Apparent timescaling of fossil diversification rates is caused by sampling bias

Bouwe R. Reijenga[1]* & Roger A. Close[1]

[1]Department of Earth Sciences, University of Oxford, Oxford OX1 3AN, UK

*Correspondence: bouwe.reijenga@earth.ox.ac.uk

Key words

speciation, extinction, macroevolution, sampling bias, fossil record, Phanerozoic

Highlights

- Diversification rates in phylogenies and the fossil record scale with time
- Rate scaling with time might signal a disconnect between micro- and macroevolution
- Trends are absent from the fossil record when controlling for sampling biases
- Rate scaling in general might be artefactual, or fossils show distinct patterns

Summary

Negative scaling relationships between both speciation and extinction rates on the one hand, and the age or duration of organismal groups on the other, are pervasive and recovered in both molecular phylogenetic and fossil time series. The consistency between molecular and fossil data hints at a universal cause, and potentially to incongruence between micro- and macroevolution. However, the existence of negative rate scaling in fossil time series has not undergone the same level of scrutiny as in molecular data. Here, we analyse the marine fossil record across the last ~535 Ma of the Phanerozoic to investigate the presence and strength of negative rate scaling. We find that negative rate scaling arises under commonly applied age range-based per-capita rates, which do not control for sampling bias, but are severely reduced or absent when metrics are used that do correct for sampling. We further show by simulation that even moderate incomplete sampling of species occurrences through time may induce rate scaling. We thus conclude that there are no significant scaling relationships present in these fossil clades, and that any apparent trend is caused by taxonomic practices and sampling artefacts. If rate scaling appears genuinely in molecular phylogenies, the absence of such a relationship in the fossil record will provide a valuable benchmark and constraint on what processes can cause it.

Results and discussion

The presence of age-rate scaling (ARS) has been a long studied phenomenon in paleobiology[1,2]. Initially, ARS arose as the result of normalising rates of origination and extinction to per-million-year units by dividing by the interval length. Raup and Sepkoski noted that this would be unnecessary given that events are not evenly spread throughout geological stages[2,3], and might be erroneous if diversification (origination – extinction) happens in pulses[4,5]. Because of this independence of the occurrence of origination and extinction events with time, negative scaling between stage duration and extinction rates was found[1].

A recent study by Henao-Diaz et al. has reinvigorated interest in ARS[6], and concluded that ARS was present in both diversification rates estimated from molecular phylogenetic trees and fossil time series. To do this, they compiled a database based on first and last appearances of 90 orders of mammals, plants, and marine animals from Sepkoski's compendium. Diversification rates were calculated via Foote's normalised per-capita metric[7–9]. In contrast to earlier work, ARS was not

studied by comparing the duration of geological stages to per-stage rates, but between the duration of taxonomic orders and their mean diversification rates through time. Their findings suggest that distinct processes operate over different time scales, and might indicate differences between micro- and macroevolutionary processes[10,11]. However, a key concern with the use of Foote's metrics in prior work in general, and when studying ARS, is that those equations are biased by sampling and taxonomic artefacts[3,12,13]. Therefore, the question is whether ARS in fossil diversification rate analyses are caused by failure to control for sampling biases.

Here, we build on the work of Foote[1] and Henao-Diaz et al.[6] by investigating how different sampling artefacts might exaggerate ARS, and if they can be addressed using more appropriate methods. There are two fundamental steps to our analysis: (*i*) We gather occurrence data for marine orders from the Paleobiology Database (n = 408 after data cleaning; http://www.paleobiodb.org). Foote's metrics are only capable of using full age ranges (first and last appearances) of taxa, but fossil occurrence data containing age, locality and identity for each taxon is now readily available. Fossil occurrences also inherently contain information about sampling intensity through time. (*ii*) We compare Foote's per-capita rates to Alroy's second-for-third metrics[12] to calculate instantaneous rates of origination and extinction. Second-for-third metrics require occurrence and use a sliding-window approach that allows the examination of sampling in small time windows [3,13]. Second-for-third metrics do not suffer from the biases that age-range metrics such as Foote's per-capita method suffer from[12,14]. Once we calculate the rates, they are correlated to the duration of their respective taxonomic order, which is the time between the first and last appearance.

We find that when Foote's per-capita metrics are used, we indeed recover strong negative rate-scaling between log-transformed estimates of diversification rates and order duration (Figures 1A-1B). These relationships (origination: $\beta = -0.52$; $R^2 = 0.27$; $P < 0.001$, extinction: $\beta = -0.52$; $R^2 = 0.27$; $P < 0.001$) are stronger than what was reported previously (e.g. origination: $\beta = -0.23$; $P < 0.001$; $R^2 = 0.15$, and extinction: $\beta = -0.25$; $P < 0.01$; $R^2 = 0.13$)[6]. However, when we compare this to Alroy's second-for-third metrics, we find only a very weak relationship for origination ($\beta = -0.16$; $P = 0.014$; $R^2 = 0.02$) and no significant relationship for extinction rate (Figures 1C-1D). Fitting non-linear least squares regressions to the untransformed data reveals a similar pattern, where second-for-third rates show far weaker relationships than per-capita rates (Figures 2A-2D). This suggests that unequal sampling through time and between clades causes the previously-observed ARS in the fossil record.

There are multiple processes relating to sampling and other methodological artefacts that could cause ARS. For instance, Foote showed that normalising rates to per-million-year units results in ARS when geological stages of various durations are compared[1]. For our focal results (Figure 1-2) we focus on instantaneous rates and do not normalise for the interval duration over which rates are calculated, so we know that ARS can arise in per-capita rates even if they are not normalised. However, it might be that normalisation would further increase the strength of ARS. We find that the response to normalisation differs between origination and extinction rate. Whereas origination shows a decrease in R^2 and shallower slopes, extinction shows steeper slopes and increase in R^2 (Figure S1A-H). This could potentially be caused by stronger clustering of extinction events compared to more continuously distributed origination events[4]. However, the effect of normalisation is marginal on the strength of ARS, and ARS is evidently still strong when rates are not normalised (Figure 1-2). Therefore, the smearing out of pulsed rates over longer periods within intervals does not appear to have a strong impact on ARS at the scale of clades, and plotting a ratio (events per Myr^{-1}) against its denominator (Myr) does not have a strong impact on ARS in the fossil record[15,16].

The occurrences that are used for the focal analysis are binned in time intervals based on geological stages. These stages vary in time span, and if a single origination and extinction pulse were

to happen within a short interval, this might result in ARS. To counteract this, we created composite time bins with approximately equal duration of ~11 Myr by lumping the same stage-level intervals[17,18]. Even when using these equal time bins, we still find strong ARS (Figure S2), although both the variance explained and the strength of ARS are decreased substantially for extinction. This is consistent with the idea that extinction more frequently occurs in pulses and origination is more continuous[4,5]. That ARS still arises even though we have binned occurrences in larger time intervals, effectively increasing the minimum age of our orders as more intervals are aggregated together, is surprising as ARS in phylogenetic studies tends to be strongest over the shortest time intervals (<10 Myr)[6,19]. Moreover, the slopes of the relationships (origination: $\beta = -0.54$; $R^2 = 0.30$ and extinction: $\beta = -0.39$; $R^2 = 0.15$) are of comparable strength to those reported for molecular data (e.g. speciation: $\beta = -0.542$; $R^2 = 0.339$ and extinction: $\beta = -0.548$; $R^2 = 0.155$)[6]. This suggests that the time scale over which ARS is measured in the fossil record is not important compared to phylogenetic studies.

What does it mean that the time units over which an ARS is measured are not important for the strength of an ARS? We argue that this indicates that a process, independent of time, affects ARS. This process is most likely to be sampling bias. Drawing a parallel again to phylogenetic studies, Louca et al.[19] argued that there are two distinct mechanisms both relating to the size of phylogenetic trees that can cause ARS. Mechanism 1 suggests that clades might be favoured that happen to be unusually large purely by chance, leading to overestimations in rates. Even if diversification rates are constant (but events are stochastic) between clades, the distribution of realised clades might be truncated in molecular phylogenetic studies. For example, if extinct clades are not considered, or rates for small clades cannot be estimated, this results in a bias towards only realising and estimating rates of clades that have stochastically experienced higher or positive diversification rates. Mechanism 2 suggests that clades that truly have higher rates might be favoured as well. This would mean that even if we accurately estimate the rates of diversification for clades, the phylogenetic trees for which we can estimate rates are those with enough species and, consequently, higher rates. Tree-size sampling bias, or the number of genera within each order, could also cause ARS in the fossil record.

One advantage of using fossil occurrences is that we incorporate extinct clades, which avoids truncating the possible diversification trajectories a clade can follow. There is also less taxonomic or researcher-focused bias – although note preservation and spatial sampling locality biases – in selecting clades to study that are above a certain threshold, as occurrences are described per paleontological collection and not by specifically sequencing a particular clade or species. However, this does not mean that estimating rates for small clades becomes significantly easier. For instance, estimating diversification rates for a monotypic order remains nonsensical. Therefore, arbitrary cut-off points might still be used for methodological reasons, but they are not directly imposed based on sampling (but see sections below).

We investigate genus-diversity threshold sampling bias by both simulation and empirical analysis. The empirical analysis focuses on the data as before, but instead of using all taxonomic orders for which rates can be estimated, we impose a minimum genus-count threshold of 10 per order, in line with previous work that has recovered ARS from fossil data[6]. This analysis suggests that there is a minimal impact on observed ARS when we look at occurrences binned at the stage-level (Figure S3). However, when occurrences are binned in equal-time bins, we do observe a substantial decrease in slope and increase in R^2(Figure S4). Likely, this is caused by there being very few young clades with low richness and low rates surviving for long enough to be sampled, because the minimum duration has increased, and few remain after filtering. Consequently, the time interval over which rates are

measured might have an important impact on how diversity-threshold sampling bias might influence ARS.

Sampling bias in the fossil record cannot be entirely deconstructed to diversity-threshold sampling bias. For species to show up in the fossil record they will need to fossilise, be preserved through time, and be sampled in the present. The simplest way of thinking about sampling a species is by attributing it a probability to be sampled per time interval – ρ, similar to speciation and extinction rates[20]. Sampling as a per-capita rate is not influenced by speciation and extinction, but the longer the duration of a species, the higher the chance will be of sampling it at least once, all else being equal. Sampling any species in an interval thus also depends on when the species arose and when it went extinct (e.g., prior to or within the bin). We hypothesise that this incomplete sampling might especially impact the accuracy of rates estimated for short-lived clades, and may potentially inflate them.

We investigate this hypothesis with a simple heuristic simulation model of species diversification and sampling. The simulation model is based on a birth-death model with speciation and extinction[14]. Simulations are started with N species (N = 1,000 in Figure 3), and speciation and extinction occur in pulses at the boundaries between intervals (E = 0.3 in Figure 3). We further assume that the number of genera is constant through time (i.e., diversity-dependence[3,21]), so that every species that is lost is immediately replaced by a new species. The sampling probability, *s*, of species within and between intervals is uniform. The point of this simulation is not to be exhaustive, nor pin-point the precise sampling artefact causing ARS, but simply to demonstrate whether incomplete sampling could conceivably cause ARS. Moreover, more complex models might induce their own artefacts leading to ARS by, for instance, inducing survivorship biases, which we seek to avoid here.

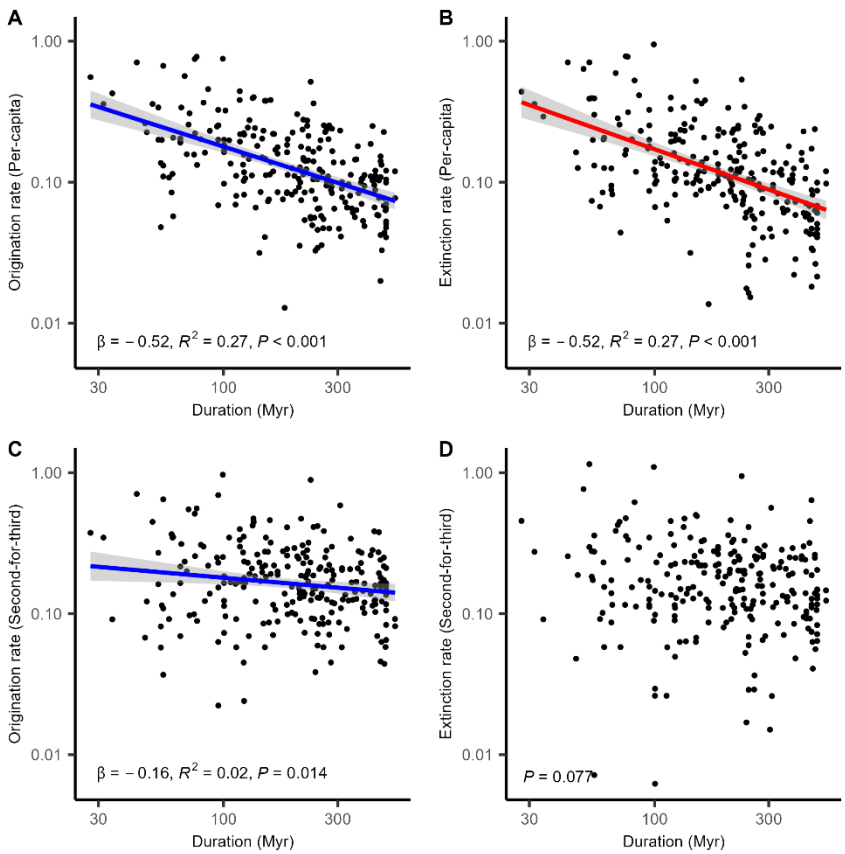
We find that under these simple assumptions, ARS is absent when all species are faithfully sampled (s = 1), and diversification is constant through time and clade duration (Figure 3). However, when we apply lower sampling probabilities (s = 0.5), ARS does manifest, but only under the per-capita rates. This result is constant under a wide breadth of parameter values (Figure S5-S6). This confirms that error in the rates for short-duration clades, caused by incomplete sampling and resulting sampling artefacts, could explain the presence of ARS in the fossil record.

Is the absence of rate scaling in fossil time series surprising? We would argue that it is not. First, rates do not seem to correlate well with the interval duration over which they are measured[3–5]. There would, therefore, be no precedent that older clades would show slower rates, especially if boundaries between stages are recognised in part by faunal turnover[22]. Second, various studies have argued that an increase in diversification rates through the Phanerozoic might have caused ARS, as young clades truly show higher rates[10,19]. Yet, if anything, apart from mass extinction events, the Phanerozoic has shown a marked decrease in the average rates of both origination and extinction with time[3,23]. Third, de Lisle and Svensson[16] have recently argued that ARS is caused by plotting a ratio against its denominator. Foote[1] already showed that this is indeed the case if normalisation to per-million-year rates is performed, and we show that while normalisation does increase the severity of ARS, both per-capita and second-for-third rates are instantaneous and time-independent and should not be *de facto* expected to cause ARS.

Why is there no rate scaling? Diversity of contemporary clades and species have often been argued to be regulated by diversity-dependence, meaning that either origination slows down or extinction increases with increasing diversity[24]. Such a pattern has been observed not only in individual clades, but also across the entire Phanerozoic [17,25]. Yet, precise expectations about how macroevolutionary diversity accumulates, and what patterns are expected to arise, are still poorly-

177 developed. Similarly, the absence of ARS in the fossil record, coupled with the potential presence of
178 the relationship in molecular phylogenies[19], might still suggest interesting avenues of research. For
179 instance, if the existence of ARS in molecular phylogenies can be upheld, why would this only be
180 observed in contemporary phylogenies? One argument about the prevalence of rate scaling for all
181 evolutionary rates has been that it highlights the incongruence between micro- and macroevolution.
182 However, why then would we expect the same pattern to arise in both fossil and molecular data?
183 Specifically, the fossil data studied here is primarily at the genus level, whereas phylogenetic
184 diversification studies most frequently focus on the species level or below. We argue that we would
185 not necessarily expect the same pattern to arise, especially if we consider the persistence of recently
186 arisen species[26]. If most species are unlikely to persist over evolutionary time, which might be the
187 case in some contemporary clades such as sticklebacks[27], would we then not expect the highest rates
188 in young clades with a substantial proportion of incipient species? This diversity, if genera or species
189 are delineated by morphology in the fossil record, would not show up in one record (fossils) but will
190 in the other (DNA). We see this incongruence as motivation for further research to uncover the
191 reasons underlying the different signals present in molecular and fossil data.

192



193

194 **Figure 1** log-transformed diversification rate scaling with clade duration. A and C show mean log-transformed genus-level
195 origination rates (blue), and B and D show mean extinction (red) rates plotted against log clade duration for 245 marine
196 orders sampled across the Phanerozoic. Rates in A and B are calculated via Foote's per-capita metric, and C and D are
197 calculated by Alroy's second-for-third metric. Coloured lines represent significant ordinary least squares regression fits, and
198 shaded areas show 95% confidence intervals. β corresponds to the standardised regression coefficient that is equal to the
199 slope of the relationship between dependent and independent variables when their variance has been standardised to 1.
200 Results are shown for occurrences binned in geological stages.

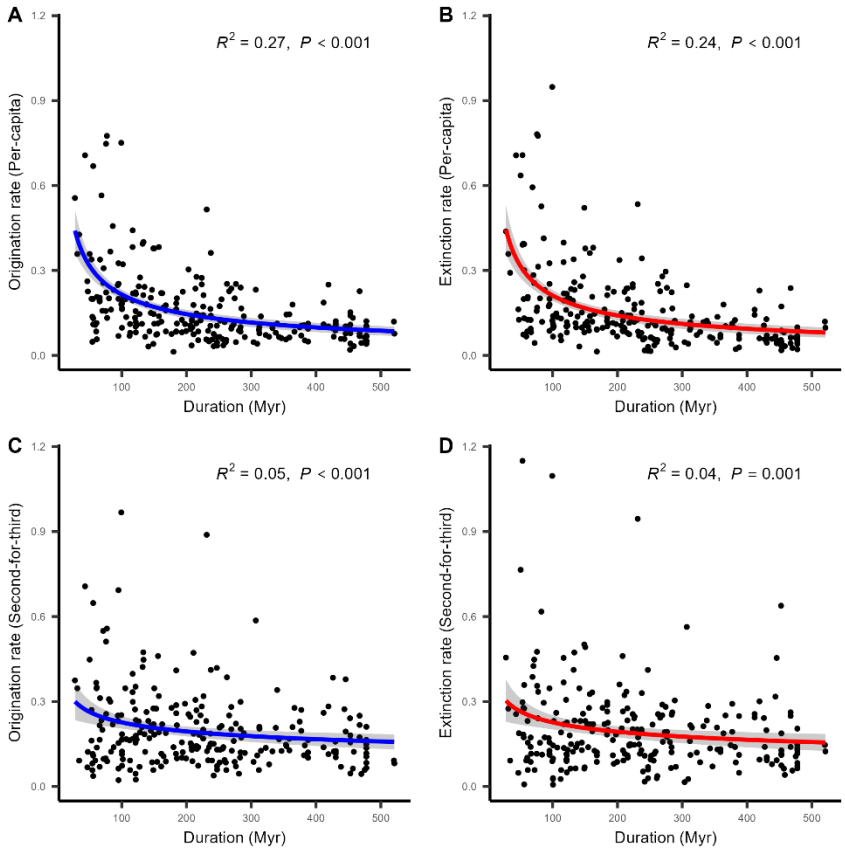


Figure 2 Diversification rate scaling with clade duration. A and C show mean genus-level origination rates (blue), and B and D show mean extinction (red) rates plotted against log clade duration for 245 marine orders sampled across the Phanerozoic. Rates in A and B are calculated via Foote's per-capita metric, and C and D are calculated by Alroy's second-for-third metric. Coloured lines represent non-linear least squares regression fits according to the function $y = ax^b$, and shaded areas represent 95% confidence intervals. P-values are obtained by comparing the non-linear regression to an intercept only model. Results are shown for occurrences binned in geological stages.

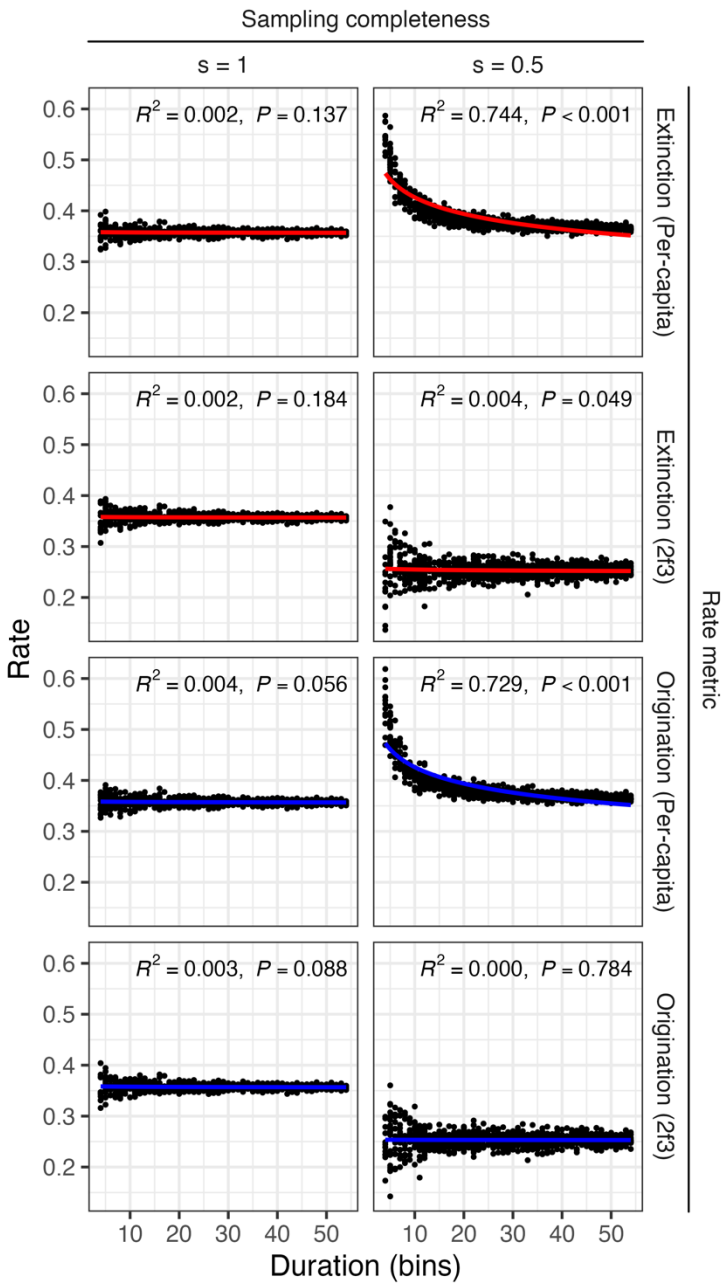


Figure 3 Diversification rate scaling with clade duration for simulations. Average extinction (red) and origination (blue) rates are plotted against their clade duration for 1000 repeats. Column 1 shows the relationships between rates and clade durations when sampling is complete, whereas column 2 shows a scenario of moderate sampling where each species has 50% chance of being preserved and sampled within a bin. Rows represent the rate metrics used for the empirical analysis. Row 1 and 3 use Foote's per-capita metrics, and row 2 and 4 use Alroy's second-for-third metrics that correct for sampling bias. Coloured lines represent non-linear least squares regression fits according to the function $y = ax^b$. 95% confidence intervals are not shown as they overlap almost completely with the coloured lines. P-values are obtained by comparing the non-linear regression to an intercept only model. Simulations are shown for a species pool of N = 1000 and a turnover probability of E = 0.3 per interval.

Methods

Fossil occurrence data assembly

We retrieved fossil occurrence data from the Paleobiology Database using the URL API in identical fashion to previous work[17,28]. We reiterate what has been stated there, but we performed a new download of the data on the 8th of May 2024. Our analyses focus on marine Animalia. For our API calls we include all fields for marine Animalia excluding Chordata, Chordata excluding Tetrapoda, and Tetrapoda. Calls were restricted to marine environments only by the command "envtype=marine". Separately, lists of occurrences were downloaded for Anthozoa, Brachiopoda, Bryozoa, Conodonta, Cephalopoda, Crinoidea, Decapoda, Echinoidea, Gastropoda, Graptolithina, marine Bivalvia, and Trilobita to define taxon sets/clades, and orders were defined by the data field "order".

Diversification rate analyses were conducted at genus level, so our empirical rates pertain to the origination and extinction of genera. Occurrences were filtered to only those that were assigned with certainty to at least genus-level using "idqual=genus_certain" in our calls. We only took into account the latest identification ("idtype=latest") of the occurrence, only valid taxa ("taxon_status=valid"), and removed form taxa (e.g., assigned based on morphological affinity but not phylogeny) or ichnotaxa (e.g., trace fossils) via "pres=regular". We further filtered the occurrences to those that had been assigned to a specific order.

Occurrence were binned via two schemes: (i) 100 geological stages, and (ii) 49 composite equal time bins. The equal time bins each spanning approximately ~11 myr were created by binning geological stages[25]. We assigned the occurrences to bins based on a majority rule. If a time bin contained >50% of the geological time range associated with the occurrence, it was assigned to that bin[25]. Our downstream diversification rate calculations depend on clades ranging over at least 3 time bins. Any order with fewer bins was removed from the analysis. After applying these sifting criteria, the two final datasets based on geological and composite time bins consisted of respectively 610,514 and 646,204 occurrences, 408 and 368 orders, a clade duration range of 5.45-520.99 and 31.60-538.79 Myr and face-value diversity counts ranging from 1 to 1267 and 1 to 1299 genera per order.

Estimating diversification rates

The quantification of changing rates of origination and extinction has been one of the central goals of palaeobiology. A substantial number of metrics have been developed over time, and have been scrutinised equally as often[1,13,29]. Alroy[7] and Foote[8,9] laid the foundation to modern rate metrics that quantify how instantaneous rates change between subsequent time intervals, i.e. geological stages and equal time bins, whereas later iterations and metrics have built on this framework[3,13,21,30]. Here, we explore the effect of Foote's[9] per-capita rate metric and Alroy's[30] second-for-third metric on the estimation of diversification rates and the impact on rate-age-relationships.

Foote's metrics rely on the age ranges of organisms, i.e. their first and last appearance in the fossil record. Four types of ranges are classified: taxa present both before and after a focal interval, crossing its base and top (N_{bt}); taxa present before and within the interval but not after, crossing only its base (N_{bL}); taxa present in and after the interval, crossing only its top; and taxa found only within the interval (N_{FL}). Summed, these counts are the range-through diversity (N_{bt}). The rates we use here are Foote's per capita instantaneous rates where origination is $\lambda = -ln\left(\frac{N_{bt}}{N_{Ft} + N_{bt}}\right)$, and extinction is $\mu = -ln\left(\frac{N_{bt}}{N_{bL} + N_{bt}}\right)$. This range based metric is still actively used[31], and is perhaps the best approach to study range based data[29]. However, range-based metrics have been argued to suffer from edge, Pull of the Recent, and Signor-Lipps effects, and taxonomic biases. In particular, this means that

extinction rates are increased in intervals preceding mass extinctions (effectively smeared out), rates gradually decline as the recent is approached, and the importance of rare taxa is overweighted when sampling is good (i.e. Pull of the Recent) or underweighted when turnover is fast (i.e. Signor-Lipps)[13,30].

Fossil occurrence data has become increasingly available, promoting a way to move beyond using range-based metrics and focussing on patterns of taxonomic sampling throughout. Alroy's second-for-third metric has been argued to not suffer from any of the above biases, and relies on a sliding-window approach, in contrast to full ranges[30]. Our focal time interval is i_0, and we assume a cohort of taxa that have only been sampled since i_{-1} up to i_2 for extinction and i_{-2} to i_1 for origination. The following counts are of importance for extinction (origination in brackets if distinct): s_1 = taxa sampled in i_{-1} and i_0 (i_0 and i_1) but no later (earlier), s_2 = taxa sampled in i_{-1} and i_1 (i_{-1} and i_1) but not in i_0 or i_2 (i_{-2} and i_0), s_3 = those sampled in i_{-1} and i_2 (i_{-2} and i_1) but not in i_0 or i_1 (i_{-1} and i_0), t_2 = those sampled in i_{-1} and i_0 (i_0 and i_1) regardless of being sampled later (earlier), and p = i_{-1} and i_1 but not in i_0. The formula for extinction and origination proportion (not rate) is then calculated as $O = E = (s_1 - s_x)/(t_2 + p)$. If sampling is uniform and given that unextinction cannot occur the counts should have the expected order $s_1 \geq s_2 \geq s_3$, and s_x would be s_3 (the Gap-Filler equation[13]). However, in practice they do not and might even show the reverse pattern $s_1 < s_2 < s_3$. To address this Alroy suggested algorithmically replacing s_x with the second lowest of the counts, e.g. $s_x = s_1$ if $s_2 > s_1 > s_3$. Conventional rates are then obtained by log-transformation where $\mu = \log\left(\frac{1}{1-E}\right)$, and $\lambda = \log\left(\frac{1}{1-O}\right)$. Second-for-third rates are obviously heuristic because of the algorithmic approach, but it incorporates as much information as possible to correct for sample size.

We focus on instantaneous rates for both Foote's and Alroy's metrics, and do not normalise for interval duration. That is, we focus on the rate at which species originate and go extinct per taxon per interval, and do not normalise our rates to be expressed per million years. Raup and Sepkoski already discussed preferring instantaneous rates in their seminal paper[2]. They observed that the number of raw extinction counts did not correlate with interval duration as they are likely clustered in time[3,29]. Foote further explored the consequences of normalisation and found that if intervals were to experience a single peak in extinction risk, this would dominate the total extinction profile of short intervals more than for long intervals, effectively causing a negative correlation between interval duration and extinction rate[1]. We repeat our analysis for time-normalised rates to understand if negative rate scaling between bins has an effect on average rate comparisons between clades.

We calculate both Foote's per capita rates and second-for-third rates using the R package *divDyn*[23]. Origination and extinction rates were calculated for all orders for both geological stages and equal time bins, and then averaged per order. In some instances, an average rate of zero is estimated. Rate estimates of zero likely are caused by sampling error or model misspecification and make little sense empirically[29] (but see ref[32]). For example, an order must have had a positive origination rate if it contains more than one species. In our focal analysis, orders with average rate estimates of zero are removed. This might bias the analysis towards observing greater rate-scaling. Regardless, average rates are subsequently correlated across all orders with their duration using (i) non-linear least squares, and (ii) linear regression by log-transforming both rate and duration, which would have removed zero estimates.

A simulation model

We investigate if incomplete sampling through time can cause ARS with a simulation model of species diversification and sampling. The simulation model is based on a prior model used in the

testing of rate metrics[14], and can be abstracted to modelling temporal turnover in species identities in a species pool. The species pool has constant species diversity, N, through time ($N = 10^2, 10^3, 10^4$), which can be justified by the seemingly strong correlation between origination and extinction rates observed in the fossil record (i.e. diversity-dependence[3,21]). At the end of each interval, species have a probability, E, to go extinct ($E = 0.15, 0.3, 0.6$). Speciation or origination is modelled as the replacement of the extinct species immediately upon extinction by a new species. Effectively, diversification thus occurs in pulses between intervals. We simulate 10^3 clades per parameter combination to calculate ARS, and these clades experience turnover for a random number of intervals drawn from a uniform distribution ranging from 3 (i.e. the minimum number of bins required to calculate per-capita rates) to 54 (i.e. roughly the reflecting the length of the Phanerozoic if each interval is 10 Myr) intervals. The realised time series of species turnover, each seen as a unique clade, are then degraded based on a per-species sampling probability per interval, s. We model sampling intensity from poor to good ($s = 0.25, 0.5, 0.75$). The complete and degraded time series are then used to calculate both Foote's per-capita and Alroy's second-for-third rates and plotted against the number of time intervals to repeat the empirical analysis (Figures 3,S5, and S6).

Our simulations assume that sampling is uniform through time and across clades. In practice, sampling within time bins will vary because of variation in preservation (e.g. substrate), sampling effort, and fossilisation potential of clades (e.g. soft or hard-bodied). Taxonomic biases might further contribute to error in the estimation of rates by over or under splitting taxa. This might in itself result from incomplete sampling through time, as infrequently sampled but temporally long-ranged taxa might be under split. Equally, taxonomic ranks are inherently arbitrary concepts. Especially when it comes to orders and other higher ranks, it might be difficult to accurately delineate when they arise, and consequently higher diversity orders might be recognised before low diversity orders. Therefore, our simulations are by no means exhaustive in terms of the possible ways that sampling could influence the estimation of diversification rates.

Acknowledgements

We thank Alex Pigot and David Murrell for discussion and encouragement. The Paleobiology Database enabled this work, and contributors are wholeheartedly thanked for their effort. B.R.R. and R.A.C. are funded through the Royal Society by a URF awarded to R.A.C. This is Paleobiology Database publication XXX.

Author contributions

B.R.R. conceptualised the study. R.A.C. curated the data. B.R.R. designed, programmed, and conducted the analysis. B.R.R. wrote the initial draft, which all authors edited and approved.

Declaration of interests

The authors declare no competing interests

References

1. Foote, M. (1994). Temporal variation in extinction risk and temporal scaling of extinction metrics. Paleobiology *20*, 424–444. https://doi.org/10.1017/S0094837300012914.

2. Raup, D.M., and Sepkoski, J.J. (1984). Periodicity of extinctions in the geologic past. Proc. Natl. Acad. Sci. *81*, 801–805. https://doi.org/10.1073/pnas.81.3.801.

3. Alroy, J. (2008). Dynamics of origination and extinction in the marine fossil record. Proc. Natl. Acad. Sci. *105*, 11536–11542. https://doi.org/10.1073/pnas.0802597105.

4. Foote, M. (2005). Pulsed origination and extinction in the marine realm. Paleobiology *31*, 6–20. https://doi.org/10.1666/0094-8373(2005)031<0006:POAEIT>2.0.CO;2.

5. Foote, M. (2007). Extinction and Quiescence in Marine Animal Genera. Paleobiology *33*, 261–272.

6. Henao Diaz, L.F., Harmon, L.J., Sugawara, M.T.C., Miller, E.T., and Pennell, M.W. (2019). Macroevolutionary diversification rates show time dependency. Proc. Natl. Acad. Sci. *116*, 7403–7408. https://doi.org/10.1073/pnas.1818058116.

7. Alroy, J. (1996). Constant extinction, constrained diversification, and uncoordinated stasis in North American Mammals. Palaeogeogr. Palaeoclimatol. Palaeoecol. *127*, 285–311. https://doi.org/10.1016/S0031-0182(96)00100-9.

8. Foote, M. (1999). Morphological Diversity in the Evolutionary Radiation of Paleozoic and Post-Paleozoic Crinoids. Paleobiology *25*, 1–115.

9. Foote, M. (2000). Origination and extinction components of taxonomic diversity: general problems. Paleobiology *26*, 74–102.

10. Harmon, L.J., Pennell, M.W., Henao-Diaz, L.F., Rolland, J., Sipley, B.N., and Uyeda, J.C. (2021). Causes and Consequences of Apparent Timescaling Across All Estimated Evolutionary Rates. Annu. Rev. Ecol. Evol. Syst. *52*, 587–609. https://doi.org/10.1146/annurev-ecolsys-011921-023644.

11. Rolland, J., Henao-Diaz, L.F., Doebeli, M., Germain, R., Harmon, L.J., Knowles, L.L., Liow, L.H., Mank, J.E., Machac, A., Otto, S.P., et al. (2023). Conceptual and empirical bridges between micro- and macroevolution. Nat. Ecol. Evol. *7*, 1181–1193. https://doi.org/10.1038/s41559-023-02116-7.

12. Alroy, J. (2015). A more precise speciation and extinction rate estimator. Paleobiology *41*, 633–639. https://doi.org/10.1017/pab.2015.26.

13. Alroy, J. (2014). Accurate and precise estimates of origination and extinction rates. Paleobiology *40*, 374–397. https://doi.org/10.1666/13036.

14. Alroy, J. (2014). Accurate and precise estimates of origination and extinction rates. Paleobiology *40*, 374–397. https://doi.org/10.1666/13036.

15. Sheets, H.D., and Mitchell, C.E. (2001). Uncorrelated change produces the apparent dependence of evolutionary rate on interval. Paleobiology *27*, 429–445. https://doi.org/10.1666/0094-8373(2001)027<0429:UCPTAD>2.0.CO;2.

16. De Lisle, S., and Svensson, E.I. (2023). Evolutionary change is remarkably time-independent across scales. Preprint, https://doi.org/10.1101/2023.12.02.569704 https://doi.org/10.1101/2023.12.02.569704.

17. Close, R.A., Benson, R.B.J., Saupe, E.E., Clapham, M.E., and Butler, R.J. (2020). The spatial structure of Phanerozoic marine animal diversity. Science *368*, 420–424. https://doi.org/10.1126/science.aay8309.

18. Alroy, J. (2010). The Shifting Balance of Diversity Among Major Marine Animal Groups. Science *329*, 1191–1194. https://doi.org/10.1126/science.1189910.

19. Louca, S., Henao-Diaz, L.F., and Pennell, M. (2022). The scaling of diversification rates with age is likely explained by sampling bias. Evolution *76*, 1625–1637. https://doi.org/10.1111/evo.14515.

20. Foote, M., and Raup, D.M. (1996). Fossil preservation and the stratigraphic ranges of taxa. Paleobiology *22*, 121–140. https://doi.org/10.1017/S0094837300016134.

21. Foote, M. (2023). Diversity-dependent diversification in the history of marine animals. Am. Nat.

22. Foote, M. (2005). Pulsed Origination and Extinction in the Marine Realm. Paleobiology *31*, 6–20.

23. Kocsis, Á.T., Reddin, C.J., Alroy, J., and Kiessling, W. (2019). The r package divDyn for quantifying diversity dynamics using fossil sampling data. Methods Ecol. Evol. *10*, 735–743. https://doi.org/10.1111/2041-210X.13161.

24. Rabosky, D.L. (2013). Diversity-Dependence, Ecological Speciation, and the Role of Competition in Macroevolution. Annu. Rev. Ecol. Evol. Syst. *44*, 481–502. https://doi.org/10.1146/annurev-ecolsys-110512-135800.

25. Close, R.A., Benson, R.B.J., Alroy, J., Carrano, M.T., Cleary, T.J., Dunne, E.M., Mannion, P.D., Uhen, M.D., and Butler, R.J. (2020). The apparent exponential radiation of Phanerozoic land vertebrates is an artefact of spatial sampling biases. Proc. R. Soc. B Biol. Sci. *287*. https://doi.org/10.1098/rspb.2020.0372.

26. Rosenblum, E.B., Sarver, B.A.J., Brown, J.W., Des Roches, S., Hardwick, K.M., Hether, T.D., Eastman, J.M., Pennell, M.W., and Harmon, L.J. (2012). Goldilocks Meets Santa Rosalia: An Ephemeral Speciation Model Explains Patterns of Diversification Across Time Scales. Evol. Biol. *39*, 255–261. https://doi.org/10.1007/s11692-012-9171-x.

27. Rabosky, D.L. (2016). Reproductive isolation and the causes of speciation rate variation in nature. Biol. J. Linn. Soc. *118*, 13–25.

28. Peters, S.E., and McClennen, M. (2016). The Paleobiology Database application programming interface. Paleobiology *42*, 1–7. https://doi.org/10.1017/pab.2015.39.

29. Alroy, J. (2010). Fair Sampling of Taxonomic Richness and Unbiased Estimation of Origination and Extinction Rates. Paleontol. Soc. Pap. *16*, 55–80. https://doi.org/10.1017/s1089332600001819.

30. Alroy, J. (2015). A more precise speciation and extinction rate estimator. Paleobiology *41*, 633–639. https://doi.org/10.1017/pab.2015.26.

31. Foote, M., Edie, S.M., and Jablonski, D. (2024). Ecological structure of diversity-dependent diversification in Phanerozoic marine bivalves. Biol. Lett. *20*, 20230475. https://doi.org/10.1098/rsbl.2023.0475.

32. Foote, M. (2007). Extinction and Quiescence in Marine Animal Genera. Paleobiology *33*, 261–272.