taxonomic levels [57].

Other valuable extensions will be to systematically investigate the effect of incomplete taxon sampling (e.g. [14]) and to develop methods that are robust to dating mismatches between hypothesized impact events and the responding nodes. Moore & Donoghue [42] recently outlined a Bayesian method that accommodates uncertainty stemming from dating, rate and event-timing estimates. Although employing pure birth [6] as their null model of diversification, the explicit incorporation of uncertainty is an important step. Finally, adaptation of existing likelihood (e.g. [8,15]) or approximate Bayesian (e.g. [30]) techniques may provide more robust tools for the detection of rule changes, and may even be able to estimate parameters like turnover rates and the absolute magnitude of the pulse.

(d) Conclusions

Our analysis is timely given the recent rapid proliferation of large phylogenies [7,59] and the recognition that diversity dynamics are probably often at least approximately equilibril [22–24]. As phylogenies become more complete and more accurate, it will be possible to mine them for information about the impact of deep-time events on diversification. Our simulations of simple models of diversity-dependent cladogenesis show that low turnover and large shifts are required if rule changes in the deep past are to be recovered robustly: there is much that we cannot know.

We thank Tim Barraclough, Ian Owens, Ally Phillimore and Alex Pigot for insightful discussions and Ally Phillimore for advice, encouragement and comments on the manuscript. Rea L. A. Kouroumioti carried out preliminary investigation for the effect of $T_2$ on pulse detection and helped develop the GAM method. Finally, we thank Irby Lovette and three anonymous reviewers for constructive comments that improved earlier versions of this manuscript. This work was funded by a Grantham studentship to L.M.

REFERENCES

9 Williams, S. T. & Duda, T. F. 2008 Did tectonic activity stimulate Oligo-Miocene speciation in the Indo-West


