



Drivers of fossil sampling heterogeneity: lessons from planktonic foraminifera

Jérémy Andréoletti¹ , Isabel Fenton^{2,3}, Hélène Morlon¹ and Erin E. Saupe² 

¹Institut de Biologie, École Normale Supérieure, Université PSL, CNRS, INSERM, Paris, France

²Department of Earth Sciences, University of Oxford, Oxford, U.K.

³The Alan Turing Institute, London, U.K.

Rapid Communication

Cite this article: Andréoletti, J., I. Fenton, H. Morlon, and E. E. Saupe (2026). Drivers of fossil sampling heterogeneity: lessons from planktonic foraminifera. *Paleobiology*, 1–9. <https://doi.org/10.1017/pab.2026.10094>

Received: 28 August 2025

Revised: 18 November 2025

Accepted: 23 December 2025

Handling Editor:

Melanie Hopkins

Corresponding author:

Jérémy Andréoletti;

Email: jeremy.andreoletti@gmail.com

Abstract

The fossil record is subject to multiple biases that can distort macroevolutionary and paleoecological inferences. Although temporal and spatial sampling biases have received substantial attention, other sources of fossil sampling heterogeneity remain less well quantified. Using the Triton database of planktonic foraminifera, we assess the influence of geographic, ecological, morphological, and methodological factors on fossil recovery rates. We first apply a temporal subsampling method to standardize fossil occurrences over geologic time, validating this approach against an expert-curated lineage-through-time trajectory. After subsampling, the occurrences remain unevenly distributed throughout species' lifetimes and inhomogeneously distributed across species, reflecting biological signal and/or persistent sampling biases.

We then investigate this residual sampling heterogeneity with a generalized additive model incorporating relevant predictors from Triton. Our results reveal that, after correcting for temporal biases, geographic predictors (paleolatitude, paleolongitude, longitudinal spread) explain nearly a third of sampling variation. Species-specific ecological and morphological attributes contribute an additional fraction, among which mean relative abundance emerges as the main factor. Additional predictors of fossil sampling rates include age-calculation methods and biostratigraphic sampling biases. Despite accounting for multiple sources of variation, 37% of the deviance remains unexplained, suggesting unmodeled biological, stratigraphic, diagenetic, or taxonomic drivers of sampling heterogeneity.

Overall, observed recovery rates question the validity of the homogeneous-sampling assumption used in most diversification models, and this heterogeneity cannot be reduced to a single dominant factor. This conclusion reinforces the need for integrated subsampling approaches and process-based models that explicitly account for heterogeneous fossilization rates to improve the reliability of macroevolutionary analyses.

Non-technical Summary

The fossil record provides a window into the history of life on Earth. However, fossils do not form or become preserved evenly across time, space, or species. These gaps and biases can distort the story we tell about how organisms evolve, spread, and go extinct. Although the effects of time and geography on fossil recovery have been studied in detail, many other sources of variation in fossil recovery remain poorly understood.

In this study, we use a large global database of planktonic foraminifera (microscopic marine organisms with a rich fossil record) to investigate why some species leave more fossils behind than others. To first correct for broad time-related effects, we applied a method that balances the number of fossil samples across different geologic periods. This revealed that, even after accounting for time, fossils remain unevenly distributed: some species are much better represented than others.

We then tested which factors best explain this residual unevenness. We found that where a species lived on the globe matters: latitude, longitude, and the extent over which a species is distributed together explain nearly a third of the differences in fossil recovery rates. Beyond geography, species' characteristics, as well as the methods used to sample and date the fossils, also predict some of the variation. Still, more than a third of the variation remained unexplained, pointing to additional biological, geologic, or human factors that may explain variation in fossil recovery.

These findings show that fossil recovery is not a uniform process. Instead, many overlapping influences shape the species we see in the fossil record. Accounting for this unevenness is crucial if we want reliable insights into the history of life and the processes that drive biodiversity change.

© The Author(s), 2026. Published by Cambridge University Press on behalf of Paleontological Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Introduction

Temporal, geographic, and preservation biases have long been recognized as key challenges in the interpretation of paleontological data (Raup 1976, 1979; Signor and Lipps 1982; Foote and Raup 1996; Smith 2001). These biases can distort observed biodiversity trends, confound

macroevolutionary analyses, and obscure true signals of origination and extinction (Smith 2001; Benson and Upchurch 2013; Vilhena and Smith 2013; Benson et al. 2021; Dunne et al. 2023; Nanglu and Cullen 2023; Antell et al. 2024; Heath et al. 2025). Consequently, several approaches have been introduced to mitigate these biases. Existing methods include rarefaction (Sanders 1968; Miller and Foote 1996; Alroy et al. 2001), coverage-based rarefaction (Chao and Jost 2012; Chiu 2023) or shareholder quorum subsampling (SQS) (Alroy 2010a,b,c), and geographic subsampling (Antell et al. 2020, 2024; Close et al. 2020a,b; Benson et al. 2021). Rarefaction adjusts for uneven sampling intensity through time by standardizing sample sizes across time bins. SQS builds an uneven but “fair” sample that tells you how many species you would find given fixed “coverage” of the underlying taxon abundance distribution. Geographic subsampling attempts to correct for the highly uneven distribution and extent of fossil localities over time by restricting geographic sampling extents.

These subsampling methods are often used to estimate changes in the richness of a clade over time (Alroy et al. 2008; Alroy 2010b; Close et al. 2020b). However, they can also be used as a preprocessing step to de-bias a fossil dataset before inferring macroevolutionary dynamics, such as patterns of origination and extinction using taxa’s first and last appearance dates in the fossil record (*per capita rates*; Alroy 1996; Foote 1999, 2000). In addition to subsampling, other methods have been developed to estimate patterns of origination and extinction from fossils while correcting for temporal biases, without necessarily subsampling the dataset (*three-timer rates*, [Alroy 2008]; *gap-filler rates* [Alroy 2014]; *PyRate* [Silvestro et al. 2014a,b, 2019]; *second-for-third substitution rates* [Alroy 2015; see Kocsis et al. 2019]; *DeepDive* [Cooper et al. 2024]). These methods attempt to correct for temporal biases by explicitly estimating sampling probabilities in each time bin, using taxon counts across multiple consecutive intervals, and incorporating these counts into rate estimations. Rates of origination and extinction can also be estimated from phylogenies with fossils included (using *fossilized birth–death models*; Stadler 2010; Heath et al. 2014; Mulvey et al. 2025). These approaches can model temporal variation in sampling rate (Gavryushkina et al. 2014; Andréoletti et al. 2022). Across paleontological and phylogenetic rate estimation methods, neglecting to account for variable fossil recovery over time has been shown to distort estimates of speciation and extinction rates (Smiley 2018; Warnock et al. 2020).

Even after accounting for temporal variation in fossil preservation, most methods for estimating speciation and extinction rates assume homogeneity in fossil sampling rates across species. This is true for the per capita (Foote 2000), three-timer (Alroy 2008), gap-filler (Alroy 2014), and second-for-third substitution rates (Alroy 2015), as well as most phylogenetic birth–death models (Stadler 2010). A recent exception is the multi-type fossilized birth–death (MTFBD) model, which introduces the possibility of distinct fossil sampling rates across discrete categories or “types” along the phylogeny (Barido-Sottani and Morlon 2026). The MTFBD thus enables heterogeneity in fossilization rate, although with a limited number of rate categories that do not offer resolution at the species level. To our knowledge, PyRate is the only tool that models species-specific preservation rates with a discretized gamma distribution to infer rates of speciation and extinction in a Bayesian framework from fossil occurrence data (Silvestro et al. 2014b). The consequences of assuming an incorrect model of fossil recovery across species have not been studied extensively.

Overall, small variations in fossil sampling rates seem to have limited impacts on macroevolutionary analyses (Silvestro et al. 2014b;

Close et al. 2018; Barido-Sottani and Morlon 2026), but the magnitude of fossilization heterogeneity has rarely been assessed in empirical datasets, and it remains unclear whether more realistic fossil sampling heterogeneity rates would lead to erroneous interpretations of diversification dynamics.

Planktonic foraminifera have one of the richest fossil records (Aze et al. 2011). They are geographically widespread, morphologically diverse, and stratigraphically well studied, making them an excellent group to assess the impact of fossil recovery rates. The Triton database (Fenton et al. 2021) compiles >500,000 planktonic foraminiferal fossil occurrences throughout the Cenozoic, each annotated with taxonomic details, paleogeographic coordinates, age models, sampling methods, and other relevant metadata. Importantly, Triton also includes expert-curated species’ stratigraphic ranges, which allow quantification of morphospecies-specific fossil sampling rates and reconstruction of a lineage-through-time (LTT) trajectory. The possibility of comparing subsampled datasets with this reference richness trajectory is a unique advantage in planktonic foraminifera.

Here, we explore sources of heterogeneity in fossil sampling in planktonic foraminifera, with implications for other taxa, particularly marine microfossils. In large fossil databases, sampling toward the present-day is often higher due to better preservation and the greater extent and area of rock outcrop (Smith 2001; Dunhill et al. 2014; Benson et al. 2021; Antell et al. 2024). This pattern is often referred to as the *Pull of the Recent* (Raup 1979; but see also Jablonski et al. 2003). Because temporal biases are known to be pervasive and have been addressed in previous studies (Alroy 2010c; Kocsis et al. 2019), we first introduce a strategy for temporal subsampling and validate this approach against the expert-based LTT trajectory. We then provide a comprehensive analysis of the remaining potential drivers of fossil sampling heterogeneity, including biological factors such as ecological and morphological traits (Aze et al. 2011), paleogeography (Fenton et al. 2021), and biostratigraphic factors such as sampling depth and sampling reason (Fenton et al. 2021).

Materials and Methods

Dataset

The Triton database (Fenton et al. 2021) compiles 512,922 fossil occurrences of planktonic foraminifera over the Cenozoic, from multiple sources (Pangaea, ocean drilling projects, Neptune, ForCenS, and newly digitized data, see [Supplementary Fig. S2](#) for the spatial distribution of sampling sites). Each record includes taxonomic information (species name, possible synonyms), age data (absolute or relative stratigraphic ages with uncertainty, computed from various biostratigraphic and magnetostratigraphic schemes), geographic and paleogeographic coordinates (drilling hole location, water depth, paleolatitude, paleolongitude), and sampling information (e.g., sediment core, sample depth, collection year, reason for sampling). Triton also provides ecological and morphological information associated with each morphospecies (e.g., relative abundance in its sample, wall texture) and their expert-curated temporal ranges (i.e., the timing of speciation and extinction compiled by taxonomic experts).

The dataset was preprocessed in Triton to exclude outliers outside the known range of the species (cutoffs at 2 Ma for the Neogene and 5 Ma for the Palaeogene, as described in Fenton et al. [2021]), as well as samples with negative ages (post-1950) and present-day ages. To avoid occurrences having the exact same age, which is not allowed under most stochastic sampling models,

such as the birth–death models mentioned earlier (Stadler 2010; Heath et al. 2014; Mulvey et al. 2025), we uniformly sampled a random age in the age uncertainty range of the occurrence, when available. If age uncertainty was not available, we added 10 kyr around age estimates to reflect homogenization in the sedimentation process (i.e., to reflect the fact that fossil assemblages represent an accumulation of mixed material from slightly different ages rather than precise instantaneous events; Patzkowsky and Holland 2012).

Identifying and Correcting Temporal Bias

The density of raw data (i.e., the number of sediment samples and number of occurrences) increased toward the present day (Fig. 1A, blue lines). To account for this temporal sampling bias, we subsampled the data temporally using a threshold-based approach

applied at the sediment sample level. In this context, a sediment sample refers to an extract from a sediment core, which is a cylindrical section of material obtained from a specific drilling hole. Each hole may contain multiple cores, and each core is subdivided into individual sediment samples, representing a finer temporal resolution. Subsampling was performed as follows:

1. Divide the Cenozoic into identical time bins of 0.2 Myr.
2. Count how many sediment samples fall within each time bin.
3. Define a reference threshold T (here, the 25th percentile of observed bin-wise sample counts).
4. For bins with sample counts n exceeding T , randomly retain samples with probability T/n . For bins below T , retain all sediment samples.

This procedure normalizes the distribution of sediment samples across geologic time. The subsampling method was validated by

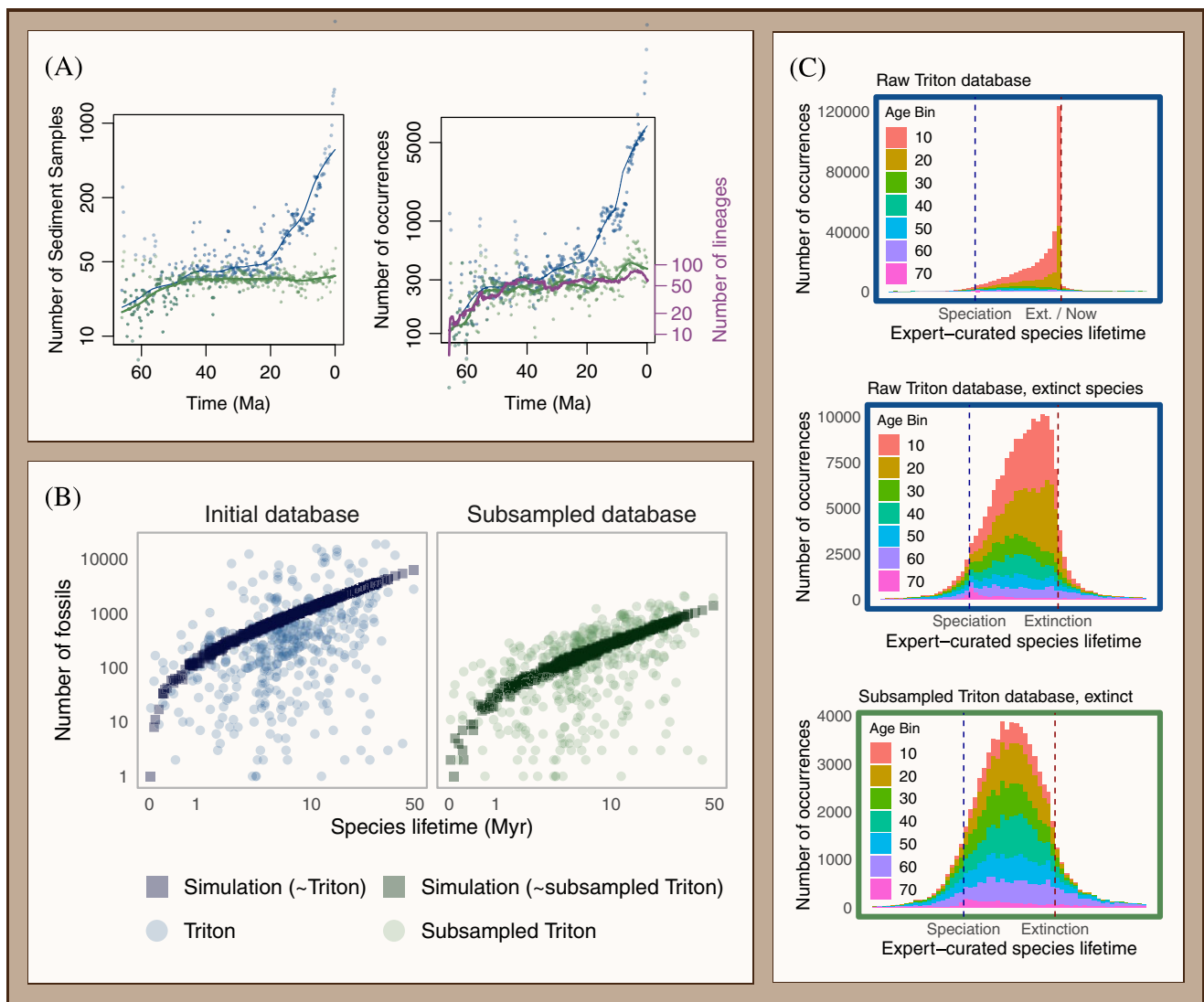


Figure 1. Impact of temporal subsampling on fossil occurrence distributions in the Triton database. (A) Distribution through time of sediment samples and occurrences before (blue) and after (green) temporal subsampling. The scaled profile of the expert-curated lineage-through-time (LTT) trajectory (purple) aligns with the distribution of subsampled occurrences, validating our temporal standardization method. (B) Levels of fossil sampling relative to species' life span. Observed fossil sampling in the Triton database (blue circles) and subsampled Triton database (green circles) shows greater heterogeneity for species of similar age range than expected under a homogeneous Poisson process (darker squares). (C) Distributions of occurrences over species' expert-curated longevity. Top, raw Triton database, illustrating the increase in the number of occurrences toward the present. Middle, Triton database restricted to extinct species, showing an excess of recent occurrences likely due to sampling biases. Bottom, subsampled database restricted to extinct species, showing a symmetric bell-shaped distribution, the tails of which indicate occurrences recorded over and beyond expected species' temporal ranges.

comparing the shape of the distribution of occurrences over time to the reference LTT trajectory (Fig. 1A) obtained as a range-through species count from expert-derived species' temporal ranges in Triton (Wade *et al.* 2011; Fenton *et al.* 2021). Under the expectation that the number of occurrences correlates with the number of species, these two curves should follow similar trajectories.

We compared the distributions of fossil occurrences over species' stratigraphic ranges in three datasets: the raw Triton database, the subset restricted to extinct species, and the temporally subsampled dataset restricted to extinct species. For each case, occurrences were rescaled relative to species' speciation and extinction dates (or the present-day for extant species) to visualize how sampling intensity varies along their lifetimes (Fig. 1C).

Modeling Fossil Sampling Heterogeneity

We examined whether species with similar stratigraphic durations are sampled at similar intensities, as expected under a homogeneous Poisson model, which is often used to represent fossil sampling in models depicting diversification as a branching birth–death process (Stadler 2010). To do so, we simulated an alternative dataset with the same number of occurrences, but with occurrences randomly attributed to a given species in proportion to the longevity of the species (derived from expert-curated speciation and extinction times). We then compared the number of fossil occurrences across species relative to their lifetimes in both the empirical and simulated datasets, before and after temporal subsampling. Any substantial remaining deviation from the Poisson predictions in the subsampled dataset suggests that additional factors may drive heterogeneity in fossil sampling.

Predictor Variables To quantify among-species variation in fossil recovery, we first estimated a sampling rate for each morphospecies based on its observed occurrences. Because species' temporal ranges are based on expert-curated origination and extinction ages that integrate multiple stratigraphic sources, we defined sampling rates as the number of occurrences N divided by the species' total duration. Some formulations use $(N - 1)$ instead of N to represent the number of waiting times between observed fossils (Solow and Smith 1997), but this applies only when durations are derived directly from the occurrences. We modeled variation of fossil recovery rates among species with a gamma distribution, which is positive valued and matches the empirical distribution of rates (Supplementary Fig. S11). We analyzed the heterogeneity in fossil sampling rates using a generalized additive model (GAM), implemented using the MASS (v. 7.3-65) and mgcv (v. 1.9-3) R packages (Venables and Ripley 2002; Wood 2011). We included 10 potential predictors of fossil sampling rate, calculated across all occurrences attributed to a given morphospecies after temporal subsampling. These predictors were selected based on their availability in the Triton database and their potential for driving fossil sampling rates.

For continuous variables (Supplementary Fig. S6), we computed the predictors over the entire stratigraphic range of each species:

- (i) the median paleolatitude of the region where the most occurrences were found (i.e., for each species, we measured the number of occurrences in the Tropics, 23.5°N–23.5°S, Northern Hemisphere, and Southern Hemisphere, and computed paleolatitude only in the region with highest counts to avoid bimodality artefacts for species spanning both hemispheres); and (ii) the circular median paleolongitude (which minimizes the total angular distance to all data points, using the R package circular v. 0.5-1

by Lund *et al.* [2024]), given that some geographic regions may have sustained more abundant populations and/or may have been sampled more extensively (see Supplementary Fig. S2);

- (iii) the variance in absolute paleolatitude and (iv) the circular variance in longitude (see Lund *et al.* 2024), given that widespread species should have higher chances of being sampled;
- (v) the mean relative abundance, given that abundant species can be expected to have richer fossil records; and
- (vi) the mean sample depth, given that species whose fossils are available in sediments near the surface (independently of age) should be easier to collect.

For categorical variables (Supplementary Figs. S3–S5), we considered:

- (vii) the morphogroup (19 types; see Aze *et al.* [2011] and Supplementary Table S1) and (viii) ecogroup (6 types; see Aze *et al.* [2011] and Supplementary Table S2), given that some morphogroups (larger or with more robust tests) or ecogroups (in depths associated with higher plankton density or lower exposure to corrosive bottom waters) may reach sediments in greater quantity;
- (ix) the most common sampling reason (biostratigraphy, community analysis, selected species, or unspecified), given that, for example, biostratigraphic indicator species may have been studied more; and
- (x) the most common age calculation method among the occurrences recorded in Triton (original, zones, magnetostratigraphic markers, interpolation, age model; Fenton *et al.* 2021), given that the choice of age calculation model reflects differences in sampling and dating approaches, notably among drilling campaigns (data from older programs with few markers only allow simple interpolations, whereas recent International Ocean Discovery Program campaigns enable age–depth curve models).

Because morphogroups and ecogroups are only defined for macroperforate foraminifera, we did not include wall texture (microperforate or macroperforate) as a separate predictor, but added a “Microperforate” category to the 19 other morphogroups. To reduce multicollinearity among predictors (see Supplementary Material: Correlations between Predictors, Multicollinearity, and Concurvity Diagnostics, Supplementary Figs. S8–S10, Supplementary Table S3), we merged the morphogroups into seven categories (S-globular-spherical, S-flat-clavate-planispiral, NS-globular, NS-muricate, NS-rotaliform, NS-spinose-planispiral, and Microperforate, with S for Spinose and NS for Non-spinose), and we restricted the ecogroups to four categories (Asymbiotic-SML, Symbiotic-SML, Thermocline and Sub-thermocline, with SML for Surface Mixed Layer; Supplementary Fig. S5).

The predictors were chosen to avoid circular relationships with fossilization intensity. For example, we observed that the oldest year of human sampling among all occurrences of a species in Triton is a predictor of its sampling rate, because species with a widespread and rich fossil record tend to have been discovered earlier. This variable was therefore excluded, as it does not offer any explanatory power on fossilization processes. Predictors such as median paleolatitude and paleolongitude or latitudinal and longitudinal spread, however, do not seem to be distorted by the level of sampling (see Supplementary Material: Robustness of Geographic Predictors to Sampling Effort, Supplementary Fig. S7). The mean relative abundance was kept in the main analysis, despite the risk of circularity as a predictor of sampling rates, because relative species abundance in a sediment sample eliminates at least potential sources of heterogeneity among samples (time, geography, local preservation

conditions). However, species-specific characteristics that increase overall preservation may inflate average relative abundance, so we also performed an alternative analysis without this predictor (Supplementary Material: Alternative GAM without Relative Abundance). We note that using absolute sediment abundance would be even more subject to circularity. If more effort is invested in counting specimens in a given sediment sample, then absolute abundance for the identified species will be inflated compared with other samples, but not their relative abundance, which we use here.

Statistical Model GAMs allow for flexible, nonlinear modeling of these predictors and their interactions in order to assess their relative influence on fossil sampling rates. They use splines, which are piecewise polynomial functions that represent the relationships between predictors and sampling rates. For paleolatitude and paleolongitude, a spherical spline was applied to model these spatial coordinates on a globe (Wood 2006). For other continuous variables, such as relative abundance, sample depth, paleolatitudinal variance, and paleolongitudinal variance, thin plate regression splines were used, as they offer smooth fits (Wood 2003). Diagnostic tools from the *mgcv* package (v. 1.9-3; Wood 2011), including residual analysis (Supplementary Fig. S12) and basis dimension checks, were applied to exclude potential over- or underfits of the model.

To assess potential multicollinearity among predictors, we computed variance inflation factor (VIF) diagnostics (Supplementary Table S3) using the performance R package (v. 0.14.0; Lüdtke et al. 2021). For terms with multiple degrees of freedom (df), the package reports the generalized VIF (GVIF; Fox and Monette 1992) together with $GVIF^{1/(2df)}$, which rescales the inflation back to the 1 – df metric so that thresholds are comparable across all terms (< 2, negligible; 2 – 5, moderate; > 5, problematic). In the context of GAMs, we also assessed *concurvity*, the nonlinear analogue of multicollinearity for smooth terms, using *mgcv*'s implementation (v. 1.9-3; Wood 2011). No variable was removed on the basis of these diagnostics.

To disentangle the influence of various factors on fossil sampling rates, we used hierarchical deviance partitioning, which quantifies the relative contributions of different predictor categories, using the R package *gam.hp* (v. 0.0-3; Lai et al. 2024).

Results

Temporal Standardization

Temporal standardization involved systematic sediment subsampling, which effectively corrected for temporal biases. Before subsampling, the occurrence distribution and the expert-derived LTT curve diverge markedly, but after subsampling their shapes become closely aligned (Fig. 1A, Pearson's correlation between the smoothed occurrence distribution and the number of lineages increases from $r = 0.5$ to $r = 0.88$).

The number of fossil occurrences per species is poorly predicted by life span, contrary to expectations under a homogeneous Poisson sampling model (represented by the simulated data in Fig. 1B). Temporal subsampling reduces some heterogeneity, but notable residual variation persists (up to three orders of magnitude between species of similar life spans).

Finally, the distribution of occurrences across species' expert-curated longevities in the raw Triton database shows a strong temporal bias, with occurrences skewed toward the present day

due to increased sampling and the absence of a declining phase for extant species (Fig. 1C). When the database is restricted to extinct species, the distribution is more symmetrical, although with a surplus of recent occurrences due to the increased ability to sample more recent rocks. After temporal subsampling, the distribution of occurrences appears Gaussian with respect to the species' life spans. Some occurrences in Triton extend over and beyond the expected species' temporal ranges, reflecting reworked records, errors in dating of individual occurrences, or overly narrow species' range assignments by experts, potentially due to the geographic incompleteness of the foraminiferal fossil record.

Geographic Heterogeneity

The GAM results (Fig. 2) indicate that no single factor overwhelmingly explains residual variation in fossil sampling rates after temporal subsampling. Among the predictors, paleogeography (paleolatitude and paleolongitude) accounts for the largest proportion of explained deviance at 16.8%. The age calculation method is also a substantial correlate, explaining 14.1% of the deviance. The species' longitudinal spread contributes 11.2%. By comparison, the latitudinal spread explains only 2.5% and is not statistically significant.

The depth of the sample explains 3.3% of the deviance. More specifically, species whose fossils are only available at greater depths in the sediment are collected at lower rates (Fig. 2A). The depth of a sample is strongly correlated with its stratigraphic age from 10 million years ago to the present day, and uncorrelated in older samples (Supplementary Fig. S9), although age is not included as a predictor (our temporal subsampling protocol was designed to remove age-related bias in sampling effort).

Relative abundance accounts for 8.8% of the deviance. Morphogroup and ecogroup variables contribute 3.6% and 1.4%, respectively, but only the microperforate group is statistically significant. The different reasons for which species are sampled only account for a small portion of the deviance (1.2%), although the oversampling of biostratigraphic indicator species is statistically significant. The residual deviance remains high, at 37.1%, suggesting there are other unmodeled processes that influence fossil occurrence patterns in planktonic foraminifera. If relative abundance is excluded from the potential drivers, similar results are observed for the other predictors (Supplementary Fig. S13).

Discussion

We used the Triton database of planktonic foraminifera as a case study to examine the sources of variation in sampling rates among species. Our results highlight the importance of considering multiple sources of bias and potential true biological variation when analyzing the fossil record. By assessing biological, geographic, stratigraphic, and other predictors of sampling heterogeneity after applying temporal subsampling, we clarify which of those factors may be the main sources of heterogeneity in fossil preservation and recovery.

Sediment Subsampling Successfully Mitigates the Pull of the Recent *sensu lato*

To account for higher sampling intensities toward the present day, we used a threshold-based subsampling procedure at the sediment-sample level to uniformly reduce sampling effort in geologically younger time intervals. This approach differs from some

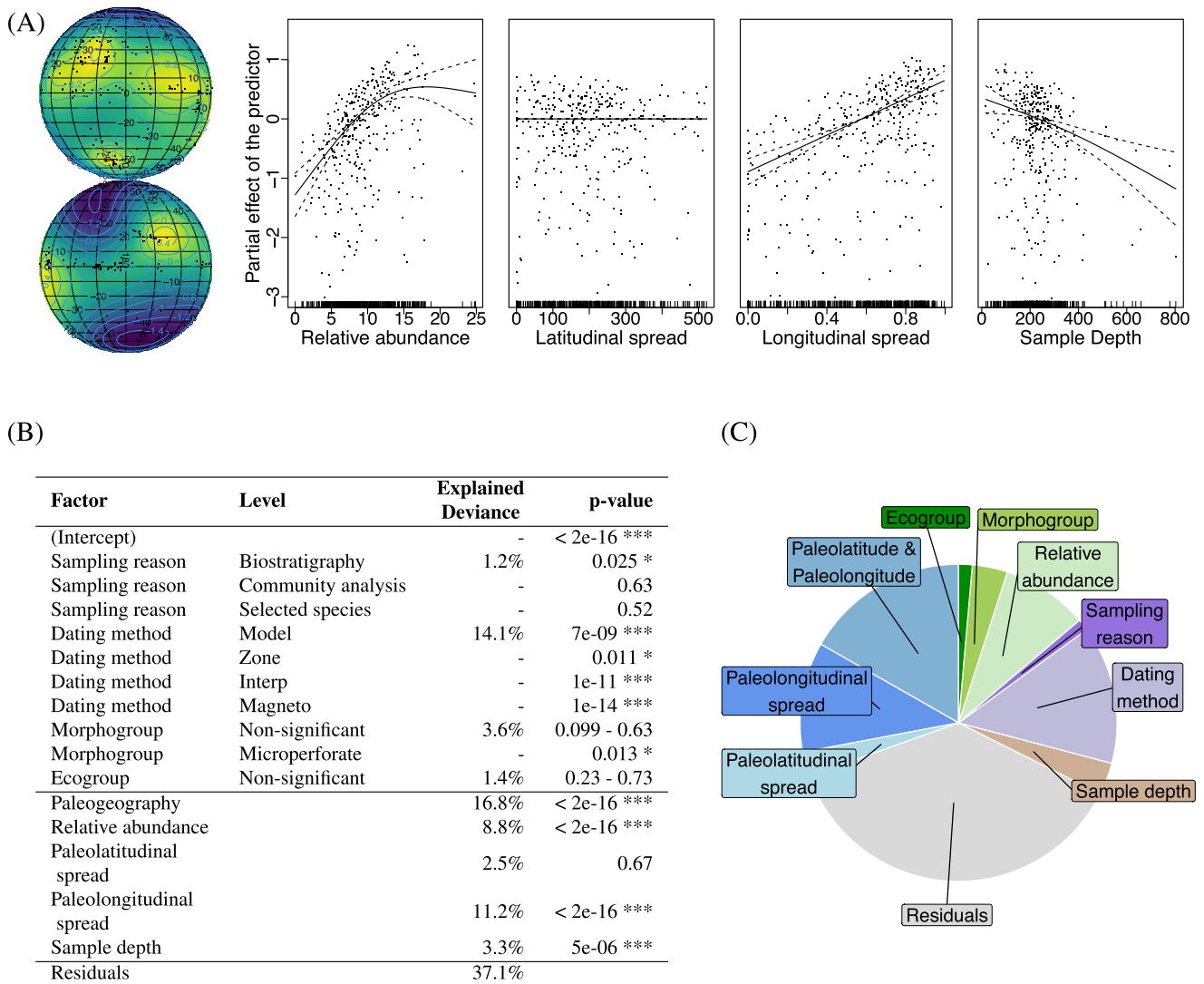


Figure 2. Heterogeneity in fossil sampling rates for planktonic foraminifera explained by geographic, ecological, stratigraphic, and other factors. (A) Spherical maps (left) show predicted sampling rates across paleolatitudes and paleolongitudes, with lighter colors representing higher sampling rates. The top sphere is centered on longitude 0, the bottom on longitude 180. The adjacent scatter plots and curves visualize partial effects of each predictor—relative abundance, paleolatitudinal spread, paleolongitudinal spread, and sample depth—on fossil sampling rates. The dashed lines represent ± 2 SE. (B) A simplified table of the proportions of explained deviance and p -values for the predictors in the generalized additive model (GAM). For factors with multiple categories, the deviance is printed only on the first row; subsequent levels show a dash. For morphogroups and ecogroups, only a single row is shown for nonsignificant groups, with the corresponding range of p -values. Significance thresholds: * $p < 0.05$; *** $p < 0.001$. (C) A pie chart displaying the proportion of deviance explained by each covariate.

conventional methods in that it homogenizes the number of sediment samples per time bin, rather than the number of occurrences per time bin. This is beneficial, because variations in fossil abundance over time are relevant as a proxy of a clade's species richness. Ideally, however, subsampling should be applied to all sediment samples from the selected sampling campaigns, including samples without identified species from the focal clade, to avoid missing true periods of low diversity. We did not implement this, because these true absence samples are not systematically recorded in Triton (99% of sediment samples contain at least one species with nonzero abundance). Assuming comparable fossil preservation probabilities and taxonomic identification effort across sediment samples, our threshold-based subsampling procedure should detect genuine fluctuations in species' abundance. Coverage-based subsampling methods, such as SQS, have been developed specifically to have this property (Alroy 2010b,c). Applying SQS

to the Triton dataset yielded a limited match to the reference LTT trajectory (Supplementary Fig. S1), which may reflect violations of core assumptions of SQS, particularly its expectation of consistent abundance distribution among assemblages. However, contrary to our approach, coverage-based methods require reliable taxonomic information (identification of specimens typically to species or genus), as they aim to standardize the proportion of the underlying frequency distribution represented by the sampled set of taxa (Close *et al.* 2018).

Our subsampled distributions align closely with the expert-derived LTT trajectory (Fig. 1A), which suggests our threshold-based sample-level standardization provides an effective way to mitigate temporal sampling biases, including the Pull of the Recent *sensu lato*, in microfossil data. This finding also supports the assumption that occurrences through time can serve as a proxy for diversity through time, at least for foraminifera, as presumed in

birth–death models that incorporate fossil occurrences (Andréoletti et al. 2022). However, despite overall alignment between occurrences through time and morphospecies diversity through time, the relationship is not one to one. A clade’s total abundance may not always scale directly with its species richness, for instance, if equilibrium dynamics constrain the total number of individuals in the community (Okie and Storch 2025), or, if at one point in time, many low-abundance specialist species are replaced by few abundant generalist species. Finally, none of these temporal subsampling methods aim to alleviate fossilization and sampling biases between species living at the same time. More generally, no subsampling method can correct for severely undersampled species.

Fossil Sampling Intensity Varies along Species’ Lifetimes

The higher number of fossils that we observe in the middle of a species’ lifetime (Fig. 1C) has already been documented in planktonic microfossils (Liow and Stenseth 2007; Liow et al. 2010) and was interpreted by those authors as a rise to dominance (high abundance and broad geographic range), followed by gradual decline. Liow et al. (2010) compared the steepness of rises and falls and found the distribution to be asymmetric; however, they did not correct for fossil sampling, and we find that the distribution asymmetry essentially disappears after temporal subsampling (Fig. 1C). Very few macroevolutionary methods explicitly account for changing fossil sampling rates over a species’ existence. A notable exception is PyRate (Silvestro et al. 2014a,b, 2019), a software that enables preservation rates to vary non-homogeneously throughout species’ life spans, capturing potential early rises and late declines in abundance, as well as rate heterogeneity through geologic time and across lineages. Ignoring these changes can lead to over- or underestimates of speciation or extinction rate trajectories, although PyRate’s BDMCMC model is generally quite robust to misspecifications (Silvestro et al. 2014b). Our findings support the importance of further exploring macroevolutionary models that relax the assumption of uniform fossil recovery throughout a lineage’s history.

Multiple Drivers of Sampling Heterogeneity

Even after temporal subsampling, fossil sampling (as shown in Fig. 1B) remains more variable than expected under a homogeneous Poisson sampling process, which is often used to represent fossil sampling in models depicting diversification as a branching birth–death process (Stadler 2010). Our results show that, for planktonic foraminifera, geographic factors, including paleolatitude/paleolongitude, longitudinal spread, and to a lesser extent latitudinal spread, account for almost a third of the deviance in the sampling rate across species (Fig. 2). These results support previous studies that highlight the extent of spatial sampling biases (see Supplementary Fig. S2) and how unequal sampling efforts can overshadow ecological signals (Smith 2001; Vilhena and Smith 2013; Close et al. 2020b; Allen et al. 2020; Jones et al. 2021; Benson et al. 2021; Dunne et al. 2023; Heath et al. 2025). They further emphasize the importance of using spatial subsampling methods that have been developed in recent years (Antell et al. 2024). Although difficult to disentangle, part of the geographic signal may reflect true ecological patterns, rather than sampling bias, such as higher abundance and therefore sampling rates across species due to favorable environmental conditions, such as upwelling zones.

Interestingly, latitudinal spread was not significant once other location-based factors were included (this is likely not due

to multicollinearity; see Supplementary Material: Correlations between Predictors, Multicollinearity, and Concurrency Diagnostics, Supplementary Figs. S8–S10, Supplementary Table S3), suggesting that broader north–south distributions by themselves do not consistently yield richer fossil records. However, broader longitudinal ranges, reflecting the presence of dominant species in multiple basins or ocean current systems, appear to lead to higher sampling rates (for circularity concerns, see Supplementary Material: Robustness of Geographic Predictors to Sampling Effort). A wider longitudinal distribution would indeed place species in a greater variety of depositional settings, increasing their likelihood of encountering favorable conditions for preservation and subsequent discovery.

We identified three ecological and morphological variables—*morphogroup*, *ecogroup* (including depth-habitat preferences), and *relative abundance*—that collectively explained around a sixth of the deviance in sampling rate. Of these biological predictors, relative abundance explained the highest proportion of the variance, implying that planktonic foraminifera species that are rarer in sediment samples, with all external factors equal, predictably also show reduced detectability globally. However, there is some risk that this predictor is partly confounded with the sampling rate (removing relative abundance from the model yielded similar results for the other predictors; see Supplementary Material: Alternative GAM without Relative Abundance). The fact that no individual macroperforate morphotype or ecotype was identified as a significant predictor of the sampling rate is encouraging for studies using these categories, as it suggests that macroevolutionary patterns using these categories may still yield meaningful results.

Among the other covariates, the age calculation method explains most of the deviance (14%). This variable likely reflects the influence of other factors correlated with how and why these samples were collected (e.g., certain sampling campaigns or drilling practices) in a database assembled from multiple drilling programs (see the justification for inclusion in the “Materials and Methods”). The fact that species used as biostratigraphic indicators are significantly more sampled is straightforward to understand, but this bias should not cause major distortions given that, overall, the reason for sampling contributed only a small share of the deviance. Sampling depth is partly correlated with age (which is not included as a predictor), but variation in sampling depth for species of similar age can arise due to differences in sedimentation rates and post-depositional compaction, which would cause some species’ fossils to be buried deeper than others depending on local depositional environments.

Altogether, the GAM left 37% of deviance unexplained, implying that additional processes—unaccounted-for biological variables, taphonomic factors, local sedimentary processes (Patzkowsky and Holland 2012; Holland et al. 2025), diagenetic histories (Allison and Bottjer 2011), and taxonomic identification biases (Lloyd et al. 2012)—remain unmodeled, or that some drivers are imperfectly modeled—for example, relative abundance is an imperfect proxy of absolute biological abundance. The unexplained deviance may also point to errors in some of the expert-derived species’ temporal ranges in the Triton database.

Because our analyses focused on marine microfossils, which differ markedly from terrestrial macro-organisms in both preservation environments and sampling regimes, the specific predictors identified here may not apply beyond similar planktonic groups. We nevertheless expect several of the potential drivers, particularly geographic and stratigraphic correlates, to generalize to other microfossil clades. More broadly, the main conclusion

that multiple processes underlie fossil sampling heterogeneity is likely to hold across taxa and should at least warrant further investigation. Finally, the analyses themselves are not easily transferable to datasets with sparse sampling, given their reliance on semi-reliable estimates of species' temporal ranges.

Conclusion

This study identifies key predictors of heterogeneity in fossil recovery for planktonic foraminifera. By leveraging the exceptional fossil record of this group, we confirm that temporal subsampling helps correct for the higher diversity toward the present-day (Pull of the Recent *sensu lato*), but still leaves other significant sources of variation in sampling rates that cannot be reduced to a single dominant driver, including potential biological, taphonomic, or other sampling factors.

In the future, macroevolutionary analyses using the fossil record should consistently apply available subsampling methods, most notably temporal and spatial subsampling tools. Furthermore, they should ensure (using simulations or sensitivity analyses) that the patterns they derive are not confounded by the other uncorrected sources of sampling heterogeneity discussed here, as new tools to correct them are still to be developed. The use and development of process-based models that permit fossil sampling rates to vary based on specific variables, or to freely vary across species, would allow these biases to be corrected without the cost of removing fossils from the dataset. Ultimately, understanding the processes driving fossil preservation and recovery is needed to improve our ability to reconstruct unbiased biodiversity patterns over deep time.

Acknowledgments. The authors thank the members of the Morlon team for their feedback on an earlier version of this paper, as well as two anonymous reviewers for their valuable suggestions. E.E.S. acknowledges funding from the Leverhulme Prize and NERC grant NE/V011405/1.

Author Contribution. J.A., I.S.F., and E.E.S. designed the study. J.A. performed the analyses and wrote the manuscript. E.E.S., I.S.F., and H.M. reviewed the manuscript and provided feedback. All authors approved the final version.

Competing Interests. The authors declare no competing interests.

Data Availability Statement. The code for reproducing the analyses is available at <https://github.com/Jeremy-Andreoletti/Foraminifera>. The Supplementary Material and an archived version of the code are available at <https://doi.org/10.17605/OSF.IO/BDU6E>.

Literature Cited

- Allen, B. J., P. B. Wignall, D. J. Hill, E. E. Saupe, and A. M. Dunhill. 2020. The latitudinal diversity gradient of tetrapods across the Permo-Triassic mass extinction and recovery interval. *Proceedings of the Royal Society B* **287**: 20201125.
- Allison, P. A., and D. J. Bottjer. 2011. Taphonomy: bias and process through time. Pp. 1–17 in P. A. Allison and D. J. Bottjer, eds. *Taphonomy: process and bias through time*. Springer, Dordrecht, Netherlands.
- Alroy, J. 1996. Constant extinction, constrained diversification, and uncoordinated stasis in North American mammals. *Palaeogeography, Palaeoclimatology, Palaeoecology* **127**:285–311.
- Alroy, J. 2008. Dynamics of origination and extinction in the marine fossil record. *Proceedings of the National Academy of Sciences USA* **105**(Suppl. 1): 11536–11542.
- Alroy, J. 2010a. Fair sampling of taxonomic richness and unbiased estimation of origination and extinction rates. *Paleontological Society Papers* **16**:55–80.
- Alroy, J. 2010b. Geographical, environmental and intrinsic biotic controls on Phanerozoic marine diversification. *Palaeontology* **53**:1211–1235.
- Alroy, J. 2010c. The shifting balance of diversity among major marine animal groups. *Science* **329**:1191–1194.
- Alroy, J. 2014. Accurate and precise estimates of origination and extinction rates. *Paleobiology* **40**:374–397.
- Alroy, J. 2015. A more precise speciation and extinction rate estimator. *Paleobiology* **41**:633–639.
- Alroy, J., C. R. Marshall, R. K. Bambach, K. Bezusko, M. Foote, F. T. Fürsich, T. A. Hansen, et al. 2001. Effects of sampling standardization on estimates of Phanerozoic marine diversification. *Proceedings of the National Academy of Sciences USA* **98**:6261–6266.
- Alroy, J., M. Aberhan, D. J. Bottjer, M. Foote, F. T. Fürsich, P. J. Harries, A. J. W. Hendy, et al. 2008. Phanerozoic trends in the global diversity of marine invertebrates. *Science* **321**:97–100.
- Andreoletti, J., A. Zwaans, R. C. M. Warnock, G. Aguirre-Fernández, J. Barido-Sottani, A. Gupta, T. Stadler, and M. Manceau. 2022. The occurrence birth–death process for combined-evidence analysis in macroevolution and epidemiology. *Systematic Biology* **71**:1440–1452.
- Antell, G. T., W. Kiessling, M. Aberhan, and E. E. Saupe. 2020. Marine biodiversity and geographic distributions are independent on large scales. *Current Biology* **30**:115–121.e5.
- Antell, G. T., R. B. J. Benson, and E. E. Saupe. 2024. Spatial standardization of taxon occurrence data—a call to action. *Paleobiology* **50**:177–193.
- Aze, T., T. H. G. Ezard, A. Purvis, H. K. Coxall, D. R. M. Stewart, B. S. Wade, and P. N. Pearson. 2011. A phylogeny of Cenozoic macroperforate planktonic foraminifera from fossil data. *Biological Reviews* **86**:900–927.
- Barido-Sottani, J., and H. Morlon. 2026. Integrating fossils samples with heterogeneous diversification rates: a combined Multi-Type Fossilized Birth-Death model. *Systematic Biology*, syag011, <https://doi.org/10.1093/sysbio/syag011>
- Benson, R. B., and P. Upchurch. 2013. Diversity trends in the establishment of terrestrial vertebrate ecosystems: interactions between spatial and temporal sampling biases. *Geology* **41**:43–46.
- Benson, R. B. J., R. Butler, R. A. Close, E. Saupe, and D. L. Rabosky. 2021. Biodiversity across space and time in the fossil record. *Current Biology* **31**: R1225–R1236.
- Chao, A., and L. Jost. 2012. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* **93**:2533–2547.
- Chiu, C.-H. 2023. Sample coverage estimation, rarefaction, and extrapolation based on sample-based abundance data. *Ecology* **104**:e4099.
- Close, R. A., S. W. Evers, J. Alroy, and R. J. Butler. 2018. How should we estimate diversity in the fossil record? Testing richness estimators using sampling-standardised discovery curves. *Methods in Ecology and Evolution* **9**:1386–1400.
- Close, R. A., R. B. J. Benson, J. Alroy, M. T. Carrano, T. J. Cleary, E. M. Dunne, P. D. Mannion, M. D. Uhen, and R. J. Butler. 2020a. The apparent exponential radiation of Phanerozoic land vertebrates is an artefact of spatial sampling biases. *Proceedings of the Royal Society B* **287**:20200372.
- Close, R. A., R. B. J. Benson, E. E. Saupe, M. E. Clapham, and R. J. Butler. 2020b. The spatial structure of Phanerozoic marine animal diversity. *Science* **368**:420–424.
- Cooper, R. B., J. T. Flannery-Sutherland, and D. Silvestro. 2024. DeepDive: estimating global biodiversity patterns through time using deep learning. *Nature Communications* **15**:4199.
- Dunhill, A. M., B. Hannisdal, and M. J. Benton. 2014. Disentangling rock record bias and common-cause from redundancy in the British fossil record. *Nature Communications* **5**:4818.
- Dunne, E. M., S. E. D. Thompson, R. J. Butler, J. Rosindell, and R. A. Close. 2023. Mechanistic neutral models show that sampling biases drive the apparent explosion of early tetrapod diversity. *Nature Ecology and Evolution* **7**:1480–1489.
- Fenton, I. S., A. Woodhouse, T. Aze, D. Lazarus, J. Renaudie, A. M. Dunhill, J. R. Young, and E. E. Saupe. 2021. Triton, a new species-level database of Cenozoic planktonic foraminiferal occurrences. *Scientific Data* **8**(1):160.
- Foote, M. 1999. Morphological diversity in the evolutionary radiation of Paleozoic and post-Paleozoic crinoids. *Paleobiology* **25**(Suppl. to No. 2):1–115.

- Foote, M. 2000. Origination and extinction components of taxonomic diversity: general problems. *Paleobiology* 26(Suppl. to No. 4):74–102.
- Foote, M., and D. M. Raup. 1996. Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology* 22:121–140.
- Fox, J., and G. Monette. 1992. Generalized collinearity diagnostics. *Journal of the American Statistical Association* 87(417):178–183.
- Gavryushkina, A., D. Welch, T. Stadler, and A. J. Drummond. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Computational Biology* 10:e1003919.
- Heath, J. A., N. Cooper, P. Upchurch, and P. D. Mannion. 2025. Accounting for sampling heterogeneity suggests a low paleolatitude origin for dinosaurs. *Current Biology* 35:941–953.e5.
- Heath, T. A., J. P. Huelsenbeck, and T. Stadler. 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences USA* 111:E2957–E2966.
- Holland, S. M., M. E. Patzkowsky, and K. M. Loughney. 2025. Stratigraphic paleobiology. *Paleobiology* 51:44–61.
- Jablonski, D., K. Roy, J. W. Valentine, R. M. Price, and P. S. Anderson. 2003. The Impact of the Pull of the Recent on the History of Marine Diversity. *Science* 300:1133–1135.
- Jones, L. A., C. D. Dean, P. D. Mannion, A. Farnsworth, and P. A. Allison. 2021. Spatial sampling heterogeneity limits the detectability of deep time latitudinal biodiversity gradients. *Proceedings of the Royal Society B* 288: 20202762.
- Kocsis, A. T., C. J. Reddin, J. Alroy, and W. Kiessling. 2019. The R package divDyn for quantifying diversity dynamics using fossil sampling data. *Methods in Ecology and Evolution* 10:735–743.
- Lai, J., J. Tang, T. Li, A. Zhang, and L. Mao. 2024. Evaluating the relative importance of predictors in Generalized Additive Models using the *gam.hp* R package. *Plant Diversity* 46:542–546.
- Liow, L. H., and N. C. Stenseth. 2007. The rise and fall of species: implications for macroevolutionary and macroecological studies. *Proceedings of the Royal Society B* 274:2745–2752.
- Liow, L. H., H. J. Skaug, T. Ergon, and T. Schweder. 2010. Global occurrence trajectories of microfossils: environmental volatility and the rise and fall of individual species. *Paleobiology* 36:224–252.
- Lloyd, G. T., J. R. Young, and A. B. Smith. 2012. Taxonomic structure of the fossil record is shaped by sampling bias. *Systematic Biology* 61:80.
- Lund, U., C. Agostinelli, H. Arai, A. Gagliardi, E. García-Portugués, D. Giunchi, J.-O. Irisson, M. Pocernich, and F. Rotolo. 2024. circular: circular statistics, R package version 0.5-2. <https://CRAN.R-project.org/package=circular>, accessed 8 November 2025.
- Lüdecke, D., M. S. Ben-Shachar, I. Patil, P. Waggoner, and D. Makowski. 2021. performance: an R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software* 6:3139.
- Miller, A. I., and M. Foote. 1996. Calibrating the Ordovician Radiation of marine life: implications for Phanerozoic diversity trends. *Paleobiology* 22:304–309.
- Mulvey, L. P. A., M. C. Nikolic, B. J. Allen, T. A. Heath, and R. C. M. Warnock. 2025. From fossils to phylogenies: exploring the integration of paleontological data into Bayesian phylogenetic inference. *Paleobiology* 51:214–236.
- Nanglu, K., and T. M. Cullen. 2023. Across space and time: a review of sampling, preservational, analytical, and anthropogenic biases in fossil data across macroecological scales. *Earth-Science Reviews* 244:104537.
- Okie, J. G., and D. Storch. 2025. The equilibrium theory of biodiversity dynamics: a general framework for scaling species richness and community abundance along environmental gradients. *American Naturalist* 205: 20–40.
- Patzkowsky, M. E., and S. M. Holland. 2012. *Stratigraphic paleobiology: understanding the distribution of fossil taxa in time and space*. University of Chicago Press, Chicago.
- Raup, D. M. 1976. Species diversity in the Phanerozoic: an interpretation. *Paleobiology* 2:289–297.
- Raup, D. M. 1979. Biases in the fossil record of species and genera. *Bulletin of Carnegie Museum of Natural History* 13:85–91.
- Sanders, H. L. 1968. Marine benthic diversity: a comparative study. *American Naturalist* 102:243–282.
- Signor, P. W., III, and J. H. Lipps. 1982. Sampling bias, gradual extinction patterns and catastrophes in the fossil record. In L. T. Silver and P. H. Schultz, eds. Geological implications of impacts of large asteroids and comets on the Earth. Geological Society of America Special Paper 190:291–296.
- Silvestro, D., N. Salamin, and J. Schnitzler. 2014a. PyRate: a new program to estimate speciation and extinction rates from incomplete fossil data. *Methods in Ecology and Evolution* 5:1126–1131.
- Silvestro, D., J. Schnitzler, L. H. Liow, A. Antonelli, and N. Salamin. 2014b. Bayesian estimation of speciation and extinction from incomplete fossil occurrence data. *Systematic Biology* 63:349–367.
- Silvestro, D., N. Salamin, A. Antonelli, and X. Meyer. 2019. Improved estimation of macroevolutionary rates from fossil data using a Bayesian framework. *Paleobiology* 45:546–570.
- Smiley, T. M. 2018. Detecting diversification rates in relation to preservation and tectonic history from simulated fossil records. *Paleobiology* 44:1–24.
- Smith, A. B. 2001. Large-scale heterogeneity of the fossil record: implications for Phanerozoic biodiversity studies. *Philosophical Transactions of the Royal Society B* 356:351–367.
- Solow, A. R., and W. Smith. 1997. On fossil preservation and the stratigraphic ranges of taxa. *Paleobiology* 23:271–277.
- Stadler, T. 2010. Sampling-through-time in birth–death trees. *Journal of Theoretical Biology* 267:396–404.
- Venables, W. N., and B. D. Ripley. 2002. *Modern applied statistics with S*, 4th ed. Statistics and Computing. Springer, New York.
- Vilhena, D. A., and A. B. Smith. 2013. Spatial bias in the marine fossil record. *PLoS ONE* 8:e74470.
- Wade, B. S., P. N. Pearson, W. A. Berggren, and H. Pälike. 2011. Review and revision of Cenozoic tropical planktonic foraminiferal biostratigraphy and calibration to the geomagnetic polarity and astronomical time scale. *Earth-Science Reviews* 104:111–142.
- Warnock, R. C. M., T. A. Heath, and T. Stadler. 2020. Assessing the impact of incomplete species sampling on estimates of speciation and extinction rates. *Paleobiology* 46:137–157.
- Wood, S. N. 2003. Thin plate regression splines. *Journal of the Royal Statistical Society B* 65:95–114.
- Wood, S. N. 2006. Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics* 62:1025–1036.
- Wood, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society B* 73:3–36.