# VARIABLE PRESERVATION POTENTIAL AND RICHNESS IN THE FOSSIL RECORD OF VERTEBRATES

*by* FIONA M. WALKER[1] (iD), ALEXANDER M. DUNHILL[2] (iD) *and* MICHAEL J. BENTON[1,*] (iD)

[1]School of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol, BS8 1RJ, UK; fiona.walker2@gmail.com, mike.benton@bristol.ac.uk
[2]School of Earth & Environment, University of Leeds, Leeds, LS2 9JT, UK; a.dunhill@leeds.ac.uk
*Corresponding author

**Abstract:** Variation in preservation and sampling probability clouds our estimates of past biodiversity. The most extreme examples are Lagerstätten faunas and floras. Although such deposits provide a wealth of information and represent true richness better than other deposits, they can create misleading diversity peaks because of their species richness. Here, we investigate how Lagerstätten formations add to time series of vertebrate richness in the UK, Germany and China. The first two nations are associated with well-studied fossil records and the last is a country where palaeontology has a much shorter history; all three nations include noted Lagerstätten in their fossil records. Lagerstätten provide a larger proportion of China's sampled richness than in Germany or the UK, despite comprising a smaller proportion of its fossiliferous deposits. The proportions of taxa that are unique to Lagerstätten vary through time and between countries. Further, in all regions, we find little overlap between the taxa occurring in Lagerstätten and in 'ordinary' formations within the same time bin, indicating that Lagerstätten preserve unusual faunas. As expected, fragile taxa make up a greater proportion of richness in Lagerstätten than the remainder of the fossil record. Surprisingly, we find that Lagerstätten account for a minority of peaks in the palaeodiversity curves of all vertebrates (18% in the UK; 36% in Germany and China), and Lagerstätten count is generally not a good overall predictor of the palaeodiversity signal. Vastly different sampling probabilities through taxa, locations and time require serious consideration when analysing palaeodiversity curves.

**Key words:** Lagerstätte, biota, diversity, sampling bias, preservation potential, linear model.

THE fossil record is a key source of information on changes in biodiversity through deep time. Correlations between diversity and environmental time series can be tested, and from these, causal relationships hypothesized at different scales (Benton *et al.* 2004; Mayhew *et al.* 2008; Hannisdal & Peters 2011). There are many problems, however, with taking global taxon counts across long spans of time at face value. Raup (1972) argued that the empirical global palaeodiversity signal was driven primarily by differing levels of sampling within each time bin.

Although palaeontologists agree that the fossil record is incomplete, there has been debate about the extent of the problem and how to deal with it (Smith 2001; Benton *et al.* 2011; Smith & McGowan 2011; Mannion *et al.* 2013; Walker *et al.* 2017). One approach is to use sampling proxies such as the number of collections, localities or formations to model geological control of the fossil record (Smith 2007; Smith & McGowan 2007; Lloyd & Friedman 2013). Problems have been noted with the use of strict proxies as they are redundant with the palaeodiversity signal (Benton *et al.* 2011), except in the case of broader sample counts; for example a count of ammonite diversity versus all marine localities. Geological map area avoids many of the redundancy problems, but even this can be problematic as a sampling proxy (Dunhill 2011, 2012; Dunhill *et al.* 2014a, b). Further, Smith & McGowan's (2007) sampling proxy and residual diversity estimates modelling approach produces contrary results and is statistically flawed (Dunhill *et al.* 2014a, b, 2018; Brocklehurst 2015; Sakamoto *et al.* 2016; Close *et al.* 2018). More useful may be subsampling approaches (rarefaction: Tipper 1979; SQS: Alroy 2010) and a Poisson sampling model (TRiPS: Starrfelt & Liow 2016), which have been employed to control for various sampling biases, with varying degrees of success.

We are interested here in the impact of Lagerstätten on palaeodiversity. Lagerstätten were defined originally as deposits or sites that contain 'an unusual amount of palaeontological information' (Seilacher 1970). Seilacher et al. (1985) identified two types: conservation Lagerstätten, which preserve exceptional fossil detail or less recalcitrant ('soft') tissues; and concentration Lagerstätten, which can include an unusually large number and density of taxa or specimens. For the rest of this study we follow the narrower definition of Muscente et al. (2017) who focused on 'exceptionally preserved fossils', so we exclude concentration deposits such as bone beds, comprising mainly hard tissues, often in abraded condition. Lagerstätten are traditionally thought of as representing unusual, localized depositional environments, but there is strong evidence for clustering in space and time, perhaps reflecting global-scale phenomena that enable several similar Lagerstätten to occur in an area or within a time bin (Muscente et al. 2017).

It is commonly assumed that Lagerstätten generate global diversity peaks, for example the Burgess Shale peak in the global Cambrian palaeodiversity plot (Sepkoski 1996), but is this always the case? At a local scale, even a small change in preservation probability (that is, one bed containing highly abundant and more complete fossils) can change the shape of the measured diversity curve in a section (Walker et al. 2017). At a larger scale, and with extremely good preservation, one might expect the effects of this heterogeneity of diversity curves to be even larger. Indeed, Benson et al. (2010) and Benson & Butler (2011) found that Lagerstätten provided more than half of the marine tetrapod taxa of the Triassic–Jurassic.

Although Lagerstätten are exceptional windows into particular facies and ecosystems, and provide a wealth of information for functional inferences and the timeline of key evolutionary innovations (Selden & Nudds 2012), they can, ironically, be a hindrance for diversity studies that require equal sampling throughout. The impact of Lagerstätten varies between taxonomic groups, and is strong for fragile, hollow-boned taxa such as pterosaurs (Dean et al. 2016), birds (Brocklehurst et al. 2012), and bats (Brown et al. 2019). Dean et al. (2016) found a causal relationship between Lagerstätten occurrences and diversity of pterosaurs, and a correlation between skeletal completeness and diversity that broke down when data from Lagerstätten were removed.

Lagerstätten are not the only cause of artificial peaks in raw diversity counts; monographic effects (detailed descriptions of a taxonomic group, occasionally limited to a single time interval or locality) can cause these too (Raup 1972; Fara 2004; Lloyd & Friedman 2013). Without referring to the original data, these may be mistaken for the Lagerstätten effect, and since a prerequisite for writing a monograph is usually an abundance of describable fossils, monographic and Lagerstätten effects may be linked. Therefore, monographic effects must be considered when investigating the diversity of a single taxonomic group, but they are probably less important in studies of numerous taxa over large geographical areas (Raup 1976).

Our aim is to address three questions: (1) Do Lagerstätten significantly alter our perceptions of the vertebrate fossil record? (2) How does this vary between countries with widely different sampling histories? (3) Do Lagerstätten contribute significantly to our understanding of new ecosystems, or do they mostly record previously identified ecosystems in greater detail? To address these questions, we explore how the palaeodiversity signal from three countries (UK, Germany, China) depends on ordinary and Lagerstätten formations, discriminating between marine and terrestrial deposits, and between fragile and robust fossil specimens. We also explore the data using model-fitting approaches to identify potential key drivers of the palaeodiversity signals.

## MATERIAL AND METHOD

### Geographic and taxonomic scaling

In this study, we focus on vertebrates, as there has been a great deal of recent work on their palaeodiversity (Barrett et al. 2009; Benson et al. 2010; Benson & Butler 2011; Fröbisch 2013; Lloyd & Friedman 2013; Close et al. 2017; Cleary et al. 2018; Driscoll et al. 2019), they span both marine and terrestrial environments, and Lagerstätten effects have been identified owing to the fragility of their remains, among, for example: pterosaurs (Butler et al. 2009, 2013; Dean et al. 2016), birds (Fountaine et al. 2005; Brocklehurst et al. 2013), lepidosaurs and lissamphibians (Fara 2002; Cleary et al. 2018) and bats (Brown et al. 2019).

We conduct this study at the scale of countries, as these allow finer subdivision of stratigraphy than global-scale data. Each country has a different combination of facies that may be more or less likely to preserve exceptional fossils. In addition, countries have national geological surveys (and in many cases, researchers) who have erected the stratigraphic frameworks by which rocks are dated, making the correlation of units easier within than between countries. Further, the palaeontological research history may be defined by national boundaries. The regional approach that we take is in line with many recently published studies of diversity and sampling bias (Crampton et al. 2003; Dunhill 2012; Dunhill et al. 2012, 2013, 2014a, b; Benson et al. 2016; Close et al. 2017; Walker et al. 2017).

*Defining Lagerstätten, biotas and formations*

It is an oversimplification to say that the fossil record is made up of Lagerstätten and ordinary deposits. Preservation probability is a continuum (Seilacher *et al.* 1985), depending on the differing taphonomic processes behind each fossiliferous deposit (Allison 1988; Briggs 1995, 2003; Sansom 2014). However, even though Lagerstätten are end members of high preservation probability, palaeontologists generally divide fossiliferous deposits into either Lagerstätten or not (Butler *et al.* 2009, 2013; Benson & Butler 2011; Dean *et al.* 2016); we follow this assumption but consider various ways of defining the cut-off between the two categories.

Geological formations are units that can be tracked laterally over mappable areas. For this study, geological survey definitions of formations are used for Germany and the UK, and 'palaeontological community' definitions (formations that have been entered into the Paleobiology Database; PBDB; https://paleobiodb.org/) are used for the Chinese data). There may be differences between these formation definitions in each country; we explore this issue on the basis of our empirical data and discuss below possible reasons for the differences in duration and definition.

Lagerstätten are usually sites or units below the scale of formation, although there are some exceptions (e.g. the units containing the Jehol Biota). We use the term 'Lagerstätte' rather than 'biota' because this is the common term in two of the countries we sampled (UK and Germany); the term biota, commonly used for Chinese examples, may correspond to a geographically and stratigraphically localized Lagerstätte, or may include numerous separate Lagerstätten (e.g. Yanliao and Jehol biotas). We standardize the scale at which Lagerstätten are designated to formation level. In fact, previously in the diversity count literature, whole formations have been described as Lagerstätten (Benson & Butler 2011; Dean *et al.* 2016), although generally a Lagerstätte is a single horizon or set of horizons within a formation. Formations clearly have different scales (Benton *et al.* 2011), but they are more comparable in scale than one group is to another or one bed is to another, for example. The time span represented by individual formations is also assessed here.

Formations that have been described as, or containing, one or more Lagerstätten are referred to here as 'Lagerstätte formations', and all other formations are termed 'ordinary formations'. Of course, there is much value in studying the variety of sediments labelled 'ordinary formations' here; the label is purely a designation based on a lack of exceptional preservation, rather than on poor preservation.

*Data collection*

Occurrence data for China, the UK, and Germany were downloaded from the PBDB (26 February 2016), searching for each country in turn and the taxon 'Vertebrata'. Over recent years there has been a concerted effort to increase the coverage of a number of vertebrate groups in the PBDB (Mannion *et al.* 2013). We cleaned the data before analysis to remove synonyms, uncertain taxa and trace fossils. Most Lagerstätten were accurately documented because of abundant publications and a focus by data enterers, but some 'ordinary' formations presented less accurate data.

These three countries were chosen because they represent different aspects of sampling. Fossil taxa from the UK and Germany have been named and systematically recorded for hundreds of years, whereas samples from China have only recently been systematically named. Famous Lagerstätten have been found in all three countries, and especially from China, where the Jehol Biota has greatly increased our knowledge of the Cretaceous Terrestrial Revolution (Benton *et al.* 2008).

Stratigraphic information was extracted from the PBDB to produce lists of fossiliferous geological formations for each country. Synonymous geological formations from these data were found and renamed, and the duration of deposition was established. Formations were coded as 'marine', 'terrestrial' or 'mixed/marginal'. For the UK, the majority of the formation names, durations and depositional environments came from the BGS Lexicon of Named Rock Units (British Geological Survey 2016), with a literature search performed for the remaining units (see Walker *et al.* 2019). For Germany, these same data were derived from Litholex (Deutsche Stratigraphische Kommission) and the Stratigraphische Tabelle von Deutschland (Menning 2002). However, such geological survey data on Chinese formations were not available, so ages from the PBDB were used. The PBDB is not the best source of stratigraphic data (Benton *et al.* 2013), even though it has been used extensively elsewhere to collect formation counts; dates relate to individual collections, so they do not always represent the full time span of a formation, and single formations may be ascribed different chronostratigraphic ages depending on the original publication opinions.

Each formation was coded as a Lagerstätte formation or an ordinary formation based on a census of the literature. Some, such as Solnhofen or Jehol, were identified as Lagerstätten in 100% of the literature, whereas others were sometimes called a Lagerstätte, sometimes not. We tried both conservative and permissive approaches and were permissive in making a maximum count of Lagerstätten. According to some, the UK has no

Lagerstätten at all, whereas others, followed here, would allow that, for example, the Early Jurassic Charmouth Mudstone Formation of Lyme Regis (Benson *et al.* 2010) and the Beacon Limestone Formation of Strawberry Bank (Williams *et al.* 2015) show evidence of exceptional preservation. Adopting this 'permissive' approach means we do not underestimate the occurrences of Lagerstätten, and in any case all, even the doubtful Lagerstätten, show higher-than-average preservation quality.

Fossil occurrences were allocated to time bins. Time bins with a mean duration of 11 myr (standard deviation = 3.8 myr) were used, the same as used by Alroy *et al.* (2001) to achieve durations that are as similar as possible, and with a sufficient sample size per bin. Where a formation in reality spans several time bins, the formation occurrence was counted in all possible bins. In other cases, multiple bins are indicated because of uncertainty over dating in the PBDB, and here we assigned each collection an age between the upper and lower age limits randomly, using a bootstrapping approach over a uniform probability distribution. This process was repeated 1000 times to obtain mean occurrence counts with confidence intervals reflecting the gross uncertainty of dating the collections. Bootstrapped counts of collections, occurrences, and genera were produced in this way, for Lagerstätte and ordinary formations, and from marine and terrestrial lithofacies in each country. All calculations were carried out in R (R Core Team 2016).

Analysis was carried out at genus level to ensure enough data could be included and to manage some forms of uncertainty. Genera have been used as a proxy for species in previous studies (Sepkoski 1996; Dunhill *et al.* 2014a); however, genera are not defined biologically, may not be monophyletic, and include widely varying numbers of species (Hendricks *et al.* 2014). This must be weighed against the volume of data required for a study such as this one. PBDB data enterers have made an effort to include data at species level where possible, but this is not uniform, and especially so for incomplete specimens, which may be recorded simply to genus level. Further, for some taxa, the distinctions between species are disputed, and in any case, most of the vertebrate taxa we consider are monospecific, meaning that generic counts per bin are not so different from species counts.

Taxonomic richness for each country-level study was considered both with and without Lagerstätten. The UK and Germany should have similar patterns of outcrop, being geographically close. China, on the other hand, has been less thoroughly sampled for fossils (being larger in geographical area and having been sampled over a shorter period of history), and should lie at a lower point on a regional-scale 'collector curve'. Here we use collector curves as a measure of research history, and in particular the accumulation of novel taxa, as done in previous palaeontological studies (Tarver *et al.* 2007; Benton 2008, 2015).

*Fragile and robust taxa: a new metric*

If Lagerstätten form an important part of the total palaeodiversity record, they are assumed to achieve this by providing better sampling of fragile taxa than in 'ordinary' fossil beds. It has been claimed that the fossil records of some delicate and small taxa, namely birds (Brocklehurst *et al.* 2012), pterosaurs (Butler *et al.* 2013; Dean *et al.* 2016), bats (Brown *et al.* 2019), and lepidosaurs and lissamphibians (Fara 2004) are especially poor (the latter two with completeness metrics of <50%; Benton 1987), and that their fossil records are dominated by occurrences in Lagerstätten. To test this, occurrences of four of these 'fragile' vertebrate taxa (lissamphibians, lepidosaurs, pterosaurs, birds) in Lagerstätte formations and ordinary formations were counted, and the proportions of 'robust' and 'fragile' taxa in each comparable time bin were calculated. Note that these designations as 'robust' and 'fragile' are arbitrary and include aspects of body size and bone fragility; lissamphibians, squamates, pterosaurs, and birds are all generally smaller than turtles, crocodilians, dinosaurs, and mammals, may have hollow bones, and lack armour plates and horns. There are many exceptions of course, including tiny mammals and large squamates, such as mosasaurs.

To document the proportions of fragile to robust taxa in Lagerstätte and ordinary formations, we developed a new ratio. We could not simply record actual numbers in each category because the values are affected by phylogeny; clades of both types may originate, diversify or become extinct within any time bin. Therefore, the ratio for each time bin was normalized by the equivalent ratio for ordinary formations in the same bin. The ratio is given in Equation 1, where $N$ is number of taxa (species or genera).

$$r_f = \frac{\left(\frac{N_{\text{fragile}}}{N_{\text{robust}}}\right)_{\text{Lagerstatten}}}{\left(\frac{N_{\text{fragile}}}{N_{\text{robust}}}\right)_{\text{Ordinary}}} - 1 \qquad (1)$$

This equation is analogous to an isotope fractionation equation; the Lagerstätte formation ratio takes the place of the 'sample' and the ordinary formation ratio takes the place of the 'standard'. This ratio is calculated for each time bin and country with enough data; i.e. in time bins where fragile and robust taxa occur in both Lagerstätten and ordinary formations.

The fragile/robust fossil ratio indicates the extent to which the hard-to-preserve groups, such as Lissamphibia,

Lepidosauria, Pterosauria and Aves, are represented in Lagerstätte formations compared to ordinary formations. It is an indicator of how much of the richness of these groups has been lost in time bins containing no Lagerstätte formations. When $r_f = 0$, the ratio of fragile to robust taxa in Lagerstätte formations is the same as in ordinary formations, when $r_f > 0$, the ratio of fragile to robust taxa is greater in Lagerstätte formations than ordinary formations, and when $r_f < 0$, the ratio of fragile to robust taxa in Lagerstätten formations is smaller than in ordinary formations. Since fragile groups such as birds and pterosaurs are more commonly preserved in Lagerstätten, the expectation is that $r_f \gg 0$ for all time bins. For each country, a mean ratio is also calculated.

## Regression modelling

It has previously been suggested that the occurrence of Lagerstätten can be a driver of diversity counts (Butler *et al*. 2009; Benson & Butler 2011; Lloyd & Friedman 2013; Dean *et al*. 2016). This hypothesis is examined using linear regression models for each country. Previously, this effect has been corrected by coding the presence or absence of Lagerstätten in each time bin (Benson & Butler 2011; Dean *et al*. 2016). Every time bin containing Lagerstätten is given equal weight, regardless of the number of Lagerstätten contained within. To improve this, we use a count of the number of Lagerstätten in each time bin, rather than binary coding. For comparison we have, however, duplicated the modelling process using a binary presence/absence time series; these results are reported in Walker *et al*. (2019).

Model fit was assessed in two ways: $R^2$ values (best fit), which do not consider model simplicity, and AICc weights (Akaike information criterion corrected for small sample sizes; best fit, model simplicity and sample size), which do (Johnson & Omland 2004; R package `AICcmodavg`, Mazerolle 2016). $R^2$ values give an indication of the fit of each model to the data points, whereas AICc weights allow models to be ranked in relation to the alternatives but do not provide an absolute test of fit. Each model was tested for normality (R function `jarque.bera.test` of the package `tseries`, Trapletti & Hornik 2016), heteroscedasticity (R function `bptest` of the package `lmtest`, Zeileis & Hothorn 2002) in the residuals, and multicollinearity (R function `vif` in the package `car` v. 3.0-4; Fox & Weisberg 2018, 2019).

The tests for multicollinearity were used to determine whether different predictor variables (e.g. formation count, rock volume) are redundant with each other; for example, time bins with more fossil-bearing formations may also include more Lagerstätten if preserved rock volume from this time bin is high. Multicollinearity can be assessed by computing the variance inflation factor (VIF), which measures how much the variance of a regression coefficient is inflated due to multicollinearity in the model. The smallest possible value of VIF is 1 (absence of multicollinearity), and a VIF value that exceeds 5 or 10 indicates a problematic amount of collinearity (James *et al*. 2013).

We used regression models to test how well various sampling proxies predict genus richness. Regression models here were not used to produce estimates of 'true' diversity; they are used to obtain a greater understanding of the interplay between diversity and sampling. Regression models were calculated for each country. The German dataset contained both marine and terrestrial Lagerstätten, so the data could be considered for each; this was not possible for the data from the UK and China. The possible controlling variables included in the models were: time bin length, number of formations, number of collections, number of Lagerstätte formations in each bin, and the $r_f$ ratio; the last as a measure of how much the absence of Lagerstätten in a time bin might explain the absence of fragile taxa.

## True richness using a Poisson sampling model (TRiPS)

To assess the effects of Lagerstätten on palaeodiversity time series, we sought a method that would provide sampling-free estimates of total, ordinary, and Lagerstätten-based counts per time bin. The TRiPS method (True richness estimated using a Poisson sampling model) uses occurrence data and the duration of a time bin to estimate original richness (Starrfelt & Liow 2016). These authors claimed that the TRiPS method provides not simply a measure of relative incompleteness of sampling per bin, but generates an absolute, corrected palaeodiversity time series.

To apply the method, sampling is modelled as a time-homogeneous Poisson process. Based on taxon abundances, a single sampling rate for a particular time interval can be estimated using maximum likelihood. In theory, this sampling rate encompasses all aspects of sampling, from decay through to preservation, diagenesis and sampling by palaeontologists. The estimate in each time bin is independent of all other time bins. Starrfelt & Liow (2016) claimed that the TRiPS method produces a 'real' diversity estimate, unlike subsampling and sampling proxy methods, which only aim to produce relative diversity estimates and therefore need to be viewed in the context of diversity estimates in other time bins. Estimates of richness were obtained by: (1) using all data; and (2) including data only derived from Lagerstätten. This is an empirical examination of the Lagerstätten effect on TRiPS.

We calculated TRiPS for each of the 1000 bootstrapped runs accommodating for uncertainty in the dating of the collections in the dataset. The mean TRiPS score was taken, and the TRiPS uncertainty was interpreted as the mean of the confidence intervals from all runs.

## RESULTS

### Formation durations and environments

Average formation duration is remarkably similar in the UK (9.81 myr) and Germany (9.85 myr); in China it is nearly twice as long (17.08 myr). This difference does not arise from differing methods used to compile the formation duration information (geological survey data for the UK and Germany and PBDB data for China) because formation durations for the UK and Germany based on PBDB data (8.21 myr and 7.82 myr respectively) are similar to their respective geological survey durations.

Depositional environments within the study time interval (early Silurian to mid-Cenozoic) differ in each country. In the UK and Germany, they are mixed (UK formations: 58% marine, 31% terrestrial; German formations: 65% marine, 23% terrestrial), whereas the majority of Chinese formations that yield vertebrates are terrestrial (86% of the formations in this dataset). It could be argued that the high proportion of terrestrial Lagerstätten in China reflects a strong focus on dinosaurs; this cannot be tested, but there has been strong collecting pressure recently to add marine vertebrates, for example from the key Triassic marine biotas (Benton *et al.* 2013). The differing proportions probably partially or largely reflect the relative proportions of outcrop of Mesozoic rocks in each of the three countries.

### UK time series

In the UK, 8% of formations are Lagerstätte formations, but these provide 14% of collections and contain 19% of genera recorded (Table 1). 8% of genera occur in both Lagerstätte formations and ordinary formations, so 11% of genera are only found in Lagerstätten.

Most formations are marginal, mixed or marine (Walker *et al.* 2019). Terrestrial genus counts are highest in the Middle Triassic, Upper Jurassic and Lower Cretaceous, with the terrestrial formations in these intervals being particularly rich in taxa and yielding large proportions of the collections derived from the time bins they occupy. None of the purely terrestrial formations counts as a Lagerstätte formation (Fig. 1).

Marine Lagerstätte formations occur in the Upper Triassic, throughout the Jurassic, the middle Cretaceous, and the Palaeogene (Fig. 1). Marine Lagerstätte formations yield a large number of collections in the Upper Jurassic. There is little overlap in the genera occurring in marine Lagerstätte formations and ordinary formations, except in the Palaeogene. Lagerstätte formations account for most of the genus richness in the marine Jurassic.

**TABLE 1.** Numbers and proportions of formations, genera and collections from each formation category (Lagerstätte formations, ordinary formations and both).

| | | Total | From Lagerstätte formations | | From ordinary formations | | From both types of formation | |
|---|---|---|---|---|---|---|---|---|
| | | No. | No. | % | No. | % | No. | % |
| UK | Formations | 115 | 9 | 7.8 | 106 | 92.2 | – | – |
| | Collections | 3580 | 516 | 14.4 | 3064 | 85.6 | – | – |
| | Genera | 730 | 140 | 19.2 | 651 | 89.2 | 61 | 8.4 |
| Germany | Formations | 89 | 12 | 13.5 | 77 | 86.5 | – | – |
| | Collections | 1964 | 466 | 23.7 | 1498 | 76.3 | – | – |
| | Genera | 579 | 175 | 30.2 | 452 | 78.1 | 48 | 8.3 |
| German marine | Formations | 50 | 7 | 14.0 | 43 | 86.0 | – | – |
| | Collections | 952 | 268 | 28.2 | 684 | 71.8 | – | – |
| | Genera | 305 | 97 | 31.8 | 236 | 77.4 | 28 | 9.2 |
| German terrestrial | Formations | 28 | 2 | 7.1 | 26 | 92.9 | – | – |
| | Collections | 829 | 110 | 13.3 | 719 | 86.7 | – | – |
| | Genera | 310 | 56 | 18.1 | 267 | 86.1 | 13 | 4.2 |
| China | Formations | 158 | 12 | 7.6 | 146 | 92.4 | – | – |
| | Collections | 2064 | 543 | 26.3 | 1521 | 73.7 | – | – |
| | Genera | 792 | 263 | 33.2 | 553 | 69.8 | 24 | 3.0 |

**FIG. 1.** The rock record of the UK (A, B), Germany (C, D) and China (E, F) through geological time. Formations and collections are split by environment (marine, terrestrial and mixed) and by preservation (Lagerstätte and ordinary formations). Bars are stacked, the height of each coloured area representing the proportion (by number) of each formation type. Blank bars indicate periods of time with no formations in the dataset used. *Abbreviations*: C, Carboniferous; D, Devonian; J, Jurassic; K, Cretaceous; Ng, Neogene; P, Permian; Pg, Palaeogene; S, Silurian; Tr, Triassic.
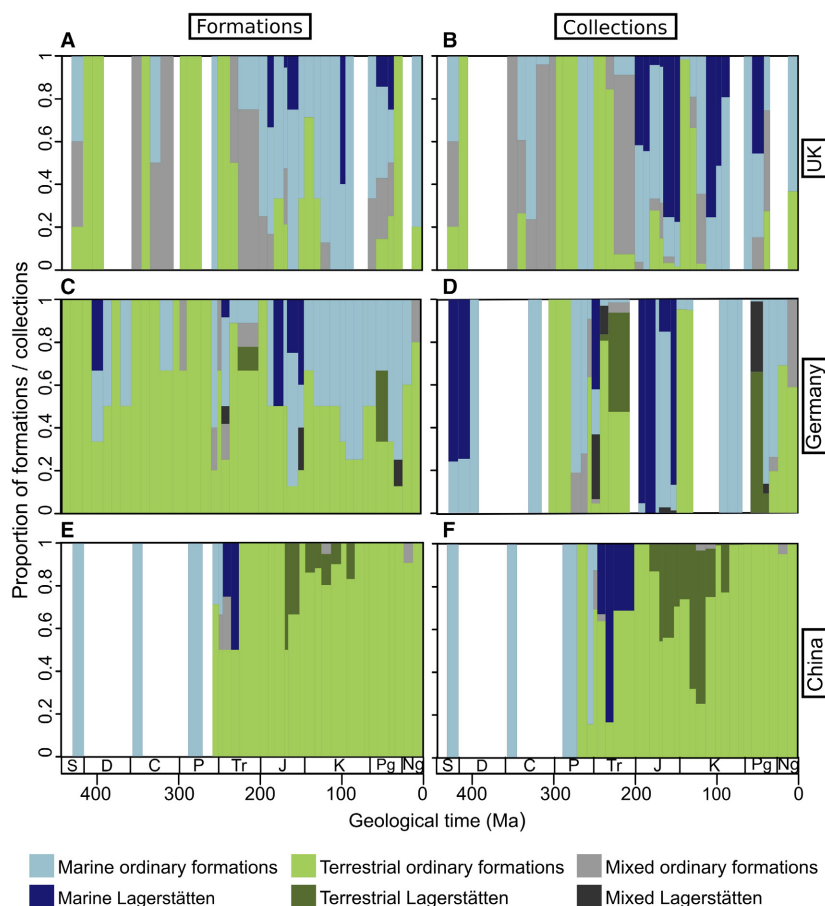


The Middle to Upper Jurassic has the most fossiliferous formations (Fig. 2B). Similarly, counts of genera and collections also peak in the Upper Jurassic, but other highs in genus and collection counts (Lower Cretaceous, Palaeogene and Neogene) do not correspond to large peaks in the number of formations (Fig 2A). There is little error introduced into the collection and genus curves as a result of uncertainty in age estimation (Walker *et al.* 2019). Lagerstätte formations contribute to the diversity peaks of the Upper Jurassic and Palaeogene, and account for all of the diversity in the Upper Cretaceous. There is little sharing of genera between Lagerstätte formations and ordinary formations.

*German time series*

In Germany, 13% of formations are Lagerstätte formations, but these provide 42% of collections and contain 30% of genera recorded (Table 1). 8% of genera occur in both Lagerstätte formations and ordinary formations, so 22% of genera are only found in Lagerstätte.

The rock record could be split into terrestrial and marine components (Fig. 1). Terrestrial sediments yielded the greatest number of collections and genera by far in the Middle–Upper Triassic, Lower Cretaceous and Neogene.

Marine Lagerstätte formations occur in the Lower Devonian, Lower Triassic, Middle–Upper Jurassic and Lower Cretaceous (Fig. 1). Marine Lagerstätte formations add greatly to the collection and genus counts in the Devonian, Triassic and Jurassic – Lower Cretaceous. There is little overlap between genera found in marine Lagerstätte formations and ordinary formations. Lagerstätte formations add most of the genus richness in the Devonian, Triassic and Jurassic time bins in the marine record of Germany.
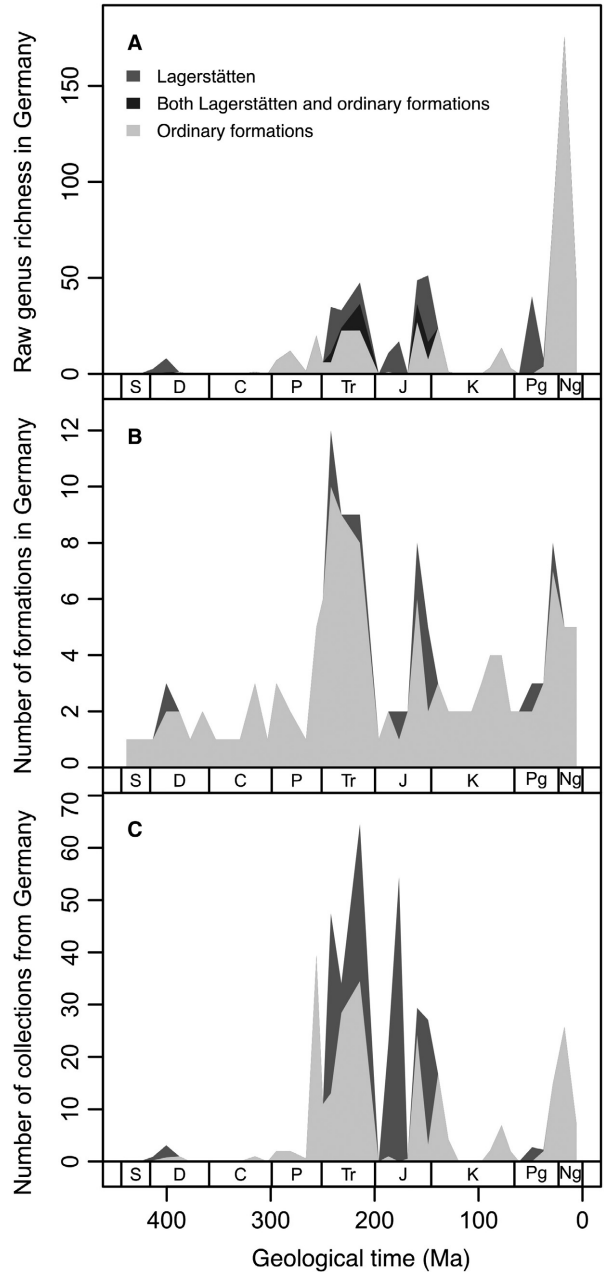
Terrestrial Lagerstätte formations occur in the Triassic and Palaeogene (Fig. 1). Deposits of these ages have yielded a large proportion of the collections and of genus richness within the respective time bins. Again, there is little overlap between genera occurring in Lagerstätte formations and ordinary formations, with some genera occurring in both types in only the Upper Triassic. There is little error introduced into the terrestrial counts in Germany, except for genus counts in the Palaeogene. There is little uncertainty introduced into specimen and genus counts as a result of age uncertainty, except in the Palaeogene. Lagerstätte formations add about half of the Triassic
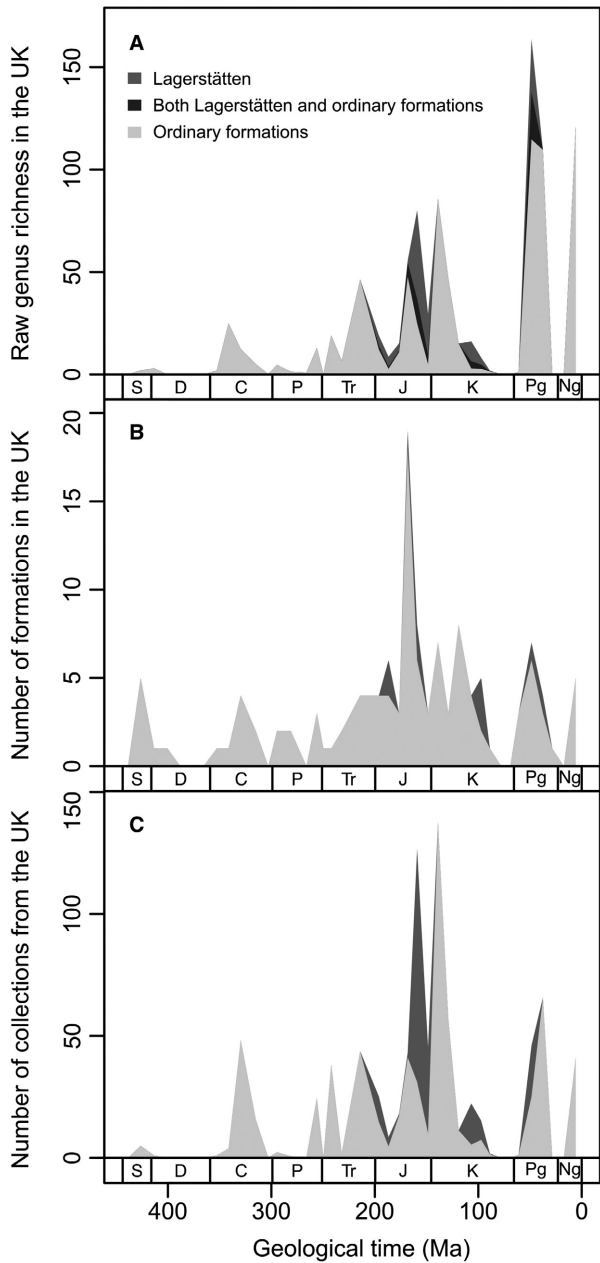
**FIG. 2.** Fossil vertebrate record in the UK. A, measured within-bin genus richness. B, formation count. C, collection count. *Abbreviations*: C, Carboniferous; D, Devonian; J, Jurassic; K, Cretaceous; Ng, Neogene; P, Permian; Pg, Palaeogene; S, Silurian; Tr, Triassic. Lines in this chart are stacked.

**FIG. 3.** Fossil vertebrate record in Germany. A, measured within-bin genus richness. B, formation count. C, collection count. *Abbreviations*: C, Carboniferous; D, Devonian; J, Jurassic; K, Cretaceous; Ng, Neogene; P, Permian; Pg, Palaeogene; S, Silurian; Tr, Triassic. Lines in this chart are stacked.

diversity in the terrestrial record, and account for most of the terrestrial Palaeogene richness.

The German rock record has the largest number of fossiliferous formations in the Triassic, with smaller peaks in the Upper Jurassic and Palaeogene–Neogene (Fig. 3). The number of collections through time largely follows this pattern. However, there are many collections in the Lower

Jurassic, all of which come from Lagerstätte formations. There is also a dearth of Devonian to Carboniferous material in the data, with most of the diversity in this period deriving from two collections of material from a single formation (the Lower Devonian Hunsrück Slate).

Generally, genus richness does not follow the same pattern as formations and collections; genus richness is

relatively low in the Palaeozoic and Mesozoic, with a large peak in the Neogene. There is little uncertainty introduced into the curves as a result of age estimate uncertainty, except in the genus counts in the Palaeogene, Devonian and Upper Jurassic. Lagerstätte formations make up a small proportion of the fossiliferous formations. Lagerstätte formations occur in the Devonian, Triassic, Jurassic, Palaeogene and Neogene. These few Lagerstätte formations have yielded large numbers of collections in the Triassic to Lower Cretaceous, and a large proportion of genus diversity in this period and in the Palaeogene. However, the large Neogene peak consists of data from formations that have not been defined as Lagerstätten. There is little sharing of genera between Lagerstätte formations and ordinary formations.

### Chinese time series

Only 7% of formations are Lagerstätte formations, but these provide 26% of collections and contain 32% of genera. Only 3% of genera occur in both Lagerstätten and ordinary formations, so 29% of genera from China are only found in Lagerstätten.

The fossil record is almost entirely made up of terrestrially-deposited formations, with the exception of examples in the Silurian, Permian–Triassic and Lower Cretaceous (Fig. 1). This contrasts with the mostly marine record of the UK and the evenly split record of Germany. Because continental deposits dominate, the terrestrial patterns are similar to the patterns seen for all formations. There are relatively small numbers of terrestrial Lagerstätte formations in the Lower Triassic, Upper Jurassic and Lower Cretaceous, but these are entirely responsible for producing the peaks in genus richness in these periods. There are only three marine Lagerstätten in this dataset.

The Chinese geological record has a large number of fossiliferous formations in the Lower Cretaceous, and in the mid-Palaeogene (Fig. 4B). Peaks and troughs in the number of collections through time match peaks and troughs in genus diversity.

### Fragile and robust taxa

The groups defined here as fragile taxa comprise a varying proportion of the diversity in each time bin and each country (see Walker *et al.* 2019). Lissamphibia, Lepidosauria and Pterosauria originated in the Triassic, and occur sporadically from the Triassic to Neogene in the UK. In Germany, fragile taxa occur in the Jurassic, Palaeogene and Neogene. In China, fragile taxa make up most of the diversity in the Lower Cretaceous, and this diversity is largely composed of birds.
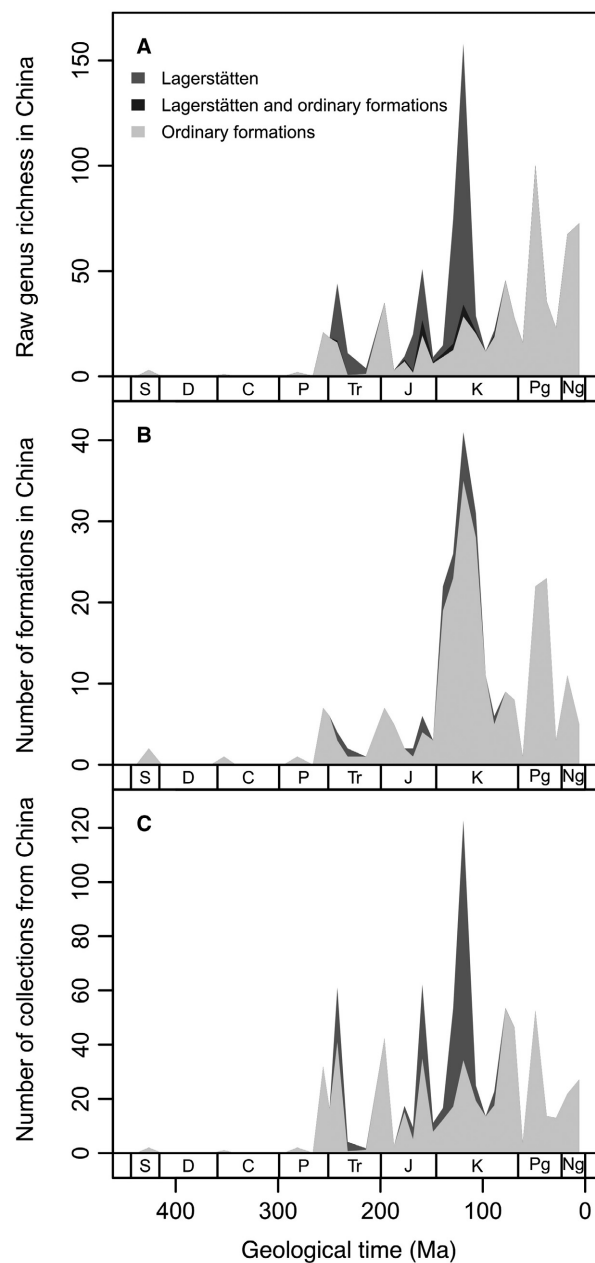


**FIG. 4.** Fossil vertebrate record in China. A, measured within-bin genus richness. B, formation count. C, collection count. *Abbreviations*: C, Carboniferous; D, Devonian; J, Jurassic; K, Cretaceous; Ng, Neogene; P, Permian; Pg, Palaeogene; S, Silurian; Tr, Triassic. Lines in this chart are stacked.

Table 2 lists the mean '$r_f$' values (Eqn 1) for the UK, Germany and China. As expected, there are no negative values of $r_f$; there are fewer fragile taxa in ordinary formations than Lagerstätte Formations. This difference is smallest in Germany, with a ratio of 1.42, but this is based on a very small sample size of time bins. The large standard deviations in the overall $r_f$ values for the UK and China result from high $r_f$ values in a single time bin

**TABLE 2.** Average $r_f$ ratios across all time bins for each country.

|  | UK | Germany | China | Total |
|---|---|---|---|---|
| Mean | 6.15 | 1.42 | 25.36 | 13.70 |
| Standard deviation | 9.86 | 0.92 | 59.12 | 38.65 |
| No. of comparable bins | 6 | 2 | 6 | 14 |

for each country (in fact, the Coniacian in both countries, arising from the very small sample sizes of only one or two specimens). If we tried to calculate $r_f$ values for individual clades, the standard deviations would probably become larger because of the patchiness of the data.

### Model fits

The best three model fits for each country, and the marine and terrestrial of Germany, are listed in Table 3 and Walker et al. (2019). In the UK, the model using the number of collections and specimens is the best for predicting raw richness (Table 3; AICc weight = 0.346, of 63 models). The second-best model includes the number of collections and specimens, plus formations. In Germany, the best model overall includes the number of specimens, collections, formations and Lagerstätten (Table 3; AICc weight = 0.158, of 63 models). When the German data are split into marine and terrestrial, marine data are best predicted by the number of collections and specimens (Table 3; AICc weight = 0.294, of 32 models), with the second-best model using only the

number of specimens. Terrestrial German data are best predicted by the model including the number of collections, specimens, formations and Lagerstätten (Table 3; AICc weight = 0.779, of 31 models), although within this model the collections time series shows a high level of multicollinearity (VIF = 5.1; Walker et al. 2019). Chinese measured richness is best predicted by the number of specimens and formations (Table 3; AICc weight = 0.271, of 63 models). In other Chinese models with less weight, specimens and collections when used together in a model display a high level of multicollinearity (VIF > 5), suggesting that they are redundant with each other (Walker et al. 2019), as might have been predicted (Benton 2015).
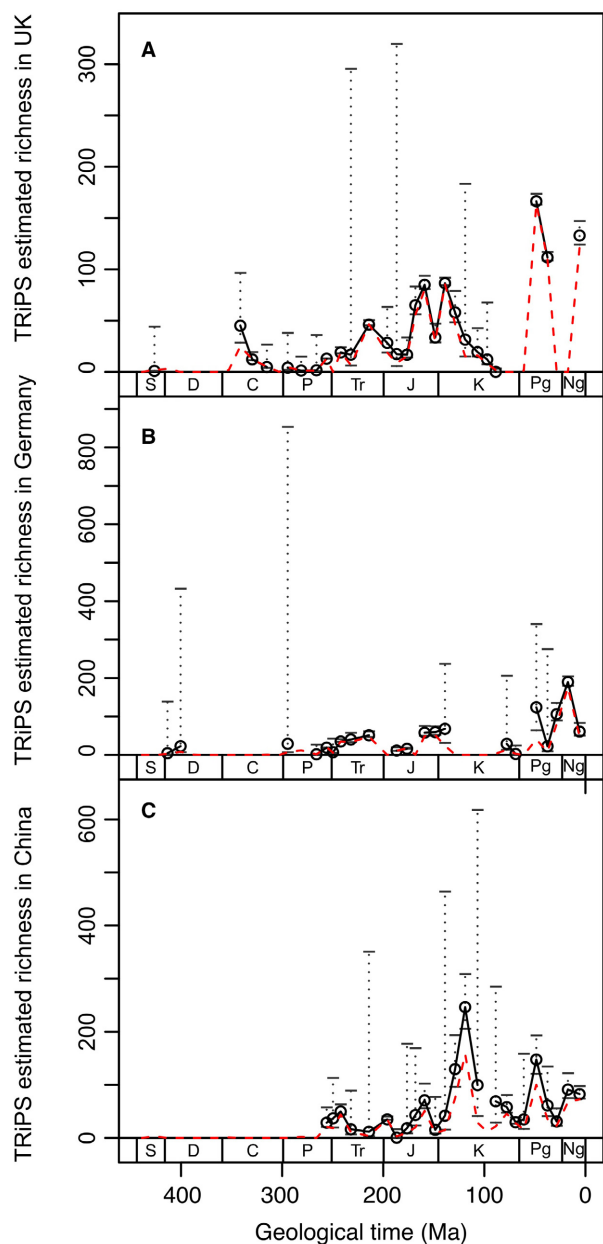
### TRiPS estimates

Generic richnesses estimated from TRiPS (Fig. 5) broadly follow the same patterns and are rarely much higher than raw measured richnesses in all three countries. The error bars in Figure 5 show that there are significant departures from the raw richness calculations in some time bins (e.g. Palaeozoic to Permian of Germany), which also have the largest error bars.

TRiPS estimates for Lagerstätten are no more similar to raw richness measurements than the estimates from the whole datasets (the $R^2$ values in Fig. 6A–C are within the same range as the same values for Fig. 6D–F). The Lagerstätten calculations also have large error bars which, in some time bins, are greater than the range of the estimated time series (Fig. 5).

**TABLE 3.** The three best-fitting models that explain aspects of the fossil genus richness palaeodiversity time series for each country, with the German record shown in total and subdivided into marine and terrestrial categories.

|  | Model | Model ranking | $R^2$ | AICc | AICc weight |
|---|---|---|---|---|---|
| UK | Cols + Specs | 1 | 0.947 | 296.03 | 0.346 |
|  | Cols + Specs + Forms | 2 | 0.947 | 298.06 | 0.125 |
|  | Cols + Specs + Lags | 3 | 0.947 | 298.58 | 0.097 |
| Germany | Cols + Specs + Forms + Lags | 1 | 0.936 | 295.16 | 0.158 |
|  | Cols + Specs + Lags | 2 | 0.930 | 295.44 | 0.137 |
|  | Cols + Specs + Forms | 3 | 0.930 | 295.54 | 0.130 |
| German marine | Cols + Specs | 1 | 0.824 | 288.74 | 0.294 |
|  | Specs | 2 | 0.807 | 289.98 | 0.158 |
|  | Cols + Specs + Forms | 3 | 0.826 | 291.08 | 0.091 |
| German terrestrial | Cols + Specs + Forms + Lags | 1 | 0.985 | 216.92 | 0.779 |
|  | Bins + Cols + Specs + Forms + Lags | 2 | 0.985 | 219.63 | 0.200 |
|  | Cols + Specs + Forms | 3 | 0.980 | 225.27 | 0.012 |
| China | Specs + Forms | 1 | 0.963 | 269.77 | 0.271 |
|  | Cols + Specs + Forms | 2 | 0.965 | 270.70 | 0.170 |
|  | Specs + Forms + $r_f$ | 3 | 0.963 | 272.08 | 0.085 |

Rankings are based on AICc weights. *Abbreviations*: Bins, time bin length; Cols, number of collections; Forms, number of geological formations; Lags, number of Lagerstätten formations; $r_f$, fragile taxon ratio (defined in the text); Specs, number of specimens. The fragile taxon ratio has only been included in the whole-country models because subsets of German marine and terrestrial data have insufficient sample size; however, note that this ratio only appears in China's top three models.

**FIG. 5.** Raw measured and TRiPS-estimated richness for: A, the UK; B, Germany; C, China. Raw richness is indicated by the dashed line, and TRiPS richness estimates with a solid line and error bars. *Abbreviations*: C, Carboniferous; D, Devonian; J, Jurassic; K, Cretaceous; Ng, Neogene; P, Permian; Pg, Palaeogene; S, Silurian; Tr, Triassic. Colour online.

## DISCUSSION

### *Formation duration*

It is remarkable that the average formation duration in China is nearly twice as long as average formation durations in the UK and Germany. This is unexpected for the

methodological reason that the Chinese duration data are derived from individual collection dates from the PBDB only, not from entire formation duration data as for Germany and the UK. There are two possible explanations, that either: (1) the concept of a formation encompasses less time and fewer sediments in the UK and Germany, when compared with China; or (2) the dating of collection records in PBDB is less precise than the dating of formations by geological surveys in Europe. The second suggestion has been rejected (see Results) as mean formation durations are the same in the PBDB and from independent survey data from the UK and Germany. Therefore, the first explanation is probably correct, that this is a historical aspect of the ways in which survey geologists worked in their respective countries, reflecting the huge size of China compared to the UK and Germany, and that exposure in China is often better than in northern Europe, so correlations can be done directly over longer distances, thus making some formations geographically and temporally huge (Benton *et al.* 2011). Further, of course, geological surveys in Europe began in the 1830s, whereas the China Geological Survey was refounded in 1999 (Jackson 2018), thus providing less time for revision and subdivision.

### *Which taxa are preserved in Lagerstätten?*

We have noted the problem of establishing a sharp definition of Lagerstätten, but Lagerstätte type might also affect how taxa co-occur in ordinary fossil beds. A conservation Lagerstätte can preserve a range of soft tissues and soft-bodied organisms, and so is likely to expand raw diversity counts, for example by 50% for pterosaurs (Dean *et al.* 2016) and by a similar amount for marine reptiles (Benson & Butler 2011). On the other hand, a concentration Lagerstätte such as a bonebed can preserve a large amount of mineralized material, but much of it may represent taxa already known from ordinary fossil beds.

Lagerstätten distribution is variable in time and space; for example, conservation Lagerstätten are particularly common in the Cambrian and Jurassic (Allison & Briggs 1993; Muscente *et al.* 2017). This variation arises from real changes in environment, making exceptional preservation more likely in some regions and time periods. As regions and taxa are sampled, the accumulation of fossiliferous formations follows a collector curve. In the early stages of investigation (e.g. in the nineteenth century in the UK), new discoveries were frequent. As collecting within a region matures, fewer new taxa are discovered, as common taxa have already been found (Benton 2008, 2015). Lagerstätten are likely to be discovered relatively early in the cycle as their fossils are abundant and spectacular. We cannot provide definitive evidence about the
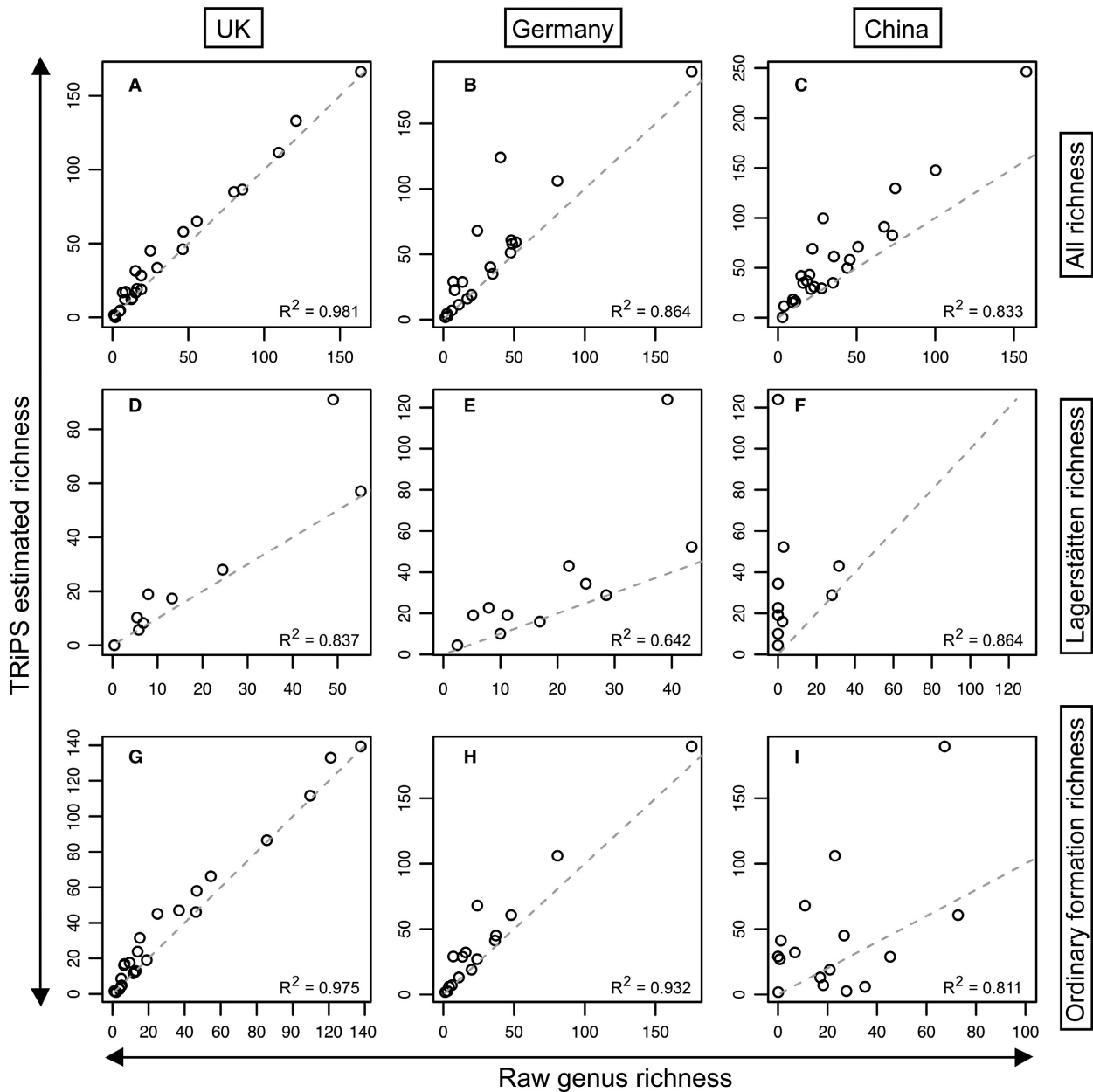
**FIG. 6.** Raw richness plotted against TRiPS estimates for the UK (A, D, G), Germany (B, E, H) and China (C, F, I). Measurements and estimates for all formations (A–C), Lagerstätte formations (D–F) and ordinary formations (G–I) are shown. $R^2$ values for each linear relationship are shown.

accumulation of Lagerstätten early and late in the collector curve, but there have been a number of recent discoveries, such as a new Burgess Shale site in Canada (Caron *et al.* 2014) and one in China (Fu *et al.* 2019), and the Tanis site in the Hell Creek Formation (DePalma *et al.* 2019). Whether these count as rare late chance finds, or whether Lagerstätten can continue to be discovered even after centuries of search, has to be quantified. Thus, the influence of Lagerstätten on diversity counts may depend on the current point along the collector curve.

In our study, for all three countries, and as expected, 'fragile' taxa make up a greater proportion of taxa in Lagerstätte formations than in ordinary formations in the same time bins. The proportions in Table 2 suggest that 'fragile' taxa are on average nearly 14 times more likely to be preserved in Lagerstätten than in 'ordinary' formations. This high figure is skewed by the Chinese data, which have a ratio of 25.4, much higher than the German ratio of 1.4.

It might be expected that Lagerstätten would preserve broadly the same taxa as in less informative deposits of

similar age and facies, but with the addition of some less easily preserved taxa. However, we find there is little overlap in the genera preserved in Lagerstätte formations and ordinary formations (Table 1), even in Germany, where the $r_f$ ratio is low, and therefore with fewer fragile taxa unique to Lagerstätten. It could be that some time bins have few Lagerstätten, so this would be expected, but this does not seem to be the case. When the time series data from each country are considered, the overlap between genera preserved in Lagerstätten and in ordinary formations remains small even in the time bins containing multiple Lagerstätten, and when Lagerstätten contain large proportions of diversity. There are three possible explanations for this:

1. Lagerstätten preserve unusual ecosystems that show little environmental and taxonomic overlap with 'ordinary' ecosystems. Examples include the Plattenkalks of Germany with their unusually high representation of flying taxa (Wellnhofer 2008), and some bird-rich locations in the Jehol Group of China (Benton *et al.* 2008).
2. Large, robust remains are excluded from Lagerstätten by a taphonomic filter. In regular depositional systems, energy decreases downstream, and larger carcasses might be dumped before the site of exceptional preservation, such as a Jehol lake, which would then include only indigenous lake-dwelling fishes and arthropods as well as smaller washed-in carcasses. A further example could be the Triassic–Jurassic Bristol fissures, which contain bonebed concentrations in which grain size is regular and small, and larger organisms or carcasses were either not washed into the cave systems because the cave mouths were too small, or dropped out upstream as current energy diminished (Whiteside *et al.* 2016).
3. Taxa that are common elsewhere are ignored when sampling Lagerstätten. Perhaps, in some cases, because of special interest or high prices, collectors select only the rare or soft-bodied fossils, a special case of the 'bonanza effect' of Raup (1977). Therefore, museum collections might feature the birds, mammals or trilobites at the expense of plants or brachiopods. This could be true of some Jehol Group localities, where collectors have specialized in recovering rare bird fossils. However, with time, palaeontologists make more controlled samples (Benton 2015), where every fossil is recovered and catalogued, and in the case of the PBDB, all taxa in collections are recorded, not just the rare ones.

In summary, it is certainly true that many Lagerstätten preserve unusual environments (1) or are subject to sedimentary and environmental filters (2), but collection-level datasets based on museum collections may reflect the bonanza effect (3).

*Differences between countries*

The impact of the Lagerstätten effect varies among the three countries. Germany has the highest proportion of Lagerstätte formations making up the rock record (Table 1; 14%) but a similar proportion of diversity from Lagerstätte formations as in China (30% and 33% respectively). The UK has few Lagerstätte formations (8% of formations), and few genera from such units (19%). China has few Lagerstätten formations (7%), but a large proportion of genera from Lagerstätten (32%). From this, in terms of genus count, Lagerstätten have yielded less diversity in the UK record than that of Germany, but in China, a large proportion of diversity derives from exceptionally preserved deposits.

The palaeodiversity curves show a number of Lagerstätten and other peaks. For the UK (Fig. 2A), only two of the 11 peaks (18%) can be ascribed to Lagerstätten. Others, such as the Middle Triassic (Anisian) peak, correspond to ordinary fossiliferous formations; here the widespread Helsby Sandstone Formation and equivalents. This unit is also likely to be well sampled due to wide exposure through quarrying for building stone (Dunhill *et al.* 2013). The Late Jurassic peaks represent the fossiliferous Oxford Clay and Kimmeridge Clay formations, which yield abundant marine reptiles and other fossils, although their identification as Lagerstätten might be queried by some. Although largely clay-dominated and so not naturally exposed extensively, these units were massively quarried for brick-making and so many fossils were collected (Dunhill *et al.* 2014*b*). The Early Cretaceous peak corresponds to the Wealden sandstones and mudstones, sources of abundant fossils, but not Lagerstätten as there is little or no soft tissue preservation. The same is true of some of the Palaeogene units such as the London Clay, with abundant fossils, which are heavily studied despite limited exposure (Dunhill *et al.* 2014*a*).

In Germany (Fig. 3A), four of the 11 peaks (36%) can be explained by Lagerstätten. These are: the Triassic peak represented largely the Muschelkalk, with rich finds in several formations over wide areas, but not all classed as Lagerstätten; the Early Jurassic peak, largely the Holzmaden Lagerstätte; the Late Jurassic peak, largely the Solnhofen Lagerstätte; and the Palaeogene peak, the Messel Lagerstätte. However, the very high diversity Neogene peak is composed of non-Lagerstätte formations.

In China (Fig. 4A), Lagerstätten peaks are also significant, explaining four of the 11 peaks (36%). These include the Middle Triassic (Panxian, Luoping, Guanling), the Late Jurassic (Yanliao) and Early Cretaceous (Jehol). The relative importance of Lagerstätten to the fossil record of China may have arisen because Germany and the UK have been more thoroughly studied than China; this could be explicitly tested by examining the collector curve for each country. Alternatively, terrestrial

Lagerstätte formations could contain a greater number of taxa than marine Lagerstätte formations. However, when German units are split into marine and terrestrial, a greater proportion of genera are sampled from Lagerstätte formations in marine environments (32%) than terrestrial ones (18%). Another explanation may be the geological and geographical characteristics of each country; China is large, and Chinese Lagerstätten such as the Middle Triassic Guanling Formation and the Early Cretaceous Jehol Group cover enormous areas, and so have yielded huge numbers of fossils from hundreds of localities each.

### Model fitting

Lagerstätte count appears as an independent variable in two of the three best German models, but in only the third-best model from the UK, and in none of the top three Chinese models. In a localized study of fishes in Great Britain, Lloyd & Friedman (2013) found that the presence of Lagerstätten in time bins was included in one of their three top models, but the best global marine reptile models of Benson & Butler (2011) all included the presence of Lagerstätten. These previous results suggest that the effect of Lagerstätten on taxic richness is variable, and the results presented here suggest a regional effect, with Lagerstätten exerting more control over diversity in Germany than in the UK and China. The results for China contradict conclusions that may be drawn from Table 1 (that a large proportion of genera come from Lagerstätten) probably because data are skewed by the high Lagerstätte genus richness count in the Lower Cretaceous.

For the German data that has been split into marine and terrestrial facies, Lagerstätte count is included in the top two terrestrial models but in none of the top three marine models. This suggests a greater Lagerstätte control in German terrestrial than marine environments.

Regression modelling has been used frequently in seeking to identify which predictor variables best explain the response variable (palaeodiversity), but we confirm other recent studies (Benton *et al.* 2011; Dunhill *et al.* 2014*a*, 2018; Benton 2015) here, that the method can be problematic. Input variables are often related, and here for several of the Chinese models, specimens and collections are redundant with each other. Additionally, the differences in AICc values between best and near-best models are often small, suggesting that, for the total German data, for example, there is little to choose between the top two or three models.

### Estimating true diversity

In all three countries, the TRiPS estimates of richness are remarkably similar to the raw richness measurements.

Additionally, in many of the time bins, the TRiPS method produces very large error bars. There are two possible explanations for this: (1) in all three countries, sampling does not change the gross palaeodiversity pattern, and only minimally reduces the total richness which is recorded in fossiliferous sediments; or (2) it is not the case that TRiPS can correct for sampling in real datasets. This inadequacy was also suggested by Close *et al.* (2018), who assessed different estimators and extrapolators of the fossil record, including TRiPS.

Lagerstätten, when compared with units with non-exceptional preservation, may provide a more accurate record of past biodiversity for the ecosystems they preserve than 'ordinary' deposits. Mineral replacement allows for less recalcitrant tissues to be preserved, a greater variety of taxa are fossilized, and preservation potential is higher in these deposits (Muscente *et al.* 2017). Additionally, palaeontological research effort is often concentrated on exceptional deposits, as 'systematists follow the fossils' (Raup 1977). Elevated preservation potential and sampling effort combine to produce the Lagerstätten effect. Because exceptional deposits have higher preservation potential and are more thoroughly sampled, it might be that raw richness measured from Lagerstätten-only data would be more similar to TRiPS estimates than the all-formations data.

Contrary to expectation, TRiPS calculations based on all data and the Lagerstätten-only data are both equally similar to their respective raw richnesses ($R^2$ values in Fig. 6). This could be because the Lagerstätten data dominate the total data sets in some way, or because the TRiPS method is more dependent on the input data, and less able to predict missing data, than has been assumed (Close *et al.* 2018).

For the TRiPS method to have utility, and to meet the claims of its authors (Starrfelt & Liow 2016), it should yield a richness count that is close to the truth, compensating for heterogeneous sampling. This is difficult to achieve because the fossil record is so patchy, with gaps in time bins that lack Lagerstätten, or even lack Lagerstätten from multiple sedimentary environments. The findings here that TRiPS values mirror raw richness counts and that the error bars produced by the method span a range greater than the time bin values (as found by Starrfelt & Liow 2016) both call into question the utility of this form of TRiPS in real palaeontological datasets.

## CONCLUSIONS

Not all of the dramatic peaks seen in the richness curves here are produced by Lagerstätten, although some regional patterns emerge: (1) a small number of Lagerstätten yield a large proportion of diversity in China; (2) in Germany,

exceptional preservation is more widely spread through the fossil record, yielding a similarly large proportion of genus diversity; (3) in the UK, a small number of Lagerstätten yield a moderate proportion of genus diversity. The fossil records of the UK and Germany (in particular the fossil records of fragile taxa in these countries) are less dominated by Lagerstätten than that of China.

The extent to which fragile taxa are preserved in Lagerstätten also varies among the three countries; the fragile taxa occurrence per robust taxon rate ($r_f$ ratio) in Chinese Lagerstätten is much higher than the UK or German equivalents. This indicates that fragile taxa are hugely undersampled in ordinary formations, or, conversely, that robust taxa are underrepresented in Lagerstätten in China; this could be further tested. We infer that some Lagerstätten truly over-sample fragile taxa because they preserve unusual ecosystems where, for example, birds or pterosaurs are especially represented; numbers of such taxa are high from China because of the great scale of some Lagerstätten deposits such as those in the Jehol Group.

It is unlikely that the TRiPS method can provide a true estimate of past richness. TRiPS produces diversity curves that are remarkably similar to the empirical curves. In exceptional deposits this result might be unsurprising, but this also occurs when using data from deposits which have not been termed 'Lagerstätten', and so the correction factors do not seem to differentiate cases where we have independent evidence for good or poor sampling.

## DATA ARCHIVING STATEMENT

Data for this study are available in the Dryad Digital Repository: https://doi.org/10.5061/dryad.f4114kq

*Editor.* Philip Mannion

## REFERENCES

ALLISON, P. A. 1988. Konservat-Lagerstätten: cause and classification. *Paleobiology*, **14**, 331–344.

—— and BRIGGS, D. E. G. 1993. Exceptional fossil record: distribution of soft-tissue preservation through the Phanerozoic. *Geology*, **21**, 527–530.

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

—— MARSHALL, C. R., BAMBACH, R. K., BEZUSKO, K., FOOTE, M., FÜRSICH, F. T., HANSEN, T. A., HOLLAND, S. M., IVANY, L. C., JABLONSKI, D., JACOBS, D. K., JONES, D. C., KOSNIK, M. A., LIDGARD, S., LOW, S., MILLER, A. I., NOVACK-GOTTSHALL, P. M., OLSZEWSKI, T. D., PATZKOWSKY, M. E., RAUP, D. M., ROY, K., SEPKOSKI, J. J., SOMMERS, M. G., WAGNER, P. J. and WEBBERE, A. 2001. Effects of sampling standardization on estimates of marine diversification. *Proceedings of the National Academy of Sciences*, **98**, 6261–6266.

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 BUTLER, R. J. 2011. Uncovering the diversification history of marine tetrapods: ecology influences the effect of geological sampling biases. 191–208. *In* McGOWAN, A. J. and SMITH, A. B. (eds). *Comparing the geological and fossil records: Implications for biodiversity studies*. Geological Society, London, Special Publications, **358**.

—— —— 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.

—— BUTLER, R. J., ALROY, J., MANNION, P. D., CARRANO, M. T. and LLOYD, G. T. 2016. Near-stasis in the long-term diversification of Mesozoic tetrapods. *PLoS Biology*, **14**, e1002359.

BENTON, M. J. 1987. Mass extinctions among families of non-marine tetrapods: the data. *Memoires de la Société Géologique de France*, **150**, 21–32.

—— 2008. How to find a dinosaur, and the role of synonymy in biodiversity studies. *Paleobiology*, **34**, 516–533.

—— 2015. Palaeodiversity and formation counts: redundancy or bias? *Palaeontology*, **58**, 1003–1029.

—— TVERDOKHLEBOV, V. P. and SURKOV, M. V. 2004. Ecosystem remodelling among vertebrates at the Permian–Triassic boundary in Russia. *Nature*, **432**, 97–100.

—— ZHOU, Z., ORR, P. J., ZHANG, F. and KEARNS, S. L. 2008. The remarkable fossils from the Early Cretaceous Jehol Biota of China and how they have changed our knowledge of Mesozoic life. *Proceedings of the Geologists' Association*, **119**, 209–228.

—— DUNHILL, A. M., LLOYD, G. T. and MARX, F. G. 2011. Assessing the quality of the fossil record: insights from vertebrates. 63–94. *In* McGOWAN, A. J. and SMITH, A. B. (eds). *Comparing the geological and fossil records: Implications for biodiversity studies*. Geological Society, London, Special Publications, **358**.

—— RUTA, M., DUNHILL, A. M. and SAKAMOTO, M. 2013. The first half of tetrapod evolution, sampling proxies, and fossil record quality. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 18–41.

BRIGGS, D. E. G. 1995. Experimental taphonomy. *Palaios*, **10**, 539–550.

—— 2003. The role of decay and mineralization in the preservation of soft-bodied fossils. *Annual Review of Earth & Planetary Sciences*, **31**, 275–301.

BRITISH GEOLOGICAL SURVEY. 2016. Lexicon of Named Rock Units. http://www.bgs.ac.uk/Lexicon/. [Accessed June 2016]

BROCKLEHURST, N. 2015. A simulation-based examination of residual diversity estimates as a method of correcting for sampling bias. *Palaeontologica Electronica*, **18**, 1–15.

—— UPCHURCH, P., MANNION, P. D. and O'CON-NOR, J. 2012. The completeness of the fossil record of Mesozoic birds: implications for early avian evolution. *PLoS One*, **7** (6), e39056.

—— KAMMERER, F. and FRÖBISCH, J. 2013. The early evolution of synapsids, and the influence of sampling on their fossil record. *Paleobiology*, **23**, 470–490.

BROWN, E. E., CASHMORE, D. D., SIMMONS, N. B. and BUTLER, R. J. 2019. Quantifying the completeness of the bat fossil record. *Palaeontology*, **62**, 757–776.

BUTLER, R. J., BARRETT, P. M., NOWBATH, S. and UPCHURCH, P. 2009. Estimating the effects of sampling biases on pterosaur diversity patterns: implications for hypotheses of bird/pterosaur competitive replacement. *Paleobiology*, **35**, 432–446.

—— BENSON, R. B. J. and BARRETT, P. M. 2013. Pterosaur diversity: untangling the influence of sampling biases, Lagerstätten, and genuine biodiversity signals. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 78–87.

CARON, J. B., GAINES, R. R., ARIA, C., MÁNGANO, M. G. and STRENG, M. 2014. A new phyllopod bed-like assemblage from the Burgess Shale of the Canadian Rockies. *Nature Communications*, **5**, 3210.

CLEARY, T. B., BENSON, R. B. J., UPCHURCH, P. and BUTLER, R. J. 2018. Lepidosaurian diversity in the Mesozoic–Palaeogene: the potential roles of sampling biases and environmental drivers. *Royal Society Open Science*, **5**, 171830.

CLOSE, R. A., BENSON, R. B. J., UPCHURCH, P. and BUTLER, R. J. 2017. Controlling for the species-area effect supports constrained long-term Mesozoic terrestrial vertebrate diversification. *Nature Communications*, **8**, e15381.

—— 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 & Evolution*, **9**, 1386–1400.

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.

DEAN, C. D., MANNION, P. D. and BUTLER, R. J. 2016. Preservational bias controls the fossil record of pterosaurs. *Palaeontology*, **59**, 225–247.

DEPALMA, R. A., SMIT, J., BURNHAM, D. A., KUIPER, K., MANNING, P. L., OLEINIK, A., LARSON, P., MAURRASSE, F. J., VELLEKOOP, J., RICHARDS, M. A. and GURCHE, L. 2019. A seismically induced onshore surge deposit at the KPg boundary, North Dakota. *Proceedings of the National Academy of Sciences*, **116**, 8190–8199.

DEUTSCHE STRATIGRAPHISCHE KOMMISSION. 2016. LithoLex. https://litholex.bgr.de/ [accessed March 2016] [in German]

DRISCOLL, D. A., DUNHILL, A. M., STUBBS, T. L. and BENTON, M. J. 2019. The mosasaur fossil record through the lens of fossil completeness. *Palaeontology*, **62**, 51–75.

DUNHILL, A. M. 2011. Using remote sensing and a geographic information system to quantify rock exposure area in England and Wales: implications for paleodiversity studies. *Geology*, **39**, 111–114.

—— 2012. Problems with using rock outcrop area as a paleontological sampling proxy: rock outcrop and exposure area compared with coastal proximity, topography, land use, and lithology. *Paleobiology*, **38**, 126–143.

—— 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**, 1155–1157.

—— —— 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.

—— HANNISDAL, B. and BENTON, M. J. 2014a. Disentangling rock record bias and common-cause from redundancy in the British fossil record. *Nature Communications*, **5**, e4818.

—— BENTON, M. J., TWITCHETT, R. J. and NEWELL, A. J. 2014b. Testing the fossil record: sampling proxies and scaling in the British Triassic-Jurassic. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **404**, 1–11.

—— 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.

FARA, E. 2002. Sea-level variations and the quality of the continental fossil record. *Journal of the Geological Society*, **159**, 489–491.

—— 2004. Estimating minimum global species diversity for groups with a poor fossil record: a case study of Late Jurassic-Eocene lissamphibians. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **207**, 59–82.

FOUNTAINE, T. M. R., BENTON, M. J., DYKE, G. J. and NUDDS, R. L. 2005. The quality of the fossil record of Mesozoic birds. *Proceedings of the Royal Society B*, **272**, 289–294.

FOX, J. and WEISBERG, S. 2018. *An R companion to applied regression*. 3rd edn. Sage, 576 pp.

—— —— 2019. Functions to Accompany J. Fox and S. Weisberg, An R Companion to Applied Regression, Third Edition, Sage, in press. R package v. 3.0-4. https://cran.r-project.org/web/packages/car/index.html

FRÖBISCH, J. 2013. Vertebrate diversity across the end-Permian mass extinction - separating biological and geological signals. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 50–61.

FU, D., TONG, G., DAI, T., LIU, W., YANG, Y., ZHANG, Y., CUI, L., LI, L., YUN, H., WU, Y. and SUN, A. 2019. The Qingjiang biota—A Burgess Shale-type fossil Lagerstätte from the early Cambrian of South China. *Science*, **363**, 1338–1342.

HANNISDAL, B. and PETERS, S. E. 2011. Phanerozoic Earth system evolution and marine biodiversity. *Science*, **334**, 1121–1124.

HENDRICKS, J. R., SAUPE, E. E., MYERS, C. E., HERMSEN, E. J. and ALLMON, W. D. 2014. The generification of the fossil record. *Paleobiology*, **40**, 511–528.

JACKSON, L. E. 2018. National geological surveys: the past, present and future. *Earth*, **63**, 8–9.

JAMES, G., WITTEN, D., HASTIE, T. and TIBSHIRANI, R. 2013. *An introduction to statistical learning: with applications in R*. Springer, 430 pp.

JOHNSON, J. B. and OMLAND, K. S. 2004. Model selection in ecology and evolution. *Trends in Ecology & Evolution*, **19**, 101–108.

LLOYD, G. T. and FRIEDMAN, M. 2013. A survey of palaeontological sampling biases in fishes based on the Phanerozoic record of Great Britain. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 5–17.

MANNION, P. D., BENSON, R. B. J. and BUTLER, R. J. 2013. Vertebrate palaeobiodiversity patterns and the impact of sampling bias. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 1–4.

MAYHEW, P. J., JENKINS, G. B. and BENTON, T. G. 2008. A long-term association between global temperature and biodiversity, origination and extinction in the fossil record. *Proceedings of the Royal Society B*, **275**, 47–53.

MAZEROLLE, M. J. 2016. AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). R package version 2.0-4. https://cran.r-project.org/package=AICcmodavg

MENNING, M. 2002. Stratigraphic Table of Germany 2002 (STG 2002). http://www.stratigraphie.de/std/ [accessed 22 February 2016] [in German]

MUSCENTE, A. D., SCHIFFBAUER, J. D., BROCE, J., LAFLAMME, M., O'DONNELL, K., BOAG, T. H., MEYER, M., HAWKINS, A. D., HUNTLEY, J. W., McNAMARA, M. and MACKENZIE, L. A. 2017. Exceptionally preserved fossil assemblages through geologic time and space. *Gondwana Research*, **48**, 164–188.

R CORE TEAM. 2016. R: a language and environment for statistical computing. v. 3.3.2. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

RAUP, D. M. 1972. Taxonomic diversity during the Phanerozoic. *Science*, **177**, 1065–1071.

—— 1976. Species diversity in the Phanerozoic: an interpretation. *Paleobiology*, **2**, 289–297.

—— 1977. Systematists follow the fossils. *Paleobiology*, **3**, 328–329.

SAKAMOTO, M., VENDITTI, C. and BENTON, M. J. 2016. "Residual diversity estimates" do not correct for sampling bias in palaeodiversity data. *Methods in Ecology & Evolution*, **2**, 453–459.

SANSOM, R. S. 2014. Experimental decay of soft tissues. 259–274. *In* LAFLAMME, M., SCHIFFBAUER, J. D. and DARROCH, S. A. F. (eds). *Reading and writing of the fossil record: Preservational pathways to exceptional fossilization*. The Paleontological Society Papers, **20**.

SEILACHER, A. 1970. Begriff und Bedeutung der fossil Lagerstätten. *Neues Jahrbuch für Geologie und Paläontologie, Monatshefte*, **1970**, 34–39.

—— REIF, W.-E., WESTPHAL, F., RIDING, R., CLARKSON, E. N. K. and WHITTINGTON, H. B. 1985. Sedimentological, ecological and temporal patterns of fossil Lagerstätten. *Philosophical Transactions of the Royal Society B*, **311**, 5–24.

SELDEN, P. and NUDDS, J. 2012. *Evolution of fossil ecosystems*, 2nd edn. Mansion Publishing, London.

SEPKOSKI, J. J. Jr 1996. Patterns of Phanerozoic extinction: a perspective from global data bases. 35–51. *In* WALLISER, O. H. (ed.) *Global events and event stratigraphy in the Phanerozoic*. Springer.

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.

—— 2007. Marine diversity through the Phanerozoic: problems and prospects. *Journal of the Geological Society*, **164**, 731–745.

—— 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**, 1–10.

—— —— 2011. The ties linking rock and fossil records and why they are important for palaeobiodiversity studies. 1–7. *In* McGOWAN, A. J. and SMITH, A. B. (eds.) *Comparing the geological and fossil records: Implications for biodiversity studies*. Geological Society, London, Special Publications, **358**.

STARRFELT, J. and LIOW, L. H. 2016. How many dinosaur species were there? True richness estimated using a poisson sampling model (TRiPS). *Philosophical Transactions of the Royal Society B*, **371**, 20150219.

TARVER, J. E., BRADDY, S. J. and BENTON, M. J. 2007. The effects of sampling bias on Palaeozoic faunas and implications for macroevolutionary studies. *Palaeontology*, **50**, 177–184.

TIPPER, J. C. 1979. Rarefaction and rarefiction—the use and abuse of a method in paleoecology. *Paleobiology*, **5**, 423–434.

TRAPLETTI, A. and HORNIK, K. 2016. tseries: Time series analysis and computational finance. R package v. 0.10-35. https://cran.r-project.org/web/packages/tseries/index.html

WALKER, F. M., DUNHILL, A. M., WOODS, M. A., NEWELL, A. J. and BENTON, M. J. 2017. Assessing sampling of the fossil record in a geographically and stratigraphically constrained dataset: the Chalk Group of Hampshire, Southern UK. *Journal of the Geological Society*, **107**, 509–521.

—— —— and BENTON M. J. 2019. Data from: Variable preservation potential and richness in the fossil record of vertebrates. *Dryad Digital Repository*. https://doi.org/10.5061/dryad.f4114kq

WELLNHOFER, P. 2008. A short history of pterosaur research. *Zitteliana Reihe B*, **28**, 7–20.

WHITESIDE, D. I., DUFFIN, C. J., GILL, P. G., MARSHALL, J. E. A. and BENTON, M. J. 2016. The Late Triassic and Early Jurassic fissure faunas from Bristol and South Wales: stratigraphy and setting. *Palaeontologia Polonica*, **67**, 257–287.

WILLIAMS, M., BENTON, M. J. and ROSS, A. 2015. The Strawberry Bank Lagerstätte reveals insights into Early Jurassic life. *Journal of the Geological Society*, **175**, 683–692.

ZEILEIS, A. and HOTHORN, T. 2002. Diagnostic checking in regression relationships. *R News*, **2**, 7–10.