“A history of the world imperfectly kept”: Will we ever know how biodiversity has changed over deep time?

Bethany J. Allen[1,2,3,*], Rachel C. M. Warnock[4], and Alexander M. Dunhill[5]

1 Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland
2 Computational Evolution Group, Swiss Institute of Bioinformatics, Lausanne, Switzerland
3 GFZ Helmholtz Centre for Geosciences, Potsdam, Germany
4 Geozentrum Nordbayern, Friedrich-Alexander University Erlangen-Nuremberg, Germany
5 School of Earth and Environment, University of Leeds, Leeds, UK
* Corresponding author email: bethany.janet.allen@gfz.de

Abstract

The fossil record is our only direct source of evidence for how life on Earth has waxed and waned over its long history. However, the fossil record is also incomplete and biased in many ways, after passing through biological, geological, and socio-economic filters. This means that we only possess snapshots of information, relating to specific places and times in Earth history, from which to try and understand large-scale biodiversity patterns. Over the last fifty years, a wide variety of methods have been developed to try and elucidate macroevolutionary patterns by accounting for fossil record structure or bias, with varying levels of success. Here we review the different approaches that have previously been applied to this problem, and discuss their strengths and weaknesses. We illustrate this by applying a selection of these methods to the global brachiopod fossil record of the Permian and Triassic. Finally, we highlight some avenues for future improvement, including (1) using simulations to investigate method efficacy, (2) designing studies around testable hypotheses, (3) embracing uncertainty, and (4) improving the integration of data from fossil and modern organisms. Although we cannot know exactly how biodiversity has changed over life’s history, it is clear that new innovations in computational palaeontology are helping us to improve the trustworthiness of our estimates of biodiversity through deep time.

The fossil record and its biases

Fossils, defined as biological presence, activity or signatures preserved in the rock record, are our sole direct source of evidence for how life existed in the geological past, with the fossil record comprising our knowledge of fossil data collectively. We know that these data are incomplete - the vast majority of organisms ever to have lived did not enter the fossil record (e.g. Marshall 2017; Žliobaitė & Fortelius 2022). Charles Darwin, when discussing the fossil record in *On the Origin of Species by Means of Natural Selection*, wrote “I look at the natural geological record as a history of the world imperfectly kept” (1859).

Among the many processes serving as filters to fossilisation, many are selective or systematic, meaning that the resulting fossil record is biased in favour of some organisms above others (Shaw et al. 2020). To what extent the nature of the fossil record is perceived as “structured” or “biased” depends on how well we understand these filtering processes, and how systematic we consider them to be (Smith 2007; Holland 2017). Regardless, it is clear that preservation in the fossil record is uneven, and this manifests in almost any possible analysis we might want to conduct using it (Kidwell & Holland 2002). For example, fossilisation potential is highly variable between palaeoenvironments, with some broadly unrepresented (e.g. Shaw et al. 2020). The soft parts of animals tend not to preserve (Shaw et al. 2020), except in Lagerstätten, meaning our perception of largely soft-bodied and/or fragile animals is highly skewed towards windows of exceptional preservation (Dean et al. 2016; Walker et al. 2020). Bias in the distribution of fossil-bearing rocks across time and space limits our potential to find fossils (Raup 1972; Smith & McGowan 2007; Wall et al. 2009). A wide range of socio-economic factors also determine where fossils are found, kept and studied, leading to major inequalities in data collection worldwide (Raja et al. 2022; Dunne et al. 2025).

As such, it is difficult to quantify diversity in the fossil record, especially to make ‘fair’ comparisons across space and time, or between clades (e.g. Dunne 2018; Dillon et al. 2023; Hammer & Harper 2024). A wide range of methods have been developed and applied over the last 50 years which attempt to do this. With the emergence of conservation paleobiology (Dillon et al. 2022; Pimiento and Antonelli 2022; Kiessling et al. 2023), and the increasing application of the fossil record to make informed predictions about current and future extinction risk (Harnik et al. 2012; Raja et al. 2021; Finnegan et al. 2024), there is particularly strong motivation to scrutinise our methods for estimating extinction rates in the past. Here, we briefly recap the main schools of thought behind methods for estimating palaeodiversity, demonstrate how some of these methods can be applied, and suggest future directions for our efforts to quantify biodiversity from the fossil record.

Inferring biodiversity in deep time

In this paper, we use the terms “diversity” and “richness” to describe the number of discrete taxonomic units present within the biological system of interest (usually globally). Approaches to estimating palaeodiversity aim to take into account the fact that we know our fossil sampling to be incomplete (e.g. Hunt & Slater 2016). They can be placed into two categories, distinguished by the key metrics(s) they aim to infer: (1) diversity estimators, which seek to infer true or relative diversity through time, and (2) rate estimators, which seek to infer changes in diversification rates, origination (or specifically speciation, when data resolution allows) and extinction, through time or across lineages (see Table 1). This major distinction determines the types of hypotheses which these different approaches can be used to test.

Typically, the raw data used to infer biodiversity in deep time are lists of fossil occurrences, denoting the presence of specific taxa in a particular place, at a particular point in geological time. Often, these data are sourced from large community databases, such as the Paleobiology Database (Uhen et al. 2023), Geobiodiversity Database (Fan et al. 2013) or Neotoma (Williams et al. 2018). The quality of such data can be suboptimal, and it is important to verify that the data are as correct and clean as possible prior to subsequent analyses (Jones et al. 2025). Even after data cleaning, the accuracy and precision of taxonomic, spatial and temporal information can be highly variable (e.g. Hopkins et al. 2018; Buffan et al. 2023) (but note model-based efforts to increase temporal data resolution, such as CONOP [Sadler & Cooper, 2008; Fan et al. 2020] and HORSE [Chu et al. 2025]). Alternatively, phylogenies can be inferred in conjunction with evolutionary rates, or be used as a template from which to estimate past diversity (e.g. Wright et al. 2022; Mulvey et al. 2025). The occurrence birth-death model combines both phylogenetic and occurrence data to infer taxonomic richness trajectories (Andréoletti et al. 2022).

Below, we discuss the philosophy behind each of the main "families" of metrics, as well as the criticisms that have been levelled against them.

Range-through analysis

One of the most apparent ways in which fossil occurrence lists are incomplete is the presence of "gaps" in taxon stratigraphic ranges, sometimes called "ghost ranges", due to the fact that the rock record is not continuous across space and time. A simple way to account for gaps in species ranges is to "range-through" taxon occurrences between their first and last appearance in the fossil record (Sepkoski et al. 2002; Sepkoski 1984). This method is based on the fundamental evolutionary assumption that a taxon is highly unlikely to have originated twice and thus must have been extant during each interval between its first and last occurrence, even if it is not sampled in all of those time bins. Although ranged-through richness accounts for gaps in ranges, it does not assess the extent to which first and last occurrences deviate from the true origination or extinction times, which may well be considerably older or younger than is preserved in the fossil record (Signor & Lipps 1982; Holland & Patzkowsky 2002) (although see methods for estimating confidence intervals on first and last occurrences, e.g. Marshall 1990, 1997; Wang et al. 2012). Because the metric relies on a cohesive taxonomic (species) concept, taxonomic uncertainty and violations of this taxonomic concept (for example, cryptic species) can lead to errors. It is also vulnerable to the pull of the recent (Raup 1972, 1979), and is inappropriate for use on regional or local data sets, as it cannot account for regional extinction.

Fair sampling

One of the first trains of thought concerning the impact of fossil record bias on estimates of biodiversity relates to "fairness". If we want to make comparisons of taxonomic diversity across space, time or clades, we ideally want to be making estimates based on equivalent samples from each of our groups of interest. The first step in this direction was a simple one: perhaps an equal number of fossils could be sampled from each time bin, with the number of taxa represented by those fossils illustrating the relative diversity through time. This approach, first implemented using the fossil record by Raup (1975), is now known as “classical rarefaction”, although the term “subsampling” is also commonly used to refer to methods that purposefully use less than the total available data. Given that the fossil record is already sparse, the discarding of data when using these methods has been a major source of criticism in their use. The opposite approach, extrapolation, has also been implemented, with the idea being that statistical models can be used to estimate biodiversity by extrapolating from low to higher sampling levels, such as by Dodson (1990) and Russell (1995). Here, criticism rests on the fact that extrapolation beyond what we know from data is heavily reliant on the correctness of the assumptions of the model and data used, which can be difficult to evaluate.

While rarefaction and extrapolation remain a viable approach for estimating past diversity, ways of determining “fair” samples have become more sophisticated. In particular, coverage-based approaches use more robust quantifications of sampling effort to determine equivalent fossil subsamples between which to compare species diversity (Wang & Dodson 2006). Shareholder quorum subsampling (Alroy 2010) uses this approach, with Hill numbers (Hill 1973) used to evaluate when equivalent levels of sampling effort (or “quorum levels”) have been sampled from each temporal or spatial bin (see Roswell et al. 2021 for a neontological review on this topic). Hill numbers are calculated using species abundance distributions, and therefore this approach utilises the abundance of each taxon, not just its presence. In the R package iNEXT (Chao & Jost 2012; Hsieh et al. 2016), coverage-based subsampling and extrapolation can both be used to estimate species diversity at a range of quorum levels. Although an improvement on simple equal subsamples, coverage-based approaches interpret the abundance of common versus rare species in taxon samples, and are therefore reliant on a fairly standardised curve relating abundances to diversity when estimating sampling effort. As a result, this approach may perform poorly if the true community possessed a relative abundance distribution which differed from that assumed by the model, or if sampling filters have substantially skewed relative abundance distributions, both of which are difficult to evaluate (Close et al. 2018).

Cohort analysis

Some of the most popular methods for estimating origination and extinction rates in deep time are those which rely on cohort analysis. First developed by Raup (1978), cohort analysis considers fossil occurrences within a series of discrete time bins. Any changes in taxonomic lists between adjacent time bins are used to determine taxon gains or losses across the bin boundary, and deemed to be indicative of speciation and extinction events. The approach has particular appeal because it fits well with the natural temporal structure of fossil data: often occurrences are dated to a named geological interval, rather than a specific numerical age, and these data are therefore suited to cohort analysis without the need for additional refinement of fossil ages. A series of metrics have been developed based on these approaches, including the boundary-crosser (Alroy 1996; Foote 1999), three-timer (Alroy 2008), gap-filler (Alroy 2014) and second-for-third (Alroy 2015) metrics. Over time, these metrics have been developed to quantify ghost ranges in increasingly sophisticated ways, and use this information to estimate (and correct for) incomplete fossil sampling; however, these developments have also resulted in the discarding of increasing amounts of occurrence data (Warnock et al. 2020). Regardless, there is large appeal in these methods because they are somewhat intuitive, and because they are relatively easy to implement, particularly since their inclusion in the R package *divDyn* (Kocsis et al. 2019).

Residual modelling using sampling proxies

Difficulties in determining how to implement "fair" subsampling methods led to the concept that additional data could be used to quantify differences in sampling intensity between time bins. A methodology was proposed by Smith and McGowan (2007), and refined by Lloyd (2012), to use a sampling proxy, i.e. an indirect measure of sampling effort, to "correct" raw palaeodiversity curves using a regression model-fitting approach. In this method, the sampling proxy is used to identify times of relatively poor and good sampling, and a regression model is fitted to determine how well the sampling signal predicts the biodiversity curve. The residuals of the regression model indicate periods of time where the model does not perfectly predict biodiversity (hence the name "residual modelling"), and can be used to highlight time periods that are more diverse or less diverse than would be expected given the level of sampling for that time interval (Smith & McGowan 2007; Lloyd 2012). Residual modelling does not require large amounts of occurrence data, has therefore been particularly popular in vertebrate palaeontology, for exploring the impact of sampling bias on raw diversity curves (e.g. Barrett et al. 2009; Benson et al. 2010; Butler et al. 2012).

The most commonly used sampling proxies have focused on geological biases, such as rock outcrop area (Smith & McGowan 2007; Dunhill et al. 2012, 2013, 2014a). However, it is unclear whether the close correlation between the amounts of rock preserved and palaeodiversity through time (Raup 1972; Na et al. 2023; Ye & Peters 2023) result from geological megabias or "common cause", whereby Earth system changes drive trends in both the rock and fossil records simultaneously (Peters 2005; Hannisdal & Peters 2011). The number of fossiliferous formations has also been a popular proxy (Barrett et al. 2009; Benson et al. 2010; Butler et al. 2012), and has been claimed capture aspects of rock volume, facies heterogeneity, geographical and temporal dispersion, and research effort (Benson & Upchurch 2013). Despite this, many studies have been critical of the use of formation counts in residual modelling, particularly when only considering formations that bear the fossil group in question, due to inherent redundancy between formation counts and the biodiversity signal which it seeks to correct (Crampton et al. 2003; Benton et al. 2011; Dunhill et al. 2014b; Benton 2015; Dunhill et al. 2018). It has been demonstrated that "correcting" palaeodiversity curves using residual modelling based on formation counts can lead to richness estimates that are further from the truth than the raw fossil record (Brocklehurst 2015; Dunhill et al. 2018). Residual modelling also performs poorly when both low diversity and poor sampling occur simultaneously, such as after a mass extinction event.

Mechanistic model-based approaches

Mechanistic models aim to describe the processes which generated our fossil data, using model parameters that have a tangible biological or geological interpretation, such as rates of sampling, origination, and extinction (e.g. Warnock et al. 2020). Each process is defined based on expectations, also known as assumptions, of how it behaves. Model performance therefore hinges on (1) whether all of the important processes have been included in the model, and (2) whether the processes have been modelled in a way which adequately represents how they work.

The most straightforward mechanistic modelling approach is to model the fossil sampling process only, without making any assumptions about the evolutionary processes. For example, TRiPS (True Richness estimated using Poison Sampling) models sampling using a Poisson process (Starrfelt & Liow 2016). Per-interval sampling rates are quantified from fossil occurrence data, then transformed into per-interval probabilities, to extrapolate an estimate of true richness. A major advantage of modelling the sampling process explicitly is that very sparse sampling, including singletons, can be an expected outcome of the process, meaning fossil occurrence data does not have to be subsampled *a priori*.

More complex models include the demographic or diversification processes, in addition to sampling. Capture-mark-recapture (CMR) models are a class of models largely used by ecologists to estimate population size dynamics from incomplete samples of contemporary populations (Nichols & Pollock 1983; Liow & Nichols 2010). CMR has been adapted for use in paleobiology, and although not yet widely applied, there have been several model extensions relevant to fossil data, such as allowing for variation in sampling across time, space and lineages (Liow & Nichols 2010; Laake 2013). In contrast, the widely used program PyRate implements a range of birth-death process models (Silvestro et al. 2014; Silvestro et al. 2019). Birth-death models have a long history in paleobiology (e.g. Raup 1985; Huelsenbeck & Wagner 1996; Foote 2000) and provide an intuitive approach for modelling branching processes resulting from origination and extinction events. These can also be combined with a model describing how the evolutionary process has been sampled. In PyRate, origination, extinction and fossil sampling are considered independent Poisson processes, with parameter values estimated from fossil occurrence data. PyRate also includes model variants that allow for different combinations of rate variation. Both CMR and PyRate allow users to test explicitly whether there has been a change in diversification rates across interval boundaries, making it easy to compare their results with those generated from simpler cohort models.

Birth-death processes, or closely related approximations, also underpin most recent phylogenetic approaches used to infer diversification history, since phylogenetic trees directly capture the inferred branching history (Wright et al. 2022; Morlon et al. 2024). Phylogenetic approaches for estimating past diversity include the software BAMM, which can be used to estimate lineage-specific evolutionary rates based on a fixed tree (Rabosky 2014; Mitchell et al. 2019), and the fossilized birth-death (FBD) process family of models (Stadler 2010; Heath et al. 2014), which can be applied in a wide range of simulation and inference contexts (see Mulvey et al. 2025 for a review of applications). Phylodynamic models can be used to infer the tree and estimate the parameters associated with the underlying tree-generating processes simultaneously (e.g. Close et al. 2015, Andréoletti et al. 2022). Being able to estimate the tree and divergence (origination) times also means we can quantify the extent of ghost lineages more comprehensively within the same model. Different variants of birth-death process models can be applied to trees with extant taxa only, trees that combine both extant and extinct taxa, or trees that include extinct taxa only (see MacPhearson et al. 2022 for an overview of the theory). While questions have been raised as to whether these models are fully identifiable (Louca & Pennell 2020, 2021), theoretical work indicates that phylogenies containing fossils should only be compatible with a single set of evolutionary rate trajectories (Truman et al. 2025). Another class of phylodynamic models are coalescent processes, based on a model of branching history in which lineages coalesce backwards in time, which have only recently been explored in the context of paleobiology (Allen et al. 2024). While phylodynamic models have mainly been used to estimate diversification rates, they can also be used to infer richness using an emerging suite of tools (Vaughan et al. 2019; Andréoletti et al. 2022; Vaughan & Stadler 2024).

Model-based approaches are often straightforward to implement in a Bayesian framework, allowing uncertainty to be quantified naturally during inference, and allowing us to take advantage of existing (prior) knowledge (Wright et al. 2022). The use of prior parameter constraints allows us to include data beyond the fossil occurrences in our models, such as information about how sampling has varied through time (see section *Residual modelling using sampling proxies*). However, these approaches tend to be more computationally expensive than their non-model-based counterparts, and can be much more difficult to use (e.g. Barido-Sottani et al. 2024). Any results are sensitive to the model assumptions, meaning that violation could lead to false conclusions, but the importance of assumptions in mechanistic modelling means they are more often explicitly stated, and can be scrutinized with more rigor using approaches like model adequacy (e.g. Duchêne et al. 2019).

Machine learning

Some of the newest techniques available to computational palaeobiologists use machine learning. This includes text mining approaches to collate fossil occurrences (e.g. Peters et al. 2014, Kopperud et al. 2019), using decision trees to infer extinction drivers (e.g. Foster et al. 2023), and identifying and discriminating morphological features using image recognition (He et al. 2024). Deep learning methods also exist for estimating species diversity in deep time. DeepDive (Cooper et al. 2024a, b) simulates species-through-time curves before and after the influence of incomplete sampling to emulate the fossil record, and trains a neural network to understand the relationship between the two. The trained network can then be applied to real fossil occurrence data, in order to infer the shape of the complete diversity curve. In contrast, the birth-death neural network (BDNN; Hauffe et al. 2024) model infers trajectories of speciation, extinction and fossilisation rates through time. The model uses a neural network to estimate these trajectories in light of a range of time series, such as environmental or sampling proxy data, providing insight into which were most influential in driving the estimated speciation and extinction rates. The limitations of these approaches lie in their complexity: they can be difficult to set up, and require more computational resources than the simpler metrics. Further, neural networks are reliant on the simulations they are trained on; should the simulations be substantially incorrect, this will also feed through into the inferred trajectories. In an attempt to address this latter point, DeepDive conducts autotuning (Cooper et al. 2024b), during which the parameters used in simulation are iteratively corrected in order to produce sampled diversity curves more in line with the provided fossil data.

Understanding geographic biases

Temporal bias has long been considered the main facet of unevenness in the fossil record. However, recent debate has turned to spatial biases as a major way in which our perception of diversity in deep time is skewed (e.g. Close et al. 2020a, Jones et al. 2021, Flannery-Sutherland et al. 2022, Raja et al. 2022, Antell et al. 2024). This includes highlighting that supposedly "global" analyses infer metrics using a set of sampled regions, which is unlikely to produce fair estimates as long as sampling is incomplete (Benson et al. 2021). Russell (1995) made an early attempt to correct for uneven geographic sampling, by estimating the amount of land area from which dinosaur fossils had already been collected, and using this information to extrapolate to the total number of dinosaurs which might have existed, should the whole Earth be sampled equivalently.

One approach to this problem is subsampling, with equivalent numbers of fossils selected from different geographic areas in order to more fairly estimate differences in biodiversity across space (e.g. Close et al. 2020a, b). The R package *divvy* aims to make this easier to implement (Antell et al. 2024). Thus far, this approach has been implemented as a data treatment, with geographic subsampling performed prior to species diversity estimation using one of the other described metrics. This means that temporal and spatial biases are typically handled in different steps within an analytical pipeline for inferring palaeodiversity: how we can consider these biases simultaneously is an important area of future research.

Another useful approach for understanding geographic sampling bias is occupancy modelling (Foote et al. 2007; Foote 2016; Kiessling & Kocsis 2016; Dean et al. 2025). This conceptual framework has been developed and applied by neontologists for decades (e.g. MacKenzie et al. 2002), but is yet to be widely used in palaeontology. Within repeatedly-sampled geographic regions (such as multiple stratigraphic levels or localities in a grid square), the model aims to differentiate between taxa missing from occurrence lists due to incomplete sampling, compared to those missing due to true (biological) absence. The model outputs provide an estimate of relative sampling adequacy between geographic regions. More sophisticated occupancy models can also be used to determine the relative importance of different drivers of sampling incompleteness, such as geological versus collection processes (Dean et al. 2025). Although useful for understanding differences in fossil sampling across space, there is currently no clear way to translate these insights into improved estimates of palaeodiversity.

Table 1. A list of methods for inferring biodiversity metrics from the fossil record which take sampling bias into account in some way. Starred citations are the first description, or application, of the method specifically relating to palaeontological data.

Name	**Metric type**	**Data needed**	**Citation**	**Description**	**Implementations**
(Classical) Rarefaction	Diversity	Taxon occurrences	Sanders (1968) *Raup (1975)	Uses subsampling to the size of the smallest sample to compare species richness between samples	divDyn R package (Kocsis et al. 2019)
Range-through	Diversity	Taxon occurrences	Raup (1972)	Corrects estimates of diversity through time by filling in ghost ranges (gaps in the temporal distribution of taxa)	
Coalescent model	Diversification rates	Phylogeny (or cladistic matrix)	Kingman (1982) *Allen et al. (2024)	Uses the branching (or “coalescence”) times in a phylogeny to infer the diversification rate, based on a Wright-Fisher population model of exponential growth	BEAST2 (Bouckaert et al. 2019)
Capture-mark-recapture (CMR)	Diversity Origination/extinction rates	Taxon occurrences (Proxy data optional)	*Nichols & Pollock (1983)	Uses comparison of repeated samples from the same pool to infer sampling probabilities, which are used to estimate total diversity	MARK (White & Burnham 1999) RMark R package (Laake 2013)
(Simple) Extrapolation	Diversity	Taxon occurrences Proxy data	Dodson (1990) Russell (1995)	Uses a metric of collection effort to estimate the asymptote of total diversity	
Abundance-based coverage estimator (ACE)	Diversity	Taxon occurrences	Chao & Lee (1992) *Anderson et al. (1996) Wang & Dodson (2006)	Uses the abundance counts of rare taxa to estimate sampling coverage (completeness) and therefore the total number of taxa	

Boundary-crosser	Origination/extinction rates	Taxon ranges	Alroy (1996) Foote (1999, 2000)	Uses cohort analysis (Raup 1978) to count the taxa which do and do not cross temporal boundaries	divDyn R package (Kocsis et al. 2019)
Three-timer	Origination/extinction rates	Taxon occurrences	Alroy (2008)	Builds on boundary-crosser by correcting based on proportion of “three-timers” (taxa that are present, absent, then present again)	divDyn R package (Kocsis et al. 2019)
Coverage-based rarefaction and extrapolation, or Shareholder quorum subsampling (SQS)	Diversity	Taxon occurrences	Alroy (2010) Chao & Jost (2012)	Uses Hill numbers (Hill 1973) to estimate sampling completeness, or “quorum”, and uses this to subsample or extrapolate to a consistent level of completeness	iNEXT R package (Hsieh et al. 2016)
Phylogenetic generalized linear mixed models	Origination/extinction rates	Phylogeny (Proxy data optional)	Hadfield (2010) *Sakamoto et al. (2016)	Estimates speciation and extinction rates from the branching times and tip dates in a fixed phylogeny	MCMCglmm R package (Hadfield 2010)
Fossilised birth-death model (FBD)	Origination/extinction rates	Phylogeny (or cladistic matrix)	Stadler (2010) Stadler et al. (2018) *Heath et al. (2014) Gavryushkina et al. (2014)	Estimates branching times and tip dates in a phylogeny, in concert with constant speciation and extinction rates, in a Bayesian framework	See review by Mulvey et al. (2025)
Gap-filler	Origination/extinction rates	Taxon occurrences	Alroy (2014)	Builds on three-timer by allowing for longer gaps between occurrences	divDyn R package (Kocsis et al. 2019)
Fossilised birth-death model, skyline version	Origination/extinction rates	Phylogeny (or cladistic matrix)	Gavryushkina et al. (2014)	Estimates branching times and tip dates in a phylogeny, in concert with piecewise-constant speciation and	See review by Mulvey et al. (2025)

				extinction rates, in a Bayesian framework	
PyRate	Origination/extinction rates	Taxon occurrences (Proxy data optional)	Silvestro et al. (2014)	Uses a range of models describing speciation, extinction, and fossil sampling, that can be used to infer rates in a Bayesian framework	PyRate Python library (Silvestro et al. 2014, 2019)
Second-for-third	Origination/extinction rates	Taxon occurrences	Alroy (2015)	Builds on gap-filler by quantifying specific temporal gaps separately	divDyn R package (Kocsis et al. 2019)
True Richness estimated using Poison Sampling (TRiPS)	Diversity	Taxon occurrences	Starrfelt & Liow (2016)	Models fossil sampling using a Poisson process to extrapolate estimates of total diversity	Supplemental R code from Starrfelt & Liow (2016)
Squares	Diversity	Taxon occurrences	Alroy (2018) *Allen et al. (2020)	Uses abundance structure within samples to determine how to extrapolate total diversity	Supplemental R code from Allen et al. (2020)
Fossilised birth-death model, treeless version	Origination/extinction rates	Taxon occurrences	Stadler et al. (2018) *Warnock et al. (2020)	Estimates speciation, extinction, and fossil sampling rates, accounting for the underlying phylogenetic branching process	DPPDiv (Heath et al. 2012; Warnock et al. 2020)
Fossil Bayesian Analysis of Macroevolutionary Mixtures (BAMM)	Speciation/Extinction rates	Phylogeny	Mitchell et al. (2019)	Estimates speciation and extinction rates from the branching times and tip dates in a fixed phylogeny, within a Bayesian framework	BAMM (Rabosky 2014) BAMMtools R package (Rabosky et al. 2014)
Occurrence birth-death model	Diversity Origination/extinction rates	Phylogeny (or cladistic matrix) Taxon occurrences	Andréoletti et al. (2022)	Estimates branching times and tip dates in a phylogeny, in concert with piecewise-constant speciation and extinction rates, plus a	OBD model in RevBayes (Höhna et al. 2016; Andréoletti et al. 2022)

				species richness trajectory through time, in a Bayesian framework	
DeepDive	Diversity	Taxon occurrences	Cooper et al. (2024)	Trains a neural network on simulated data matching the real data, to infer total diversity	DeepDive Python library (Cooper et al. 2024a) DeepDiveR R package (Cooper et al. 2024b)
Birth-Death Neural Network (BDNN)	Speciation/extinction rates	Taxon occurrences Proxy data	Hauffe et al. (2024)	Trains a neural network on simulated data matching the real data, alongside proxy time series, to infer speciation and extinction rates, and their drivers	PyRate Python library (Silvestro et al. 2014) Simulations in BDNNsim Python library (Hauffe et al. 2024)
Fossilised birth-death model, multi-type version	Origination/extinction rates	Phylogeny (or cladistic matrix)	Barido-Sottani & Morlon (2025)	Estimates branching times and tip dates in a phylogeny, in concert with per-branch speciation and extinction rates, in a Bayesian framework	MSBD package in BEAST2 (Barido-Sottani et al. 2020a; Barido-Sottani & Morlon 2025)
Fossilised birth-death diffusion model	Diversification rates	Phylogeny	Quintero et al. (2025)	Uses data augmentation to infer viable complete trees, allowing per-branch diversification rates to be estimated, using a model where these rates evolve along the tree under a diffusion process	Tapestree Julia package (Quintero & Landis 2020; Quintero et al. 2025)

Case study: brachiopod diversity in the Permian and Triassic

Methods

To illustrate the differences between diversity estimates generated by the different metrics described here, we applied a range of them to the fossil record of brachiopods across the Permian and Early/Middle Triassic. Brachiopod occurrences from this time interval identified to the species level were downloaded from the Paleobiology Database (Uhen et al. 2023). Occurrences were placed into stage-level bins, with any dated to a coarser temporal resolution discarded. Collections allocated to the same stage and location (matching modern-day latitude and longitude to two decimal places) were pooled into a single "locality". The resulting dataset included 25,678 occurrences across 4,404 pooled collections, containing 4,710 species within 948 genera.

The different metrics were then calculated in R (R Core Team, 2024), on a per-stage basis, at both species and genus levels, using the following approaches:

1) *Raw*. A raw diversity curve was calculated by counting the number of unique taxa sampled within each stage bin.
2) *Range-through.* The R package *palaeoverse* (Jones et al. 2023) was used to calculate ranged-through diversity. First, the function *tax_range_time()* was used to convert the occurrence data into stratigraphic ranges for each unique taxon. Then, the function *tax_expand_time()* was used to create pseudo-occurrences for each taxon within each stage of their temporal range. The total number of pseudo-occurrences for each stage was then counted.
3) *Simple rarefaction*. Simple rarefaction was conducted by calculating diversity within an equal number of pooled collections for each stage bin. First, the stage with the smallest number of collections was identified: this was the Olenekian, with 93 collections. For each stage bin, 93 pooled collections were sampled (without replacement), and the taxonomic diversity within this sample was quantified; this was repeated 100 times for each stage.
4) *Squares.* Diversity within each stage bin was estimated using the Squares extrapolator (Alroy, 2018) based on the code of Allen et al. (2020).
5) *Coverage-based rarefaction and extrapolation*. The function *estimateD()* in the R package *iNEXT* (Hsieh et al. 2016) was used to estimate diversity, at a quorum level of 0.8, for each stage bin. Hsieh et al. (2016) recommend that extrapolated values above twice the observed sample size are discarded; 0.8 was the highest quorum level at which this was not necessary.
6) *Residual modelling.* The number of geological formations included within the occurrence dataset for each stage was counted, for use as the sampling proxy for residual modelling. Code from Lloyd (2012) was then used to calculate residual diversity between the raw taxon curve and the number of formations through time.

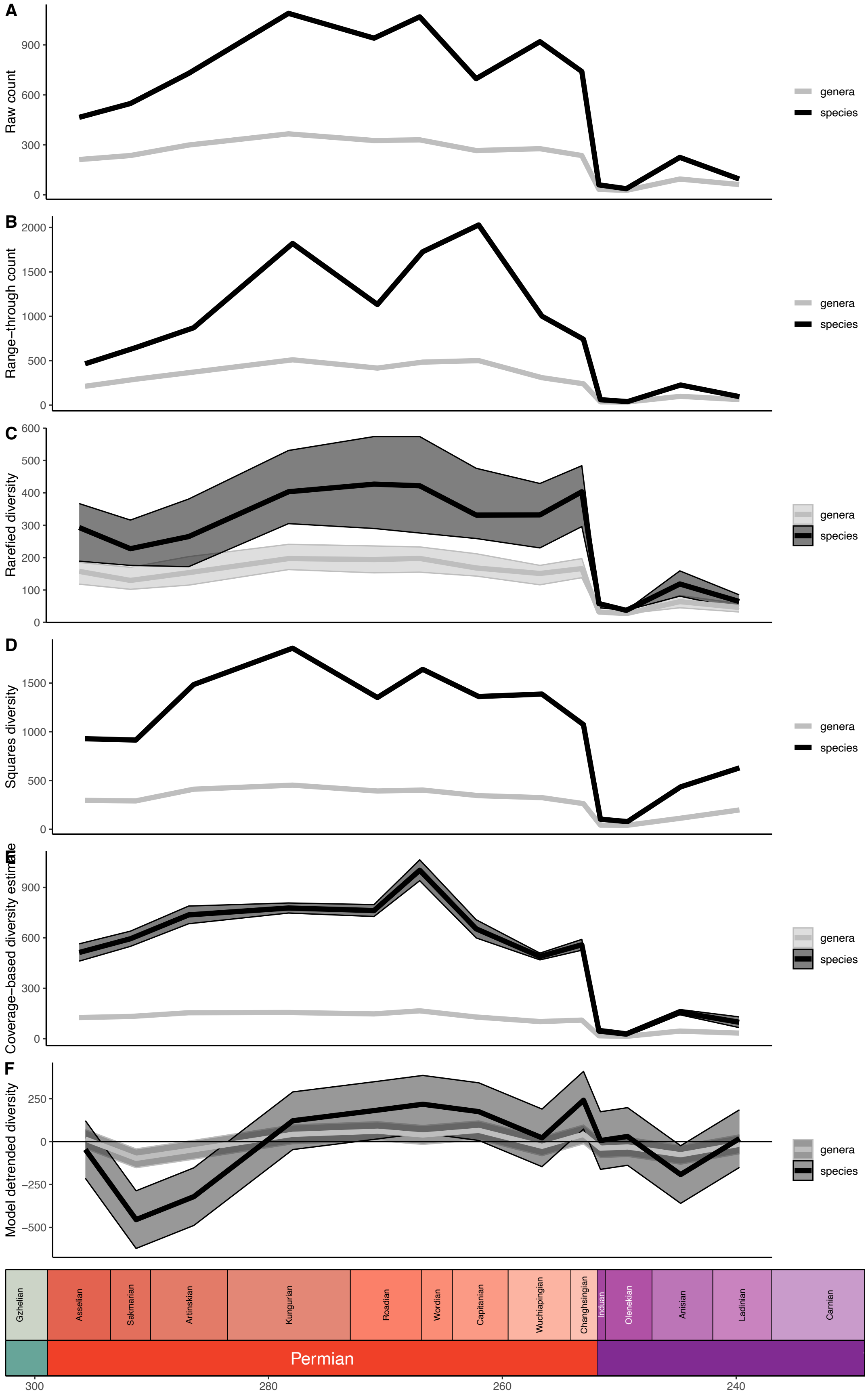
For each metric (except residual modelling), the proportional change in diversity across each stage boundary was also calculated.

Results

The different metrics produce comparable trends at the broad scale, showing roughly constant diversity throughout the Permian, followed by a steep drop at the end of the Permian, with some limited recovery during the Middle Triassic (Figure 1). However, the timing, strength, and even direction of diversity fluctuations varies considerably based on the metric chosen. All metrics show a comparably-severe mass extinction at the end of the Permian (species losses of 92% for raw data, 92% for range-through, 86% for rarefied, 90% for Squares, and 92% for coverage-based estimates; see Table 2). By contrast, the metrics vary considerably in their estimates of diversity trends at the end of the Capitanian (species gains of 32% for raw data, losses of 51% for range-through, gains of 1% for rarefied, gains of 2% for Squares, and losses of 25% for coverage-based estimates; see Table 2). The curve produced by residual modelling also shows comparable diversity loss between the Early and Middle Triassic, and across the Permian-Triassic boundary (Figure 1). These overall trends are observable at both species and genus level (Figure 1, 2). Clapham et al. (2009) also reported that sampling standardisation of fossil occurrences resulted in observing a gradual diversity decline from the Wordian to the Induan for marine invertebrates, due to reduced origination rates in the Capitanian and Wuchiapingian.

Figure 1. Brachiopod diversity through the Permian and Triassic estimated using different methods. Black (or dark grey) lines show the species-level data, while grey (or light grey) lines show the genus-level data. Panels show (a) raw diversity, (b) range-through diversity, (c) diversity rarefied using locality counts, with line showing median values and error envelope showing full range across 100 repetitions, (d) diversity estimates using the Squares extrapolator, (e) diversity estimated using coverage-based rarefaction and extrapolation, to a quorum level of 0.8, with 95% confidence intervals, and (f) model-detrended diversity based on the number of formations sampled through time, with error envelope showing standard deviation. The geological timescale axis was added using *deeptime* (Gearty, 2024).

Figure 2. Brachiopod generic diversity through the Permian and Triassic estimated using different methods. See caption for Figure 1 for panel details.


A
Raw count
900
600
300
0
genera
species
B
Range−through count
2000
1500
1000
500
0
C
Rarefied diversity
600
500
400
300
200
100
D
Squares diversity
E
Coverage−based diversity estimate
F
Model detrended diversity
250
−250
−500
Gzhelian
Asselian
Sakmarian
Artinskian
Kungurian
Roadian
Wordian
Capitanian
Wuchiapingian
Changhsingian
Induan
Olenekian
Anisian
Ladinian
Carnian
Permian
300
280
260
240

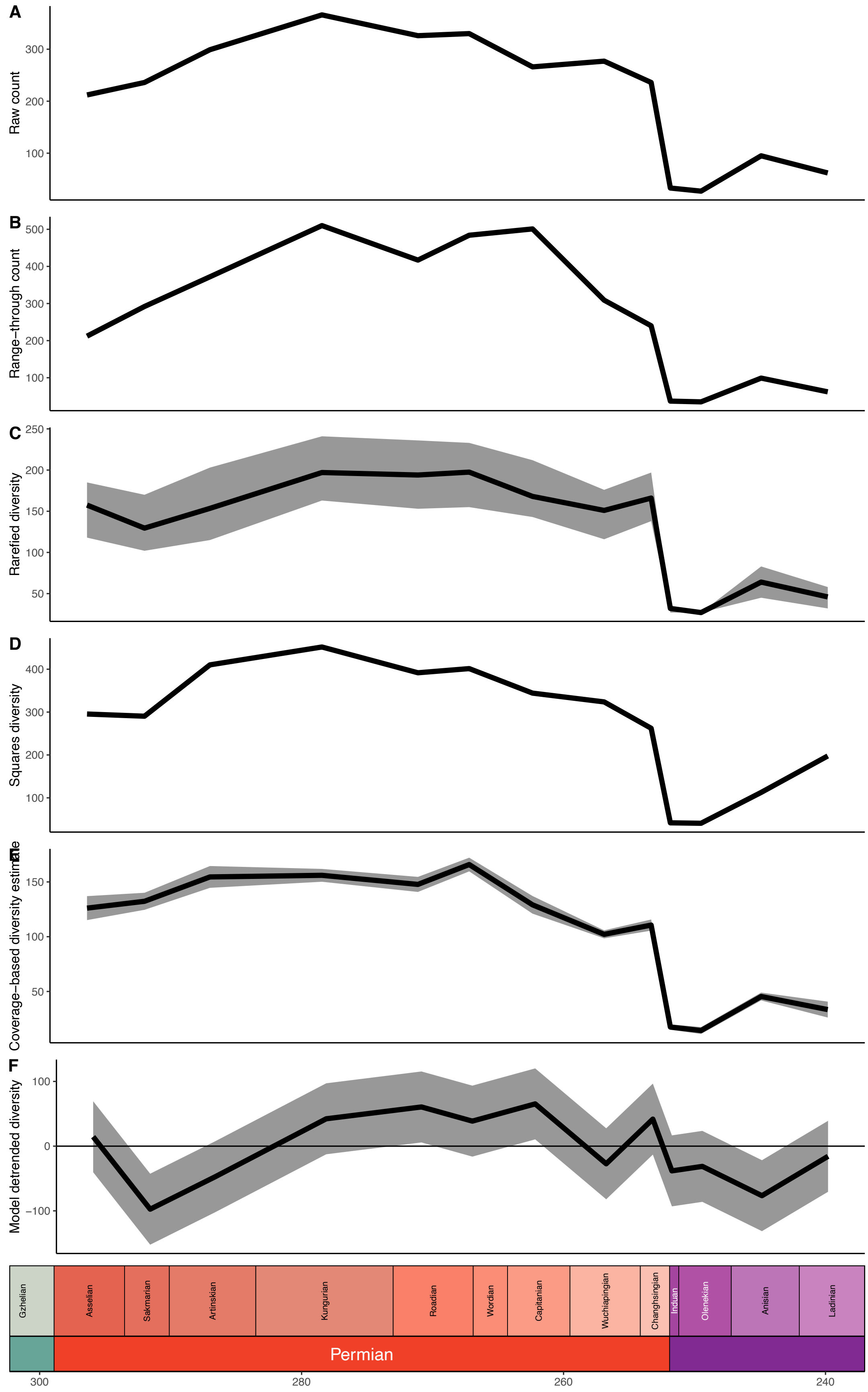

A
Raw count
100
200
300
B
Range−through count
100
200
300
400
500
C
Rarefied diversity
50
100
150
200
250
D
Squares diversity
100
200
300
400
E
Coverage−based diversity estimate
50
100
150
F
Model detrended diversity
−100
0
100
Gzhelian
Asselian
Sakmarian
Artinskian
Kungurian
Roadian
Wordian
Capitanian
Wuchiapingian
Changhsingian
Induan
Olenekian
Anisian
Ladinian
Permian
300
280
260
240

Table 2. Change in diversity estimates across stage boundaries, as a percentage of diversity in the older stage.

Stages	Level	Raw	Range-through	Rarefied	Squares	Coverage-based
Asselian to Sakmarian	genera	11.32	37.74	-15.38	-1.73	4.94
	species	17.85	39.35	-21.21	-1.44	15.91
Sakmarian to Artinskian	genera	26.69	27.4	17.42	41.24	16.81
	species	33.03	34.41	15.1	62.22	23.9
Artinskian to Kungurian	genera	22.41	37.1	26.45	10.16	0.95
	species	49.38	109.07	52.09	25.17	5.55
Kungurian to Roadian	genera	-10.93	-18.24	-1.28	-13.31	-5.31
	species	-13.68	-37.78	4.38	-27.27	-1.95
Roadian to Wordian	genera	1.23	16.07	1.29	2.48	12.28
	species	13.62	52.43	-0.24	21.42	31.45
Wordian to Capitanian	genera	-19.39	3.51	-12.24	-14.25	-22.18
	species	-34.74	17.49	-19.33	-17.01	-34.77
Capitanian to Wuchiapingian	genera	4.14	-38.32	-13.95	-5.95	-20.86
	species	31.85	-50.62	0.74	1.86	-25.18
Wuchiapingian to Changhsingian	genera	-14.8	-22.33	13.85	-18.99	8.29
	species	-19.48	-25.95	19.65	-22.6	14.23
Changhsingian to Induan	genera	-86.02	-84.58	-81.01	-84.1	-84.16
	species	-91.89	-91.78	-85.68	-90.44	-91.87
Induan to Olenekian	genera	-18.18	-5.41	-15.62	-2.43	-18.11
	species	-38.33	-36.07	-36.21	-24.28	-39.75
Olenekian to Anisian	genera	251.85	182.86	153.7	176.73	216.48
	species	508.11	479.49	240.54	458.62	477.88
Anisian to Ladinian	genera	-34.74	-37.37	-29.93	75.53	-26.29
	species	-57.78	-57.96	-45.63	44.88	-37.99

The future of palaeobiodiversity metrics

Utilising simulations

Although all of the aforementioned methods and metrics were proposed with a logical basis, a major barrier to reliable inference of past biodiversity is that we have a poor understanding of whether any of these methods work well, i.e. we do not know if they bring us closer to the truth than the raw data. An increasing number of studies have attempted to compare the performance of different biodiversity metrics (e.g. Lane et al. 2005; Warnock et al. 2017; Close et al. 2018; Smiley, 2018; Alroy 2020), but it is clear that there is scope for much more rigorous testing.

While we might not know how true biodiversity changed over deep time as a marker against which to test our metric performance, we can use the next best thing, which is simulations (Barido-Sottani et al. 2020b). Simulations enable us to generate believable biodiversity data, often using a mechanistic model describing hypothesised evolutionary processes, that can then be compared to trends in empirical data (e.g. Saupe et al. 2019, 2020; Dunne et al. 2023). We can also subsample this data, illustrating how the fossil record may obscure our view (Liow et al. 2010; Dunhill et al. 2014; Brocklehurst 2015; Barido-Sottani et al. 2019). We can then apply our metrics to the simulated fossil data, and compare the inferences to the true values used in simulation. As well as generally enabling us to quantify how correct our metrics are, a particular advantage of this approach is that simulations can be designed to emulate data from different clades or preservational systems, allowing us to investigate the contexts in which different metrics perform better or more poorly (e.g. Allen et al. 2023). In fact, tuning simulations to replicate specific empirical datasets, prior to or within analyses, can be a powerful tool for exploring uncertainty (e.g. Cooper et al. 2024b; Quintero et al. 2025).

Acknowledging and embracing uncertainty

Methods which provide an estimate of uncertainty around their average value should ultimately be preferred over those which do not. Such error bars or confidence intervals are essential for making informed interpretations concerning the reliability of our diversity estimates. All methods that involve subsampling data can be repeated on multiple samples, producing a range of estimates which can then be summarised. Methods that use a Bayesian approach produce posterior distributions for each model parameter, describing the range of values obtained over the course of a Markov chain Monte Carlo (MCMC) inference (Barido-Sottani et al. 2024).

Estimates of uncertainty can be interpreted in many useful ways. First, they can be used to understand the range of values obtained across an analysis. A wide range might suggest a low amount of certainty in a mean or median value, and cast doubt on the fact that this average should be interpreted at face value. However, the range of values can also be viewed as a way of excluding extremes. For example, if we assume that our model is performing well, we can have relatively high certainty that the true diversity does not lie outside of our confidence interval. Second, particularly for Bayesian analyses, we can interpret the shape of the posterior distribution obtained for our focal parameters. Closer investigation might reveal a multimodal distribution, in which case a mean or median might be a poor representation of the distribution as a whole, instead placing considerable likelihood on two or more different values. Such distributions may result from nonidentifiability, meaning that the priors and data are insufficient to be able to determine a single best-fitting mode (e.g. Louca & Pennell 2020). In this case, further analysis with a larger dataset and/or an alternative model may converge on a single mode (Barido-Sottani et al. 2024).

Having a clear, testable hypothesis

Sometimes deep time diversity is investigated in a manner closer to an open question, e.g. “What was the diversity of my clade through time?” Answering such a question requires diversity estimates that are both accurate and precise, across a large number of data points, which is ambitious given the sparsity of data we have available to us. However, most computational analyses aim to either support or disprove a given hypothesis concerning the system which the data describe (e.g. Hammer & Harper 2024), and adhering to this in the case of diversity estimation is also a good idea. For example, we can investigate whether diversification was positive (net speciation) or negative (net extinction) within a given interval (e.g. Allen et al. 2024). Although our exact estimate of the diversification rate might be inaccurate, or vary across different methods, our conclusion will only be incorrect if this inaccuracy means that the point crosses the zero line. Using methods which provide an estimate of uncertainty can also help us to evaluate the probability with which diversification was positive or negative, including whether one of these solutions is fully excluded from the posterior.

Comparison can also be a useful approach: for example, we could test whether our clade was more diverse in interval X or interval Y. Although sampling incompleteness affects the known fossil record in both intervals, any bias only becomes concerning if it substantially differs between the two time bins. If desired, we can also interpret this further, such as considering that the larger the difference between estimates for the two intervals, the more certainty we might have in which was more diverse.

Improving the quality and volume of data

Ultimately, having more data will give us more statistical power for estimating diversity in deep time. Alongside simply collecting more fossils, we also need these fossils to be described and identified, before being entered in global databases (Alroy 2003, Marshall et al. 2018). Text mining shows promise as a method for facilitating biodiversity data collation more rapidly (e.g. Peters et al. 2014, Kopperud et al. 2019). However, it is important that we maximise efforts to collate data from across the world, but also that researchers in every country have access to, and stakeholdership in, this data (Dunne et al. 2025).

At present, data limitations are a major influence on the methodologies used to estimate palaeodiversity. For example, many studies focus on generic diversity, as such patterns are perceived to be more robust in comparison with the sparser data available for individual species (Hendricks et al. 2016). However, we cannot assume that generic and species-level patterns are related or a proxy for one another, and this must be considered when making comparisons between clades, and with present-day diversity. Further, differences in data availability contribute to the preferential use of occurrence-based methods for invertebrate studies, and phylogenetic methods for vertebrates (e.g. Mulvey et al. 2025). This contributes to difficulties comparing studies across this divide. In future, increased volumes of vertebrate occurrence data, alongside the development of more morphological matrices for invertebrates, may help to close the gap.

Integration with modern data, for extant organisms, is also an area which can be further developed (Dillon et al. 2023; Liow et al. 2023). The use of modern diversity estimates to condition inferred diversity curves can have a large impact on richness estimates through time, particularly in the recent past (e.g. Cooper et al. 2024b). Although modern and fossil occurrence data have their own nuances and biases, these can be modelled separately in order to account for any discrepancies: for example, in the fossilized birth-death model, Ψ describes fossil sampling, while ρ describes extant sampling, allowing these two ways of observing the evolutionary process to be incorporated fairly into the model (Mulvey et al. 2025). In addition, modern data can prove useful as a template through which we can investigate the potential influences of sampling bias on diversity estimates (e.g. Barr & Wood 2024, Krone et al. 2024).

Conclusions

In the absence of a time machine, it is difficult for us to have complete faith in our estimates of biodiversity in deep time. However, a long history of methodological development has provided us with an abundance of ideas and approaches for attempting to estimate palaeodiversity. Choosing a metric to use depends on the hypothesis being tested, the data available, and the computational resources available. It is not guaranteed that different metrics will agree with each other, but we do have tools available to us to make an informed choice about which metrics to use, and comparison between metrics should be conducted where possible. Collecting more data, testing methods with simulations, and being careful about how to interpret our results, can all contribute to making better inferences of palaeodiversity. The future development of new methods may also yet provide more accurate and precise estimates. Such approaches may allow us to address a wide range of questions in macroevolution with increased confidence in our results.

Data availability

Data and code for the presented case study are available at https://github.com/bethany-j-allen/brachiopod_diversity/tree/main.

Acknowledgements

We thank the Computational Evolution group at ETH Zurich for helpful discussion. The contributions of BJA were supported by ETH Zurich, and by Deutsche Forschungsgemeinschaft Collaborative Research Centre 1211 (grant no. 268236062). We thank all those who contributed to compiling the Paleobiology Database data used in this manuscript. This is Paleobiology Database Official Publication No. XXXX.

References

Allen, B. J., Wignall, P. B., Hill, D. J., Saupe, E. E. and Dunhill, A. M. 2020. The latitudinal diversity gradient of tetrapods across the Permo-Triassic mass extinction and recovery interval. *Proceedings of the Royal Society B*, **287**, 20201125.

Allen, B. J., Clapham, M. E., Saupe, E. E., Wignall, P. B., Hill, D. J. and Dunhill, A. M. 2023. Estimating spatial variation in origination and extinction in deep time: a case study using the Permian-Triassic marine invertebrate fossil record. *Paleobiology*, **49**, 509-526.

Allen, B. J., Volkova Oliveria, M. V., Stadler, T., Vaughan, T. G. and Warnock, R. C. M. 2024. Mechanistic phylodynamic models do not provide conclusive evidence that non-avian dinosaurs were in decline before their final extinction. *Cambridge Prisms: Extinction*, **2**, e6.

Alroy, J. 1996. Constant extinction, constrained diversification, and uncoordinated stasis in North American mammals. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **127**, 285-311.

Alroy, J. 2003. Global databases will yield reliable measures of global biodiversity. *Paleobiology*, **29**, 26-29.

Alroy, J. 2008. Dynamics of origination and extinction in the marine fossil record. *Proceedings of the National Academy of Sciences USA*, **105**, 11536-11542.

Alroy, J. 2010. Geographical, environmental and intrinsic biotic controls on Phanerozoic marine diversification. *Palaeontology*, **53**, 1211-1235.

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. 2018. Limits to species richness in terrestrial communities. *Ecology Letters*, **21**, 1781-1789.

Alroy, J. 2020. On four measures of taxonomic richness. *Paleobiology*, **46**, 158-175.

Anderson, J., Anderson, H., Fatti, P. and Sichel. H. 1996. The Triassic Explosion (?): a statistical model for extrapolating biodiversity based on the terrestrial Molteno Formation. *Paleobiology*, **22**, 318-328.

Andréoletti, J., Zwaans, A., Warnock, R. C. M., Aguirre-Fernández, G., Barido-Sottani, J., Gupta, A., Stadler, T. and Manceau, M. 2022. The occurrence birth-death process for combined evidence analysis in macroevolution and epidemiology. *Systematic Biology*, **71**, 1440-1452.

Antell, G. T., Benson, R. B. J. and Saupe, E. E. 2024. Spatial standardization of taxon occurrence data – a call to action. *Paleobiology*, **50**, 177-193.

Barido-Sottani, J., Pett, W., O'Reilly, J. and Warnock, R. C. M. 2019. FossilSim: an R package for simulating fossil occurrence data under mechanistic models of preservation and recovery. *Methods in Ecology and Evolution*, **10**, 835-840.

Barido-Sottani, J. Vaughan, T. G. and Stadler, T. 2020a. A multitype birth-death model for Bayesian inference of lineage-specific birth and death rates. *Systematic Biology*, **69**, 973-986.

Barido-Sottani, J., Saupe, E. E., Smiley, T. M., Soul, L. C., Wright, A. M. and Warnock, R. C. M. 2020b. Seven rules for simulations in paleobiology. *Paleobiology*, **46**, 435-444.

Barido-Sottani, J., Schwery, O., Warnock, R. C. M., Zhang, C. and Wright, A. M. 2024. Practical guidelines for Bayesian phylogenetic inference using Markov chain Monte Carlo. *Open Research Europe*, **3**, 204.

Barido-Sottani, J. and Morlon, H. 2025. Integrating fossil samples with heterogeneous diversification rates: a combined multi-type fossilized birth-death model. *bioRxiv*, 2025.02.03.636219.

Barr, W. A. and Wood, B. 2024. Spatial sampling bias influences our understanding of early hominin evolution in eastern Africa. *Nature Ecology and Evolution*, **8**, 2113-2120.

Barrett, P. M., McGowan, A. J. and Page, V. 2009. Dinosaur diversity and the rock record. *Proceedings of the Royal Society B*, **276**, 2667-2674.

Benson, R. B. J., and Upchurch, P. 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., Butler, R. J., Lindgren, J. and Smith, A. S. 2010. Mesozoic marine tetrapod diversity: mass extinctions and temporal heterogeneity in geological megabiases affecting vertebrates. *Proceedings of the Royal Society B*, **277**, 829-834.

Benton, M. J., Dunhill, A. M., Lloyd, G. T., and Marx, F. G. 2011. Assessing the quality of the fossil record: insights from vertebrates. *Geological Society of London Special Publications*, **358**, 63-94.

Benton, M. J. 2015. Palaeodiversity and formation counts: redundancy or bias? *Palaeontology*, **58**, 1003-1029.

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M. A., Wu, C.-H., Xie, D., Zhang, C., Stadler, T., Drummond, A. J. 2019. BEAST2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537.

Brocklehurst, N. 2015. A simulation-based examination of residual diversity estimates as a method of correcting for sampling bias. *Palaeontologia Electronica*, 18.3.7T.

Buffan, L., Jones, L. A., Domeier, M., Scotese, C. R., Zahirovic, S., and Varela, S. 2023. Mind the uncertainty: Global plate model choice impacts deep-time palaeobiological studies. *Methods in Ecology and Evolution*, **14**, 3007-3019.

Butler, R. J., Brusatte, S. L., Andres, B., and Benson, R. B. J. 2012. How do geological sampling biases affect studies of morphological evolution in deep time? A case study of pterosaur (Reptilia: Archosauria) disparity. *Evolution*, **66**, 147-162.

Chao, A. and Lee, S.-M. 1992. Estimating the number of classes via sample coverage. *Journal of the American Statistical Association*, **87**, 210-217.

Chao, A. and Jost, L. 2012. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology*, **93**, 2533-2547.

Chu, T., Huang, H., Fan, J., Deng, Y., Xu, T., Qian, C., Xue, K., Sheets, H. D., Stephenson, M. H., Shi, Y. and Hou, X. 2025. HORSE: harmonize regional and global stratigraphic records through horizon sequencing. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **670**, 112976.

Clapham, M. E., Shen, S. and Bottjer, D. J. 2009. The double mass extinction revisited: reassessing the severity, selectivity, and causes of the end-Guadalupian biotic crisis (Late Permian). *Paleobiology*, **35**, 32-50.

Close, R. A., Friedman, M., Lloyd, G. T., and Benson, R. B. 2015. Evidence for a mid-Jurassic adaptive radiation in mammals. *Current Biology*, **25**, 2137-2142.

Close, R. A., Evers, S. W., Alroy, J. and Butler, R. J. 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., Benson, R. B. J., Saupe, E. E., Clapham, M. E. and Butler, R. B. J. 2020a. The spatial structure of Phanerozoic marine animal diversity. *Science*, **368**, 420-424.

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. 2020b. The apparent exponential radiation of Phanerozoic land vertebrates is an artefact of spatial sampling biases. *Proceedings of the Royal Society B*, **287**, 20200372.

Cooper, R. B., Flannery-Sutherland, J. T. and Silvestro, D. 2024a. DeepDive: estimating global biodiversity patterns through time using deep learning. *Nature Communications*, **15**, 4199.

Cooper, R. B., Allen, B. J. and Silvestro, D. 2024b. DeepDiveR – a software for deep learning estimation of palaeodiversity from fossil occurrences. *bioRxiv*.

Crampton, J. S., Beu, A. G., Cooper, R. A., Jones, C. M., Marshall, B., and Maxwell, P. A. 2003. Estimating the rock volume bias in paleobiodiversity studies. *Science*, **301**, 358-360.

Darwin, C. R. 1859. *On the Origin of Species by Means of Natural Selection*. John Murray, London, United Kingdom. 502pp.

Dean, C. D., Mannion, P. D. and Butler, R. J. 2016. Preservational bias controls the fossil record of pterosaurs. *Palaeontology*, **59**, 225-247.

Dillon, E. M., Pier, J. Q., Smith, J. A., Raja, N. B., Dimitrijević, D., Austin, E. L., Cybulski, J. D., De Entrambasaguas, J., Durham, S. R., Grether, C. M., H. S. Haldar, Kocáková, K., Lin, C.-H., Mazzini, I., Mychajliw, A. M., Ollendorf, A. L., Pimiento, C., Regalado Fernández, O. R., Smith, I. E. and Dietl, G. P. 2022. What is conservation paleobiology? Tracking 20 years of research and development. *Frontiers in Ecology and Evolution*, **10**, 1031483.

Dillon, E. M., Dunne, E. M., Womack, T. M., Kouvari, M., Larina, E., Claytor, J. R., Ivkić, A., Juhn, M., Milla Carmona, P. S., Robson, S. V., Saha, A., Villafaña, J. A. and Zill, M. E. 2023. Challenges and directions in analytical paleobiology. *Paleobiology*, **49**, 377-393.

Dodson, P. 1990. Counting dinosaurs: how many kinds were there? *Proceedings of the National Academy of Sciences USA*, **87**, 7608-7612.

Duchêne, S., Bouckaert, R., Duchene, D. A., Stadler, T. and Drummond, A. J. 2019. Phylodynamic model adequacy using posterior predictive simulations. *Systematic Biology*, **68**, 358-364.

Dunhill, A. M., Benton, M. J., Twitchett, R. J., and Newell, A. J. 2012. Completeness of the fossil record and the validity of sampling proxies at outcrop level. *Palaeontology*, **55**, 1155-1175.

Dunhill, A. M., Benton, M. J., Newell, A. J., and Twitchett, R. J. 2013. Completeness of the fossil record and the validity of sampling proxies: a case study from the Triassic of England and Wales. *Journal of the Geological Society*, **170**, 291-300.

Dunhill, A. M., Benton, M. J., Twitchett, R. J., and Newell, A. J. 2014a. Testing the fossil record: sampling proxies and scaling in the British Triassic-Jurassic. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **404**, 1-11.

Dunhill, A. M., Hannisdal, B., and Benton, M. J. 2014b. Disentangling rock record bias and common-cause from redundancy in the British fossil record. *Nature Communications*, **5**, 4818.

Dunhill, A. M., Hannisdal, B., Brocklehurst, N., and Benton, M. J. 2018. On formation-based sampling proxies and why they should not be used to correct the fossil record. *Palaeontology*, **61**, 119-132.

Dunne, E. M. 2018. Patterns in palaeontology: how do we measure biodiversity in the past? *Palaeontology Online*, **8**, 1-9.

Dunne, E. M., Thompson, S. E. D., Butler, R. J., Rosindell, J. and Close, R. A. 2023. Mechanistic neutral models show that sampling biases drive the apparent explosion of early tetrapod diversity. *Nature Ecology and Evolution*, **7**, 1480-1489.

Dunne, E. M., Chattopadhyay, D., Dean, C. D., Dillon, E. M., Dowding, E. M., Godoy, P. L., Smith, J. A. and Raja, N. B. 2025. Data equity in paleobiology: progress, challenges, and future outlook. *Paleobiology*, **51**, 237-249.

Fan, J., Chen, Q., Hou, X., Miller, A. I., Melchin, M. J., Shen, S., Wu, S., Goldman, D., Mitchell, C. E., Yang, Q., Zhang, Y., Zhan, R., Wang, J., Leng, Q., Zhang, H. and Zhang, L. 2013. Geobiodiversity Database: a comprehensive section-based integration of stratigraphic and paleontological data. *Newsletters on Stratigraphy*, **46**, 111-136.

Fan, J., Shen, S., Erwin, D. H., Sadler, P. M., MacLeod, N., Cheng, X., Hou, X., Yang, J., Wang, X., Wang, Y., Zhang, H., Chen, X., Li, G., Zhang, Y., Shi, Y., Yuan, D., Chen, Q., Zhang, L., Li, C. and Zhao, Y. 2020. A high-resolution summary of Cambrian to Early Triassic marine invertebrate biodiversity. *Science*, **367**, 272-277.

Finnegan, S., Harnik, P. G., Lockwood, R., Lotze, H. K., McClenachan, L. and Kahanamoku, S. S. 2024. Using the fossil record to understand extinction risk and inform marine conservation in a changing world. *Annual Review of Marine Science*, **16**, 307-333.

Flannery-Sutherland, J. T., Silvestro, D. and Benton, M. J. 2022. Global diversity dynamics in the fossil record are regionally heterogeneous. *Nature Communications*, **13**, 2751.

Foote, M. 1999. Morphological diversity in the evolutionary radiation of Paleozoic and post-Paleozoic crinoids. *Paleobiology*, **25**, 1-115.

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

Foote, M., Crampton, J. S., Beu, A. G., Marshall, B. A., Cooper, R. A., Maxwell, P. A. and Matcham, I. 2007. Rise and fall of species occupancy in Cenozoic fossil mollusks. *Science*, **318**, 1131-1134.

Foote, M. 2016. On the measurement of occupancy in ecology and paleontology. *Paleobiology*, **42**, 707-729.

Foster, W. J., Allen, B. J., Kitzmann, N. H., Münchmeyer, J., Rettelbach, T., Witts, J. D., Whittle, R. J., Larina, E., Clapham, M. E. and Dunhill, A. M. 2023. How predictable are mass extinction events? *Royal Society Open Science*, **10**, 221507.

Gavryushkina, A., Welch, D., Stadler, T. and Drummond, A. J. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Computational Biology*, **10**, e1003919.

Gearty, W. 2024. deeptime: an R package that facilitates highly customizable visualizations of data over geological time intervals. *EarthArXiv*, X5841N.

Hadfield, J. D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, **33**, 1-22.

Hammer, Ø., and Harper, D. A. T. 2024. Paleontological Data Analysis (Second Edition). Wiley. pp1-376.

Hannisdal, B., and Peters, S. E. 2011. Phanerozoic Earth system evolution and marine biodiversity. *Science*, **334**, 1121-1124.

Harnik, P. G., Lotze, H. K., Anderson, S. C., Finkel, Z. V., Finnegan, S., Lindberg, D. R., Liow, L. H., Lockwood, R., McClain, C. R., McGuire, J. L., O'Dea, A., Pandolfi, J. M., Simpson, C. and Tittensor, D. P. 2012. Extinctions in ancient and modern seas. *Trends in Ecology and Evolution*, **27**, 608-617.

Hauffe, T., Cantalapiedra, J. L. and Silvestro, D. 2024. Trait-mediated speciation and human-driven extinctions in proboscideans revealed by unsupervised Bayesian neural networks. *Science Advances*, **10**, eadl2643.

He, Y., Mulqueeney, J. M., Watt, E. C., Salili-James, A., Barber, N. S., Camaiti, M., Hunt, E. S. E., Kippax-Chui, O., Knapp, A., Lanzett, A., Rangel-de Lázaro, G., McMinn, J. K., Minus, J., Mohan, A. V., Roberts, L. E., Adhami, D., Grisan, E., Gu, Q., Herridge, V., Poon, S. T. S., West, T. and Goswami, A. 2024. Opportunities and challenges in applying AI to evolutionary morphology. *Integrative Organismal Biology*, obae036.

Heath, T. A., Holder, M. T. and Huelsenbeck, J. P. 2012. A Dirichlet process prior for estimating lineage-specific substitution rates. *Molecular Biology and Evolution*, **29**, 939-955.

Heath, T. A., Huelsenbeck, J. P. and Stadler, T. 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences USA*, **111**, E2957-E2966.

Hendricks, J. R., Saupe, E. E., Myers, C. E., Hermsen, E. J. and Allmon, W. D. 2016. The generification of the fossil record. *Paleobiology*, **40**, 511-528.

Hill, M. O. 1973. Diversity and evenness: a unifying notation and its consequences. *Ecology*, **54**, 427-432.

Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., Huelsenbeck, J. P. and Ronquist, F. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology*, **65**, 726-736.

Holland, S. M. 2017. Structure, not bias. *Journal of Paleontology*, **91**, 1315-1317.

Hopkins, M. J., Bapst, D. W., Simpson, C. and Warnock, R. C. M. 2018. The inseparability of sampling and time and its influence on attempts to unify the molecular and fossil records. *Paleobiology*, **44**, 561-574.

Hsieh, T. C., Ma, K. H. and Chao, A. 2016. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, **7**, 1451-1456.

Huelsenbeck, J. P. and Wagner, P. J. 1996. Stratigraphic Data and Estimates of Phylogenies: Contrasting the Efficiency of Different Methods Using Simulations. *The Paleontological Society Special Publications*, **8**, 184-184.

Hunt, G. and Slater, G. 2016. Integrating paleontological and phylogenetic approaches to macroevolution. *Annual Review of Ecology, Evolution, and Systematics*, **47**, 189-213.

Jones, L. A., Dean, C. D., Mannion, P. D., Farnsworth, A. and Allison, P. A. 2021. Spatial sampling heterogeneity limits the detectability of deep time latitudinal biodiversity gradients. *Proceedings of the Royal Society B*, **288**, 20202762.

Jones, L. A., Gearty, W., Allen, B. J., Eichenseer, K., Dean, C. D., Galván, S., Kouvari, M., Godoy, P. L., Nicholl, C. S. C., Buffan, L., Dillon, E. M., Flannery-Sutherland, J. T. and Chiarenza, A. A. 2023. palaeoverse: a community-driven R package to support palaeobiological analysis. *Methods in Ecology and Evolution*, **14**, 2205-2215.

Jones, L. A., Dean, C. D., Allen, B. J., Drage, H. B., Flannery-Sutherland, J. T., Gearty, W., Chiarenza, A. A., Dillon, E. M., Farina, B. M. and Godoy, P. L. 2025. Ten simple rules to follow when cleaning occurrence data in palaeobiology. *EcoEvoRxiv*, X2FS8M.

Kidwell, S. M. and Holland, S. M. 2002. The quality of the fossil record: implications for evolutionary analyses. *Annual Review of Ecology, Evolution, and Systematics*, **33**, 561-588.

Kiessling, W. and Kocsis, Á. T. 2016. Adding fossil occupancy trajectories to the assessment of modern extinction risk. *Biology Letters*, **12**, 20150813.

Kiessling, W., Smith, J. A. and Raja, N. B. 2023. Improving the relevance of paleontology to climate change policy. *Proceedings of the National Academy of Sciences*, **120**, e2201926119.

Kingman, J. F. C. 1982. The coalescent. *Stochastic Processes and their Applications*, **13**, 235-248.

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 in Ecology and Evolution*, **10**, 735-743.

Kopperud, B. T., Lidgard, S. and Liow, L. H. 2019. Text-mined fossil biodiversity dynamics using machine learning. *Proceedings of the Royal Society B*, **286**, 20190022.

Krone, I. W., Magoulick, K. M. and Yohler, R. M. 2024. All the Earth will not remember: how geographic gaps structure the record of diversity and extinction. *Paleobiology*, **50**, 214-225.

Laake, J. 2013. RMark: an R interface for analysis of capture-recapture data with MARK.

Lane, A., Janis, C. M. and Sepkoski Jr., J. J. 2005. Estimating paleodiversities: a test of the taxic and phylogenetic methods. *Paleobiology*, **31**, 21-34.

Liow, L. H. and Nichols, J. D. 2010. Estimating rates and probabilities of origination and extinction using taxonomic occurrence data: capture-mark-recapture (CMR) approaches. *The Paleontological Society Papers*, **16**, 81-94.

Liow, L. H., Quental, T. B. and Marshall, C. R. 2010. When can decreasing diversification rates be detected with molecular phylogenies and the fossil record? *Systematic Biology*, **59**, 646-659.

Liow, L. H., Uyeda, J. and Hunt, G. 2023. Cross-disciplinary information for understanding macroevolution. *Trends in Ecology and Evolution*, **38**, 250-260.

Lloyd, G. T. 2012. A refined modelling approach to assess the influence of sampling on palaeobiodiversity curves: new support for declining Cretaceous dinosaur richness. *Biology Letters*, **8**, 123-126.

Louca, S. and Pennell, M. W. 2020. Extant timetrees are consistent with a myriad of diversification histories. *Nature*, **580**, 502-505.

Louca, S. and Pennell, M. W. 2021. Why extinction estimates from extant phylogenies are so often zero. *Current Biology*, **31**, 3168-3173.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A. and Langtimm, C. A. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248-2255.

MacPherson, A., Louca, S., McLaughlin, A., Joy, J. B. and Pennell, M. W. 2022. Unifying phylogenetic birth–death models in epidemiology and macroevolution. *Systematic Biology*, **71**, 172-189.

Marshall, C. R. 1990. Confidence intervals on stratigraphic ranges. *Paleobiology*, **16**, 1-10.

Marshall, C. R. 1997. Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. *Paleobiology*, **23**, 165-173.

Marshall, C. R. 2017. Five palaeobiological laws needed to understand the evolution of the living biota. *Nature Ecology and Evolution*, **1**, 0165.

Marshall, C. R., Finnegan, S., Clites, E. C., Holroyd, P. A., Bonuso, N., Cortez, C., Davis, E., Dietl, G. P., Druckenmiller, P. S., Eng, R. C., Garcia, C., Estes-Smargiassi, K., Hendy, A., Hollis, K. A., Little, H., Nesbitt, E. A., Roopnarine, P., Skibinski, L., Vendetti, J. and White, L. D. 2018. Quantifying the dark data in museum fossil collections as palaeontology undergoes a second digital revolution. *Biology Letters*, **14**, 20180431.

Mitchell, J. S., Etienne, R. S. and Rabosky, D. L. 2019. Inferring diversification rate variation from phylogenies with fossils. *Systematic Biology*, **68**, 1-18.

Morlon, H., Andréoletti, J., Barido-Sottani, J., Lambert, S., Perez-Lamarque, B., Quintero, I., Senderov, V. and Veron, P. 2024. Phylogenetic insights into diversification. *Annual Review of Ecology, Evolution, and Systematics*, **55**, 1-21.

Mulvey, L. P. A., Nikolic, M. C., Allen, B. J., Heath, T. A. and Warnock, R. C. M. 2025. From fossils to phylogenies: exploring the integration of paleontological data into Bayesian phylogenetic inference. *Paleobiology*, **51**, 214-236.

Na, L., Li, Q., Krause, C., Zhu, M. and Kiessling, W. 2023. Revisiting the Phanerozoic rock-diversity relationship. *Geological Magazine*, **160**, 1853-1862.

Nichols, J. D. and Pollock, K. H. 1983. Estimating taxonomic diversity, extinction rates, and speciation rates from fossil data using capture-recapture models. *Paleobiology*, **9**, 150-163.

Peters, S. E. 2005. Geologic constraints on the macroevolutionary history of marine animals. *Proceedings of the National Academy of Sciences USA*, **102**, 12326-12331.

Peters, S. E., Zhang, C. Livny, M. and Ré, C. 2014. A machine reading system for assembling synthetic paleontological databases. *PLoS One*, **9**, e113523.

Pimiento, C. and Antonelli, A. 2022. Integrating deep-time palaeontology in conservation prioritisation. *Frontiers in Ecology and Evolution*, **10**, 959364.

Quintero, I. and Landis, M. J. 2020. Interdependent phenotypic and biogeographic evolution driven by biotic interactions. *Systematic Biology*, **69**, 739-755.

Quintero, I., Andréoletti, J., Silvestro, D. and Morlon, H. 2025. The rise, decline and fall of clades. *bioRxiv*, 2025.03.20.644316.

R Core Team. 2024. R: A Language and Environment for Statistical Computing. Version 4.4.2.

Rabosky, D. L. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS One*, **9**, e89543.

Rabosky, D. L., Grundler, M., Anderson, C., Title, P., Shi, J. J., Brown, J. W., Huang, H. and Larson, J. G. 2014. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution*, **5**, 701-707.

Raja, N. B., Lauchstedt, A., Pandolfi, J. M., Kim, S. W., Budd, A. F. and Kiessling, W. 2021. Morphological traits of reef corals predict extinction risk but not conservation status. *Global Ecology and Biogeography*, **30**, 1597-1608.

Raja, N. B., Dunne, E. M., Matiwane, A., Khan, T. M., Nätscher, P. S., Ghilardi, A. M. and Chattopadhyay, D. 2022. Colonial history and global economics distort our understanding of deep-time biodiversity. *Nature Ecology & Evolution*, **6**, 145-154.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G. and Suchard, M. A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, **67**, 901-904.

Raup, D. M. 1972. Taxonomic diversity during the Phanerozoic. *Science*, **177**, 1065-1071.

Raup, D. M. 1975. Taxonomic diversity estimation using rarefaction. *Paleobiology*, **1**, 333-342.

Raup, D. M. 1978. Cohort analysis of generic survivorship. *Paleobiology*, **4**, 1-15.

Raup, D. M. 1979. Biases in the fossil record of species and genera. *Bulletin of Carnegie Museum of Natural History*, **13**, 85.

Raup, D. M. 1985. Mathematical models of cladogenesis. *Paleobiology*, **11**, 42-52.

Roswell, M., Dushoff, J. and Winfree, R. 2021. A conceptual guide to measuring species diversity. *Oikos*, **130**, 321-338.

Russell, D. A. 1995. China and the lost worlds of the dinosaurian era. *Historical Biology*, **10**, 3-12.

Sadler, P. M. and Cooper, R. A. 2008. Best-fit intervals and consensus sequences. *Topics in Geobiology*, **21**, 49-94.

Sakamoto, M., Benton M. J. and Venditti, C. 2016. Dinosaurs in decline tens of millions of years before their final extinction. *Proceedings of the National Academy of Sciences USA*, **113**, 5036-5040.

Sanders, H. L. 1968. Marine benthic diversity: a comparative study. *The American Naturalist*, **102**, 243-282.

Sepkoski, J. J. 1984. A kinetic model of Phanerozoic taxonomic diversity. III. Post-Paleozoic families and mass extinctions. *Paleobiology*, **10**, 246-267.

Sepkoski, J. J. 2002. A compendium of fossil marine animal genera. *Bulletins of American Paleontology*, **363**, 243-282.

Saupe, E. E., Myers, C. E., Peterson, A. T., Soberón, J., Singarayer, J., Valdes, P. and Qiao, H. 2019. Spatio-temporal climate change contributes to latitudinal diversity gradients. *Nature Ecology and Evolution*, **3**, 1419-1429.

Saupe, E. E., Qiao, H., Donnadieu, Y., Farnsworth, A., Kennedy-Asser, A. T., Ladant, J.-B. Lunt, D. J., Pohl, A., Valdes, P. and Finnegan, S. 2020. Extinction intensity during Ordovician and Cenozoic glaciations explained by cooling and palaeogeography. *Nature Geoscience*, **13**, 65-70.

Shaw, J. O., Briggs, D. E. G. and Hull, P. M. 2020. Fossilization potential of marine assemblages and environments. *Geology*, **49**, 258-262.

Silvestro, D., Salamin, N. and Schnitzler, J. 2014. PyRate: a new program to estimate speciation and extinction rates from incomplete fossil data. *Methods in Ecology and Evolution*, **2014**, 1126-1131.

Silvestro, D., Salamin, N., Antonelli, A., and Meyer, X. 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. 2007. Marine diversity through the Phanerozoic: problems and prospects. *Journal of the Geological Society*, **164**, 731-745.

Smith, A. B. and McGowan, A. J. 2007. The shape of the Phanerozoic marine palaeodiversity curve: how much can be predicted from the sedimentary rock record of Western Europe? *Palaeontology*, **50**, 765-774.

Stadler, T. 2010. Sampling-through-time in birth–death trees. *Journal of Theoretical Biology*, **267**, 396-404.

Stadler, T., Gavryushkina, A., Warnock, R. C. M., Drummond, A. J., and Heath, T. A. 2018. The fossilized birth-death model for the analysis of stratigraphic range data under different speciation modes. *Journal of Theoretical Biology*, **447**, 41-55.

Starrfelt, J. and Liow, L. H. 2016. How many dinosaur species were there? Fossil bias and true richness estimated using a Poisson sampling model. *Philosophical Transactions of the Royal Society B*, **371**, 20150219.

Truman, K., Vaughan, T. G., Gavryushkin, A. and Gavryushkina, A. S. 2025. The fossilized birth–death model is identifiable. *Systematic Biology*, **74**, 112-123.

Uhen, M. D., Allen, B. J., Behboudi, N., Clapham, M. E., Dunne, E. M., Hendy, A., Holroyd, P. A., Hopkins, M., Mannion, P., Novack-Gottshall, P., Pimiento, C. and Wagner, P. 2023. Paleobiology Database User Guide Version 1.0. *PaleoBios*, **40**, 1-56.

Vaughan, T. G., Leventhal, G. E., Rasmussen, D. A., Drummond, A. J., Welch, D. and Stadler, T. 2019. Estimating epidemic incidence and prevalence from genomic data. *Molecular Biology and Evolution*, **36**, 1804-1816.

Vaughan, T. G. and Stadler, T. 2024. Bayesian phylodynamic inference of multi-type population trajectories using genomic data. *bioRxiv*, 2024.11.26.625381.

Walker, F. M., Dunhill, A. M. and Benton, M. J. 2020. Variable preservation potential and richness in the fossil record of vertebrates. *Palaeontology*, **63**, 313-329.

Wall, P. D., Ivany, L. C. and Wilkinson, B. H. 2009. Revisiting Raup: exploring the influence of outcrop area on diversity in light of modern sample-standardization techniques. *Paleobiology*, **35**, 146-167.

Wang, S. C. and Dodson, P. 2006. Estimating the diversity of dinosaurs. *Proceedings of the National Academy of Sciences USA*, **103**, 13601-13605.

Wang, S. C., Zimmerman, A. E., McVeigh, B. S., Everson, P. J., and Wong, H. 2012. Confidence intervals for the duration of a mass extinction. *Paleobiology*, **38**, 265-277.

Warnock, R. C. M., Yang, Z. and Donoghue, P. C. J. 2017. Testing the molecular clock using mechanistic models of fossil preservation and molecular evolution. *Proceedings of the Royal Society B*, **284**, 20170227.

Warnock, R. C. M., Heath, T. A., and Stadler, T. 2020. Assessing the impact of incomplete species sampling on estimates of speciation and extinction rates. *Paleobiology*, **46**, 137-157.

White, G. C. and Burnham, K. P. 1999. Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46**, S120-S138.

Williams, J. W., Grimm, E. C., Blois, J. L., Charles, D. F., Davis, E. B., Goring, S. J., Graham, R. W., Smith, A. J., Anderson, M., Arroyo-Cabrales, J., Ashworth, A. C., Betancourt, J. L., Bills, B. W., Booth, R. K., Buckland, P. I., Curry, B. B., Giesecke, T., Jackson, S. T., Latorre, C., Nichols, J., Purdum, T., Roth, R. E., Stryker, M. and Takahara, H. 2018. The Neotoma Paleoecology Database, a multiproxy, international, community-curated data resource. *Quaternary Research*, **89**, 156-177.

Wright, A. M., Bapst, D. W., Barido-Sottani, J., Warnock, R. C. M. 2022. Integrating fossil observations into phylogenetics using the fossilized birth-death model. *Annual Review of Ecology, Evolution, and Systematics*, **53**, 251-273.

Ye, S. and Peters, S. E. 2023. Bedrock geological map predictions for Phanerozoic fossil occurrences. *Paleobiology*, **49**, 394-413.

Žliobaitė, I. and Fortelius, M. 2022. On calibrating the completometer for the mammalian fossil record. *Paleobiology*, **48**, 1-11.