

Geological Society, London, Special Publications

## Quantifying the deep-sea rock and fossil record bias using coccolithophores

Graeme T. Lloyd, Andrew B. Smith and Jeremy R. Young

*Geological Society, London, Special Publications* 2011; v. 358; p. 167-177  
doi: 10.1144/SP358.11

---

### Email alerting service

click [here](#) to receive free e-mail alerts when new articles cite this article

### Permission request

click [here](#) to seek permission to re-use all or part of this article

### Subscribe

click [here](#) to subscribe to Geological Society, London, Special Publications or the Lyell Collection

---

### Notes

Downloaded by guest on December 16, 2011

---

# Quantifying the deep-sea rock and fossil record bias using coccolithophores

GRAEME T. LLOYD\*, ANDREW B. SMITH & JEREMY R. YOUNG

*Natural History Museum, Cromwell Road, London SW7 5BD, UK*

*\*Corresponding author (e-mail: graemelloyd@gmail.com)*

**Abstract:** While many studies show a correlation between observed taxonomic richness and various measures of geological sampling, all have been based on the same record of terrestrial and marine sediments collected from the land. Here we present the first analyses of how rock and fossil records vary in the deep-sea. We have developed a novel database of species occurrences of coccolithophores sampled during major drilling programs of the North Atlantic, including the Mediterranean and Caribbean. Our sampling proxy, the number of deep-sea sites sampled – perhaps the most direct measure of sampling used so far – shows an exponential rise towards the Recent. Over the same period species-richness has grown in an approximately linear fashion, but genus-level richness shows a sharp initial increase followed by a much slower decline. However, correlations between both richness measures and sampling are extremely strong and a model assuming true diversity to be constant accurately predicts much of observed richness. We conclude that the deep-sea fossil record, like its land-based counterpart, bears a rock record bias.

Fossils collected and recorded from sedimentary rocks provide the empirical evidence from which the history of life over geological time is reconstructed. However, we also know that the sedimentary rock record we can access on land does not represent a time series with uniform sampling opportunity; both the areal extent and environmental representation of rocks at outcrop vary in a non-trivial way, even over relatively short time intervals. Major sea-level cycles (for example) have driven the ratio of terrestrial to marine rocks that are preserved and accessible to palaeontologists at outcrop (Smith 2001). As a result we now know that the classic diversity curves derived from raw counts of fossils through time (e.g. Sepkoski *et al.* 1981; Benton 1995) are a product of the complex interplay between original biological diversity and sampling effort. Recent work to standardize Phanerozoic diversity curves for sampling effort (Alroy *et al.* 2001, 2008; Alroy 2010) indicates that such corrections can dramatically modify their shape, suggesting that sampling bias in the fossil record is non-negligible.

Over the last 10 years there has been significant effort first to quantify this bias in the rock record and then to compare the quality of the rock record against sampled fossil diversity curves (Peters & Foote 2001, 2002; Smith 2001; Crampton *et al.* 2003; Smith & McGowan 2007; McGowan & Smith 2008; Barrett *et al.* 2009; Butler *et al.* 2009; Marx 2009; Wall *et al.* 2009; Benson *et al.* 2010; Mannion *et al.* 2011). While these studies have

demonstrated that there is a strong correlation between the quality of the rock and fossil records, the significance of this correlation remains problematic (Smith 2007; Peters 2008; Butler *et al.* in press; Mannion *et al.* 2011; Upchurch *et al.* 2011; Benson & Butler 2011). This is because these studies have used proxies for sampling effort rather than measuring sampling effort itself, and the factors that drive the quality of the rock record also drive the fossil sampling potential. Thus sea-level change may be driving biological diversity on the land and in shallow epicontinental shelf seas (through a species-area effect), while at the same time altering availability of rock at outcrop. This is the ‘common cause’ hypothesis of Peters (2005). Alternatively, parts of the sampled diversity curve might be recording little more than sampling effort. To truly test the effect of sampling effort on recorded diversity we need to turn to a record where there is not such a close coupling between the area of original habitat occupied and area of preserved rock outcrop remaining. To do this we turn to the deep-sea rock and microfossil records.

Pelagic sediments accumulating in the deep-sea are composed of the calcareous or siliceous skeletons of microscopic plankton that lived and died in the overlying surface waters. These microplankton achieve extremely wide geographical distribution within ocean basins (Winter *et al.* 1994; Ziveri *et al.* 2004) and a few grams of rock sample can yield 1000s of individual microfossils for analysis. Furthermore, in some places the rock record

created can be near complete for relatively long geological time spans (Wang *et al.* 2003; Ebra *et al.* 2010). Not surprisingly therefore the deep-sea microfossil record is considered by some to be amongst the best we have (Ebra 2004; McGowran 2005; Suchéras-Marx *et al.* 2010).

Accessing the deep-sea rock and fossil records is, of course, difficult, and it is largely through the Deep Sea (DSDP; <http://www.deepseadrilling.org/>) and Ocean Drilling Program (ODP; <http://www-odp.tamu.edu/>) that knowledge of the age distribution and fossil content of deposits in the deep-sea has been acquired. Such limited access has advantages as well as disadvantages. While the volume of rock recovered is tiny compared to the available land-based record, the numbers of cores drilled and the number of samples from which microfossils have been recovered from each core retrieved allow an accurate measure of sampling effort to be quantified. In contrast to the land-based record, therefore, we are able to measure both sampling effort and taxonomic diversity directly.

There is less reason to expect a tight correspondence between recorded taxonomic diversity and the amount of rock sampled from the deep-sea for particular time intervals. This is partly because groups such as the coccolithophores are effectively cosmopolitan in their distribution (Winter *et al.* 1994; Ziveri *et al.* 2004) and thus, unlike most shelf or terrestrial taxa, in theory a global record can be accessed from a single core. Highest biomass productivity (and hence bioclastic sediment delivery to the deep-sea) is associated with low species-richness counts, whereas highest species-richness occurs at intermediate levels of phytoplankton biomass (Irigoien *et al.* 2005). Although locally sections can be astonishingly complete, the global basin record of pelagic sediments in the deep-sea can be patchy and far from perfect, and in many sections hiatuses abound (Spencer-Cervato 1998). Variation in the carbonate compensation depth (CCD; Murray & Renard 1891) over time is a major factor in controlling whether calcareous oozes accumulate at specific sites and at specific times in the geological past, and the CCD is controlled by temperature and CO<sub>2</sub> concentration of the water (Gomitz 2009). So while the abundance of microplankton in surface waters might be expected to affect the quantity of skeletal debris being delivered to the sea floor far below, no strong link is expected connecting taxonomic diversity and extent of rock record sampling through time. The deep-sea rock and fossil records thus offer an ideal opportunity to test the strength of correlation between sampling intensity and recorded diversity over geological time.

Strategies to answer questions about how the rock and fossil records compare have evolved

through time. The earliest comparisons (e.g. Raup 1972, 1976; Peters & Foote 2001; Smith 2001) involved the plotting of independently created databases of taxonomic diversity and sedimentary rock through time. More recently work has shifted to creating new, more sophisticated databases for rock measures, such as North American gap-bound packages (Peters 2006) or Western European maps (Smith & McGowan 2007). However, in order to make direct comparisons it is preferable to collect the rock and fossil data together, as many vertebrate workers have done (Frobisch 2008; Barrett *et al.* 2009; Butler *et al.* in press; Mannion *et al.* 2011; Benson & Butler 2011; Upchurch *et al.* 2011), or assign the fossils to specific rock packages *a posteriori* (Heim & Peters 2011). Here we adopt the former approach and create a completely novel database that houses information on both fossil occurrences and lithological information with all data coming from the published records of the DSDP and ODP. We use this data to test the correlation between deep-sea sampling effort and the taxonomic diversity of coccolithophorids recorded from Atlantic sites spanning the last 150 Ma.

## Material and methods

### *The database*

As it was not tractable to enter data from the world's oceans as a whole we limited ourselves to the North Atlantic, which we define as 90°N to 20°S and including both the Mediterranean and Caribbean. This area offers the key advantage of being relatively densely sampled by the DSDP/ODP, compared to say, the Pacific. However, our database is explicitly designed to be easily expandable in future and already incorporates the basic information on all DSDP/ODP holes (e.g. leg number, site number, depth below sea-level and latitude–longitude).

The DSDP/ODP volumes present data on several microfossil groups, but we chose to limit ourselves to the two major calcareous planktonic groups, the coccolithophores and planktic foraminifera. These are commonly occurring, frequently recorded and sufficiently abundant and speciose to be appropriate for palaeobiodiversity studies. In this paper we shall only deal with the coccolithophores.

Unlike the pre-existing NEPTUNE database (Lazarus 1994) our fundamental unit is not an individual sample, but a biozone within a specific DSDP or ODP hole. This decision was made to greatly reduce the amount of data entry required, in order to allow more sites to be entered. Here a biozone is either a nannofossil or planktic foraminifera zone. Our dates come from Gradstein *et al.*

(2004), and specifically the TimeScale Creator program (<https://engineering.purdue.edu/Stratigraphy/tscreator/>). In practice some complications arise from using zones. An initial problem is slumping which can lead to a zone occurring more than once in the same DSDP/ODP hole. This was solved by additionally defining a unit by its top and bottom as a depth in metres below sea floor (mbsf), thus allowing a zone to occur twice in the same hole. In other cases precise allocation to a single zone is not possible and instead a range is given. This is accommodated in the database by having separate fields for oldest and youngest possible zone. Finally, in some cases both nannofossil and planktic foraminiferal zones are available. Where possible, these were used to split up the units more finely based on areas of overlap that ultimately lead to more precise dating (see methods, below).

Once defined stratigraphically, other data can then be assigned to that unit. Most important amongst this data are the taxonomic occurrences that come from the distribution charts in the scientific results of the DSDP/ODP volumes. Apart from the species name, additional data recorded include the taxon's highest abundance within that unit (if recorded) and whether or not the occurrence is considered questionable. The abundance and preservation quality for the unit as a whole is also recorded where given and is based on the best sample in the unit. This is done for coccolithophores and planktic foraminifera separately, as are counts of the total number of samples within the unit as well as the number that are fossiliferous. Additional data assigned to a unit include: the presence of other taxa, including reworked or indeterminate coccolithophores/planktic foraminifera, the reference(s) from which the data came, the taxonomist(s) responsible for the species occurrence data and lithological data, including the presence of glauconite. Reference data is linked to a separate bibliography file created in BibTeX (<http://www.bibtex.org/>).

Data entry proceeded with a mixture of manual entry and parsing from online resources. Several pre-existing web depositories house digitized versions of DSDP/ODP distribution charts, including NEPTUNE (<http://paleodb.org/cgi-bin/bridge.pl?a=displayDownloadNeptuneForm>), JANUS (<http://www-odp.tamu.edu/database/>) and ODSN (<http://www.odsn.de/>). These were parsed into the correct format for our database using custom written R (version 2.11.1; R Development Core Team 2010) code and checked for potential errors. Error checking was done by using a conservative range-through database provided to us by Paul Bown (pers. comm., 2010), the results of which are published in Bown *et al.* (2004). We flagged

up all occurrences for shared taxa that definitively lie outside of the range of the Bown *et al.* (2004) data. We then consulted the original DSDP/ODP scientific results to check if the data was entered correctly (the range extension thus being considered 'real') or not. Using this procedure we found an alarmingly high error rate (*c.* 18% of all occurrences outside the Bown *et al.* 2004 range were not accurate representations of the original DSDP/ODP data). Some of these were simple misdatings of a unit and a handful were questionable occurrences that hadn't been flagged as such. However, the majority (*c.* 15%) were completely erroneous, bearing no resemblance to the published distribution chart. All errors were overwhelmingly concentrated in the NEPTUNE data and likely reflect the lengthy and complex history of this database. However, presently there is an effort underway to overhaul NEPTUNE (Lazarus 2011) and we have passed on our findings to them. For our database all such errors were corrected, in many cases leading to whole charts or units being re-entered.

At present the database is implemented in Microsoft Access with the data used in the analyses here coming from three separate queries: (1) a table of all species occurrences, (2) a table of units entered and, (3) a list of taxa and their statuses (see below). At present data entry is still underway, but we hope to ultimately migrate the database to MySQL and make it freely available online.

### Analytical methods

*Taxonomic standardization.* Before analysing our data we standardize our taxonomy using a new list of valid, invalid and synonymized taxa originally based on the NEPTUNE database, but significantly overhauled by one of the authors (JRY). In the process of manual data entry we have additionally uncovered many names not included in the NEPTUNE list making our global nannofossil synonymy list the most comprehensive and up-to-date presently available. This list is stored in the main database, allowing data entry to proceed using the original names from the distribution charts. This procedure thus allows for a different future taxonomy to still be used should opinions on synonymy etc. change. For data analysis we adopt the following procedure: (1) synonyms are replaced with their senior counterparts, (2) any resulting duplicate occurrences are removed and, (3) invalid taxa, questionable occurrences, taxa whose status is presently considered unknown and cf. or aff. taxa are removed.

*Creating time bins of equal duration and calculating error bars.* Units are given numerical dates based on Gradstein *et al.* (2004) and TimeScale

Creator as follows. If only a nannofossil or planktic foraminiferal zonation is known then the top of the youngest and bottom of the oldest are used. If both zone types are present then the dates of the overlap are used, conferring greater precision. In some cases, however, the two zonations do not overlap (implying uncertainty). When this happens then the maximum possible age range is used. Finally, if the uncertainty between the maximum and minimum possible dates is large (>15 million years) then we remove that unit and its constituent taxa from the analyses.

As we are interested in counts of species-richness through time an appropriate time binning approach is required. However, nannofossil or foraminiferal zones are problematic to use as they vary considerably in length and are thus likely to give misleading results, with more taxa likely to accumulate in a longer bin than a shorter one. This problem was identified by Sepkoski & Koch (1996) who recommended using time bins of roughly equal length. Alroy *et al.* (2008) adopted such an approach by combining geological stages to get roughly 11 million-year time bins. Although this is appropriate for an overview of Phanerozoic macrofossils, or poorly time-constrained taxa such as dinosaurs (Lloyd *et al.* 2008) such coarse binning is unnecessary for the data used here. Instead we adopt the Alroy *et al.* (2008) approach, but combine biozones, instead of geological stages, to make roughly 3 million-year time bins. We made an additional modification to this approach however, which is to enforce the inclusion of major geological boundaries (the Jurassic/Cretaceous, Cretaceous/Palaeogene, Eocene/Oligocene and Palaeogene/Neogene). This is because these are often associated with major turnover events and a bin spanning such a boundary is thus likely to have artificially inflated diversity because of an extinction and recovery fauna being time-averaged together.

Even after clumping zones together it is inevitable that some units will lack the precise dating required to assign them to a single time bin. Previous workers have had diametrically opposed solutions to this quandry. For example, Alroy *et al.* (2008) simply ignore taxa that cannot be assigned to a single bin and don't count them. By contrast, vertebrate workers have tended to treat uncertainty instead as the range of a taxon, counting it in all bins it could *possibly* be in (e.g. Benton 1995; see Upchurch & Barrett 2005 for a justification of this approach). Here we regard both solutions to be somewhat extreme and prefer instead a method that is intended to better quantify this uncertainty. Firstly we assume that each unit really does belong to a single bin and assign it based on a randomization approach. This is done by picking a random number from a uniform distribution between the

oldest and youngest possible dates for the unit. We then assign the unit to a time bin based on this single date and perform all of the counts outlined below. We then repeat this procedure 1000 times and record the resulting mean and 95% confidence intervals for our counts. When plotted it can thus be clearly seen whether a rise or fall between successive bins is likely to be empirically real or within dating error. It is this procedure that is used to create the error bars on the graphs presented here and the mean values that are used in the tests below.

*Picking a sampling proxy in the deep-sea.* Here we use the number of DSDP/ODP sites that have yielded sediments dated to a specific time bin as a measure of sampling. Sites can be considered a good measure of sampling, as they are decided on *a priori* by the researchers on the DSDP/ODP leg. We use sites rather than holes, as although multiple holes are typically drilled at the same site appearing to represent additional sampling, these are actually just additional attempts to recover from a particular horizon. Furthermore there is zero redundancy in the database: there are no cases of entered core being drilled from the same depth at the same site. Consequently, our within bin sampling measure is the number of sites recording fossil-bearing rock of that age.

*Correlation tests.* In comparing our sampling measure against taxonomic richness we employ a Spearman rank test. These are performed for both the raw data (long-term correlation) and detrended data (short-term correlation). Here we use two different detrending methods, a first differences and a 5-bin moving average (following Smith & McGowan 2007). In both cases raw values were logged for plotting purposes, but this does not affect rank-based correlation. Here and elsewhere when zero-values were encountered these were treated as non-applicable, essentially removing them from the analysis. This was done to avoid the problem of a value of infinity being returned, is not expected to affect the correlations, and only occurs in the late Early Jurassic to early Middle Jurassic and once in the Upper Jurassic (9 out of 66 time-bins).

*Modelling.* Assuming that sampling is a major factor in producing observed taxonomic richness curves an interesting follow up question is: how much of the observed richness is unexplained by sampling? Smith & McGowan (2007) introduced a procedure to tackle this question that starts from the notion that sampling perfectly predicts observed richness. In other words, the smallest sample is matched up with the lowest observed richness, the second smallest with the second lowest and so on. A linear

model is then fitted to this new data from which a function can be derived that allows us to predict the richness for a given sample. Subtracting these predicted values from those observed gives residuals that show either higher than expected (positive) or lower than expected (negative) richness. To further test if any residuals were significantly different than expected Barrett *et al.* (2009) simply looked for points outside two standard deviations of the mean.

Here we extend the method of Smith & McGowan (2007) and Barrett *et al.* (2009) as follows. Firstly, we consider more than just a linear model: we also fit logarithmic, exponential, hyperbolic, sigmoidal and polynomial models. The best model is then chosen using the sample size corrected Akaike information Criterion (AIC; Akaike 1973), the  $AIC_c$  (Johnson & Omland 2004). The  $AIC_c$  weighs both a model's fit (a close fit being best) and its complexity (a simple model being best). Second, we use the standard error and standard deviation from the model-fitting process instead of the standard deviation of the residuals (after subtracting the modelled estimate from the observed values) to assess statistically significant diversions. For further details see Lloyd (in press).

**Subsampling.** An alternative way to remove a sampling bias from a species-richness curve is to rarefy or subsample (e.g. Alroy *et al.* 2001, 2008). For each time bin a list is compiled by combining the species occurrences from each site that is represented. This is the occurrence list that the subsampling is based on and is different to the full inventory as we do not allow an occurrence to be counted more than once simply because multiple units from a site are present in that bin. An additional problem arises from the fact stated earlier that some units span more than one time bin, complicating the question of whether a unit (and its constituent taxa) is represented in that time bin. Previous workers (e.g. Alroy *et al.* 2001, 2008) have simply ignored such occurrences, removing them from the analysis. We dub this a 'minimum' approach. An alternative is to multiply this unit so that it occurs in every possible bin, a 'maximum' approach. Here we perform both in order to ascertain what difference it makes.

Having compiled these two sets of occurrence lists subsampling proceeded in the usual manner (see rarefaction by occurrences in Bush *et al.* 2004). Here 1000 replicates are performed in order to get a mean and 95% confidence interval that make up a rarefaction curve (samples v. species-richness). Diversity curves were then produced by recording the species-richness when the number of samples taken is equal to the bin with the fewest samples. We then removed the bin with the fewest

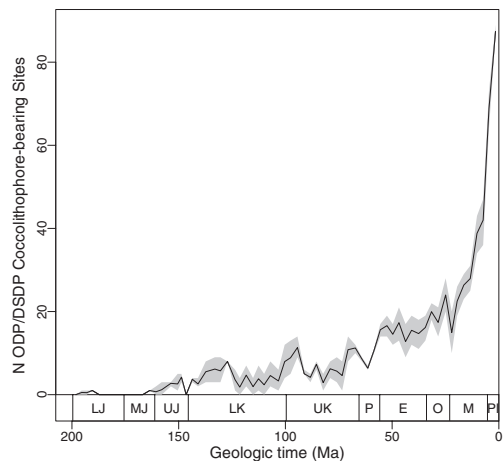
samples and repeated the process for the bin with the next fewest and so on until only a single bin (with the most samples) remains. The same procedure was applied to both the minimum and maximum lists.

All analyses were performed by importing the three SQL queries outlined above into the freely-available statistical programming package R 2.11.1 (R Development Core Team 2010). Custom-written code is available on request from one of the authors (GTL).

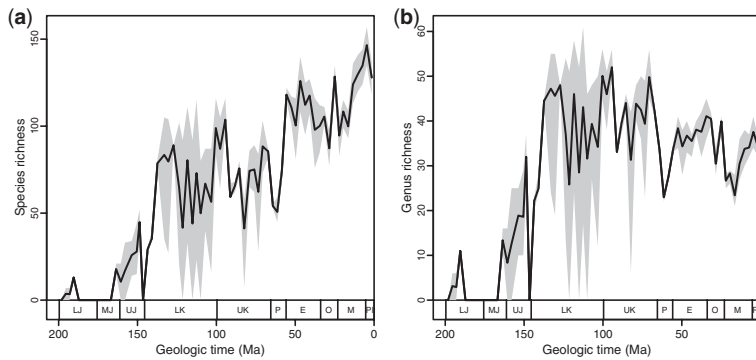
## Results

### *Empirical pattern*

Figure 1 shows our sampling proxy, number of ODP and DSDP sites, through time. Low Mesozoic values rise slightly going into the Palaeogene and then exponentially in the Neogene, with shorter-term fluctuations also apparent. Figure 2a shows species-richness through time. Intermittent Jurassic preservation is replaced by a sharp rise going into the Early Cretaceous followed by a period of large fluctuations before a latest Cretaceous drop. Richness rises sharply in the Palaeocene and then follows a plateau before rising again in the Miocene before a final Plio-Pleistocene drop. Short-term fluctuations are also evident. Genus-level richness, shown in Figure 2b, shows a different pattern. Again there is a sharp rise going into the Early Cretaceous, but this is something of a peak, with



**Fig. 1.** Sampling proxy over time – the number of DSDP and ODP sites from which coccolithophore-bearing rock in a c. 3 Myr time bin have been recovered. Grey polygon shows 95% confidence interval based on 1000 iterations where uncertainly dated units are assigned at random to a time bin.



**Fig. 2.** Sampled coccolithophore richness through time: (a) species and (b) genera. Grey polygon shows 95% confidence interval as in Figure 1.

later richness slowly dropping off, with notable troughs in the Early Palaeocene and Early Miocene. Short-term fluctuations are again evident.

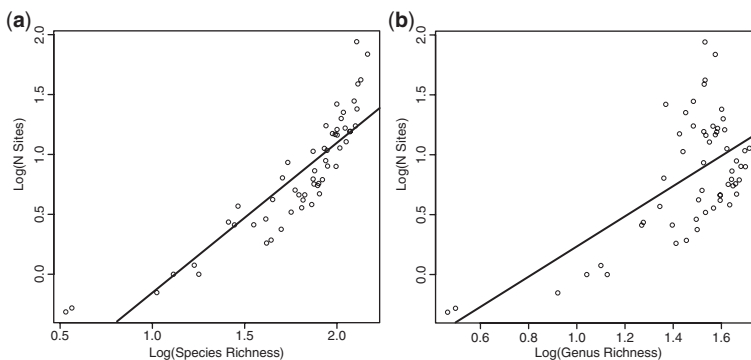
### Correlation

Figure 3 shows the raw correlation between the number of sites and: (a) species-richness (Spearman  $\rho = 0.94$ ,  $P < 0.001$ ) and, (b) genus-level richness (Spearman  $\rho = 0.37$ ,  $P = 0.005$ ). Figure 4a, b shows the first differences – the absolute rise or fall over time – of number of sites (blue line) and: (a) species-richness (yellow line; Spearman  $\rho = 0.87$ ,  $P < 0.001$ ) and, (b) genus-level richness (black line; Spearman  $\rho = 0.84$ ,  $P < 0.001$ ). Figure 4c, d shows differences from a five-bin moving average for number of sites (blue line) and: (c) species-richness (yellow line; Spearman  $\rho = 0.90$ ,  $P < 0.001$ ) and, (d) genus-level richness (yellow line; Spearman  $\rho = 0.87$ ,  $P < 0.001$ ). Overall correlation is very high – very unlikely to be due to chance

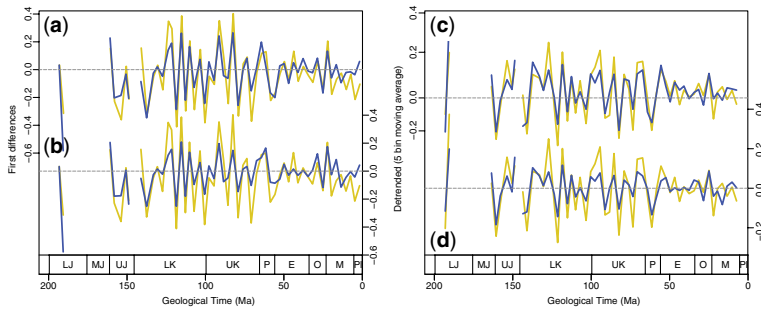
alone – and consistent regardless of taxonomic level or detrending method. The only exception here is raw genus-level richness, which although much less strongly correlated with sampling than species-richness, is still statistically significant.

### Corrected pattern

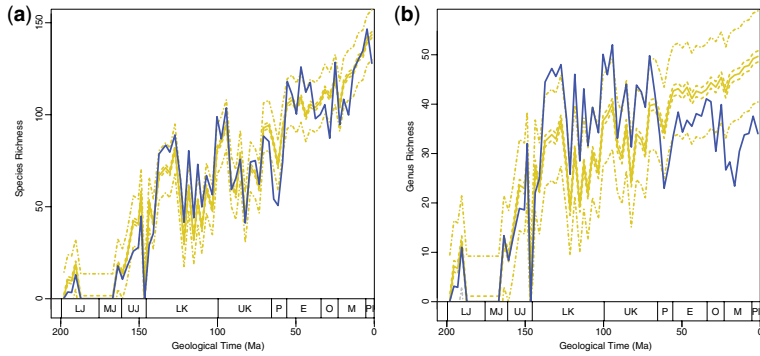
Figure 5 shows taxon richness ((a) = species, (b) = genus) estimates modelled from sampling, assuming that true diversity is constant and hence driven purely by sampling. Figure 6 shows the same data, but with observed diversity detrended by the model predictions. Figure 7 summarizes the subsampled results. Here we choose to show only two plots, both based on the maximum approach and recording the results for a relatively high sample count ( $N = 109$ ). We chose this figure as it gives continuous results (a result for successive bins) from the Early Cretaceous to present. (Higher figures meant bins with insufficient numbers of



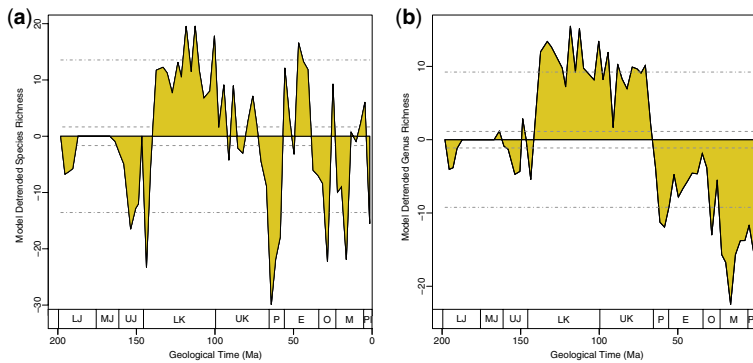
**Fig. 3.** Long term correlation –  $\log_{10}$  number of DSDP and ODP sites against: (a)  $\log_{10}$  species-level richness and (b)  $\log_{10}$  genus-level richness.



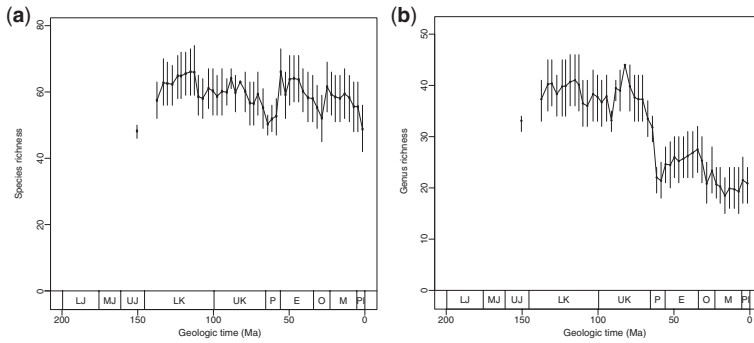
**Fig. 4.** Short term correlation time series – first differences of number of sites (blue) and: (a) species-richness (yellow) and (b) genus-level richness (yellow); 5-bin moving average detrended number of DSDP and ODP sites (blue) and: (c) species-richness (yellow) and (d) genus-level richness (yellow).



**Fig. 5.** Predicted richness based on a model (see Lloyd, in press) (yellow) against sampled richness (blue) for: (a) species and (b) genera. Dashed yellow line is a 95% confidence interval based on 1.96 standard errors, dash-dot yellow line is a 95% confidence interval based on 1.96 standard deviations (see Lloyd, in press).



**Fig. 6.** Sampling model corrected richness for: (a) species and (b) genera. Confidence intervals are as in Figure 5.



**Fig. 7.** Subsampling corrected richness at 109 species occurrences based on the maximum or optimistic approach (see text) for: (a) species and (b) genera. Vertical lines indicate 95% confidence interval based on 1000 iterations.

species occurrences could not be used, creating gaps that make interval-to-interval changes harder to interpret.) The pattern of roughly flat richness for species and declining richness for genera is still present in the minimum approach results, but there are a greater number of bins that return no result.

## Discussion

### *Sampling and observed diversity are highly correlated*

The clearest conclusion from our results is that our sampling proxy and observed diversity are very highly correlated. Indeed, this correlation is stronger than those obtained for marine invertebrates (Peters & Foote 2001; Smith & McGowan 2007) and even exceeds those for many terrestrial vertebrates (Barrett *et al.* 2009; Butler *et al.* 2009; Mannion *et al.* 2011). There are likely multiple explanations for this pattern. Firstly, unlike Peters & Foote (2001) and Smith & McGowan (2007) our sampling proxy and taxonomic richness are taken from the same database, and thus this is not a regional *v.* global comparison that is unlikely to be particularly strongly correlated. Secondly, our sampling proxy is, we contend, a much more appropriate measure of sampling (as discussed above) than those used previously and thus we are more likely to be capturing a true sampling signal rather than some other geological or anthropogenic measure. Thirdly, our data does not include Lagerstätten that can introduce outliers that confound correlation and give the false impression that correlation isn't as strong as it really is. (This is particularly true for vertebrates and future work requires methods for dealing with this such as that proposed in Cavin & Forey 2007.) The strength of this correlation, and its pervasiveness over both the short- and long-term, is the strongest evidence yet that the fossil record

should not be interpreted at face value and further that the deep-sea record is beset by the same biases as that of the land-based record.

### *Corrected diversity and biological signal*

Despite these strong correlations our data do show evidence of sampling-independent diversity excursions. Even in the uncorrected data it is clear that in the Neogene, where sampling is exceptionally high, observed species-richness starts to become decoupled from sampling. For example, the Miocene shows a progressive rise in sampling (Fig. 1), but something of a dip in richness (Fig. 2) and in the Pliocene–Pleistocene where sampling rockets upwards (Fig. 1) richness shows a downturn (Fig. 2). This inversion of ups and downs is more clearly visible in the short-term correlation plots (Fig. 4). It appears then, that correlations would be diminished if the Neogene were considered separately, and this notion is supported by a Mesozoic/Cenozoic partition. When considered separately the relatively poorly sampled Mesozoic shows a higher Spearman  $\rho$  (0.94) compared to the better-sampled Cenozoic (0.73). (The Neogene was not considered on its own as it includes too few data points for a meaningful statistical correlation.) These results are thus consistent with the notion that when sampling passes a particular threshold it no longer influences the pattern.

Additional sampling-independent signals are evident from the model-corrected measure used here. These include lower than expected (based on the standard deviation confidence interval) species-richness (Fig. 6a) in the middle Upper Jurassic, lowest Cretaceous, lowest Palaeogene, middle Oligocene, middle Miocene and in the Pliocene–Pleistocene. Many of these correspond to either hypothesized extinction events or periods of low speciation (see Bown *et al.* 2004; their Fig. 3) and the largest drop occurs with the largest extinction

at the Cretaceous/Palaeogene boundary. Comparatively few points show significantly greater than expected diversity, however: just three in the Cretaceous and one in the Eocene. For genus-level richness the picture is notably different with a clear three-phase signal in the residuals (Fig. 6b) indicating that the model is not accurately capturing the data. The first phase, corresponding to the Jurassic, is one of constant diversity with the model fitting well. However, the second is a plateau of greater than expected diversity that lasts for almost the entire Cretaceous with several points significantly higher than expected. The third and final phase is one of lower than expected diversity, with a clear declining pattern that lasts for the entire Cenozoic. Again many points are significant excursions from the model. Taken together these results bear comparison with other studies of coccolithophore diversity (e.g. Bown *et al.* 2004), where diversity rises to a Late Cretaceous peak followed by a more stable Cenozoic low (their Fig. 2).

Some of these signals are also evident in the subsampling approach, although this is perhaps the most conservative correction (Alroy 2010). The species-richness curve (Fig. 7a) seems consistent with a good fit of the constant diversity model with a near flat trajectory from the Lower Cretaceous to the Plio-Pleistocene, suggesting the empirical graph of rising species-richness (Fig. 2a) is highly misleading. However, there are some medium-term features of note. These include a slight rising trend in the middle part of the Lower Cretaceous, a declining trend in the Upper Cretaceous, depressed diversity in the Palaeocene, a declining trend through the Eocene and Oligocene and finally a declining trend in the Neogene. However, in most cases the 95% confidence intervals between successive bins are overlapping. More substantial trends are apparent in the genus-level richness curve (Fig. 7b), most obviously a dichotomy between a Mesozoic plateau of high richness and a Cenozoic plateau of significantly lower richness. Interestingly though, the biggest drop between these two levels actually occurs within the Palaeocene, suggesting this shift is not directly linked to the Cretaceous–Palaeogene extinction. Finer scale changes include more volatile short-term trends in the Upper Cretaceous, a rising trend through the Palaeogene that is curtailed at the Eocene/Oligocene boundary and generally lower richness in the Neogene than at any preceding time. Comparing the two richness corrected curves suggests greater congruence at the genus- than species-level, perhaps due to the conservative nature of the subsampling.

An interesting conclusion from the corrected approaches used here is that all of them suggest that there is either no rise or a fall in both species

and genus-level richness between the Jurassic and the Pleistocene. However, there is no doubting that coccolithophores are a morphologically diverse group and consequently a strict biological interpretation of this pattern suggests that peak coccolithophore diversity must have been established very early in the clade's history. This conclusion stands in stark contrast to the results of face-value range-through curves such as those of Bown *et al.* (1992, 2004) that show a gradual, additive rise to a peak in species-richness in the Upper Cretaceous. Our results thus support similar findings by Alroy *et al.* (2008) when comparing marine invertebrate curves. This study thus suggest that the genus-level richness curve (Fig. 2b) may be a more accurate description of coccolithophore diversity trends, with a relatively rapid initial rise followed by a gradual depletion of diversity, seemingly supporting the use of higher taxa in palaeobiodiversity estimates.

The authors would like to thank A. J. McGowan and R. Benson for their helpful reviews and S. Peters for the R function used to add the geological epochs to the x-axes of Figures 1, 2 and 4–7. This research was supported by NERC grant NE/F016905/1 to ABS, JRY and P. Pearson.

## References

- AKAIKE, H. 1973. Information theory and an extension of the maximum likelihood principle. In: PETROV, B. N. & CSÁKI, F. (eds) *2nd International Symposium on Information Theory*. Akadémiai Kiadó, Budapest, 267–281.
- ALROY, J. 2010. The shifting balance of diversity among major marine animal groups. *Science*, **329**, 1191–1194.
- ALROY, J., MARSHALL, C. R. *ET AL.* 2001. Effects of sampling standardization on estimates of Phanerozoic marine diversification. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 6261–6266.
- ALROY, J., ABERHAN, M. *ET AL.* 2008. Phanerozoic trends in the global diversity of marine invertebrates. *Science*, **321**, 97–100.
- BARRETT, P. M., MCGOWAN, A. J. & PAGE, V. 2009. Dinosaur diversity and the rock record. *Proceedings of the Royal Society, B*, **276**, 2667–2674.
- BENSON, R. B. J. & BUTLER, R. J. 2011. Uncovering the diversification history of marine tetrapods: ecology influences the effect of geological sampling bias. In: MCGOWAN, A. J. & SMITH, A. B. (eds) *Comparing the Geological and Fossil Records: Implications for Biodiversity Studies*. Geological Society, London, Special Publications, **358**, 191–207.
- BENSON, R. B. J., BUTLER, R. J., LINDGREN, J. & 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. 1995. Diversity and extinction in the history of life. *Science*, **268**, 52–58.

- BOWN, P. R., BURNETT, J. A. & GALLAGHER, L. T. 1992. Calcareous nannoplankton evolution. *Memorie di Scienze Geologiche*, **43**, 1–17.
- BOWN, P. R., LEES, J. A. & YOUNG, J. R. 2004. Calcareous nannoplankton evolution and diversity through time. In: THIERSTEIN, H. R. & YOUNG, J. R. (eds) *Coccolithophores: from Molecular Processes to Global Impact*. Springer Verlag, Berlin, Germany, 481–508.
- BUSH, A. M., MARKEY, M. J. & MARSHALL, C. R. 2004. Removing bias from diversity curves: the effects of spatially organized biodiversity on sampling-standardization. *Paleobiology*, **30**, 666–686.
- BUTLER, R. J., BENSON, R. J., NOWBATH, S. & 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.
- BUTLER, R. J., BARRETT, P. M., CARRANO, M. T., MANNION, P. D. & UPCHURCH, P. in press. Sea level, dinosaur diversity and sampling biases: investigating the ‘common cause’ hypothesis in the terrestrial realm. *Proceedings of the Royal Society, B*. (doi: 10.1098/rspb.2010.1754)
- CAVIN, L. & FOREY, P. L. 2007. Using ghost lineages to identify diversification events in the fossil record. *Biology Letters*, **3**, 201–204.
- CRAMPTON, J. S., BEU, A. G., COOPER, R. A., JONES, C. M., MARSHALL, B. & MAXWELL, P. A. 2003. Estimating the rock volume bias in paleobiodiversity studies. *Science*, **301**, 358–360.
- EBRA, E. 2004. Calcareous nannofossils and Mesozoic oceanic anoxic events. *Marine Micropaleontology*, **52**, 85–106.
- EBRA, E., BOTTINI, C., WEISSERT, H. J. & KELLER, C. E. 2010. Calcareous nannoplankton response to surface water acidification around Anoxic Event 1a. *Science*, **329**, 428–432.
- FROBISCH, J. 2008. Global taxonomic diversity of anomodonts (Tetrapoda, Therapsida) and the terrestrial rock record across the Permo-Triassic boundary. *PLoS One*, **3**, e3733.
- GOMITZ, V. (ed.) 2009. *Encyclopedia of Paleoclimatology and Ancient Environments*. Springer, Dordrecht.
- GRADSTEIN, F. M., OGG, J. G. & SMITH, A. G. (eds) 2004. *A Geologic Time Scale 2004*. Cambridge University Press, Cambridge.
- IRIGOIEN, X., HUISMAN, J. & HARRIS, R. P. 2005. Global biodiversity patterns of marine phytoplankton and zooplankton. *Nature*, **429**, 863–867.
- JOHNSON, J. B. & OMLAND, K. S. 2004. Model selection in ecology and evolution. *Trends in Ecology and Evolution*, **19**, 101–108.
- LAZARUS, D. 1994. Neptune: a marine micropaleontology database. *Mathematical Geology*, **26**, 817–832.
- LAZARUS, D. B. 2011. The deep-sea microfossil record of macroevolutionary change in plankton and its study. In: MCGOWAN, A. J. & SMITH, A. B. (eds) *Comparing the Geological and Fossil Records: Implications for Biodiversity Studies*. Geological Society, London, Special Publications, **358**, 141–165.
- LLOYD, G. T. in press. A refined modelling approach to assess the influence of sampling on palaeobiodiversity curves. *Biology Letters*.
- LLOYD, G. T., DAVIS, K. E. ET AL. 2008. Dinosaurs and the Cretaceous Terrestrial Revolution. *Proceedings of the Royal Society, B*, **275**, 2483–2490.
- MANNION, P. D., UPCHURCH, P., CARRANO, M. T. & BARRETT, P. M. 2011. Testing the effect of the rock record on diversity: a multidisciplinary approach to elucidating the generic richness of sauripodomorph dinosaurs through time. *Biological Reviews*, **86**, 157–181; doi: 10.1111/j.1469-185X.2010.00139.x
- MARX, F. G. 2009. Marine mammals through time: when less is more in studying palaeodiversity. *Proceedings of the Royal Society, B*, **276**, 887–892.
- MCGOWAN, B. 2005. *Biostratigraphy: Microfossils and Geological Time*. Cambridge University Press, Cambridge.
- MCGOWAN, A. J. & SMITH, A. B. 2008. Are global Phanerozoic marine diversity curves truly global? A study of the relationship between regional rock records and global Phanerozoic marine diversity. *Paleobiology*, **34**, 80–103.
- MURRAY, J. & RENARD, A. F. 1891. *Report on deep-sea deposits based on the specimens collected during the voyage of H.M.S. Challenger in the years 1872 to 1876*. Challenger Reports, H.M.S.O., Edinburgh.
- PETERS, S. E. & HEIM, N. A. 2011. Macrostratigraphy and macroevolution in marine environments: testing the common-cause hypothesis. In: MCGOWAN, A. J. & SMITH, A. B. (eds) *Comparing the Geological and Fossil Records: Implications for Biodiversity Studies*. Geological Society, London, Special Publications, **358**, 95–104.
- PETERS, S. E. 2005. Geologic constraints on the macroevolutionary history of marine animals. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 12326–12331.
- PETERS, S. E. 2006. Macrostratigraphy of North America. *Journal of Geology*, **114**, 391–412.
- PETERS, S. E. 2008. Environmental determinants of extinction selectivity in the fossil record. *Nature*, **454**, 626–629.
- PETERS, S. E. & FOOTE, M. 2001. Biodiversity in the Phanerozoic: a reinterpretation. *Paleobiology*, **27**, 583–601.
- PETERS, S. E. & FOOTE, M. 2002. Determinants of extinction in the fossil record. *Nature*, **416**, 420–424.
- R DEVELOPMENT CORE TEAM. 2010. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- RAUP, D. M. 1972. Taxonomic diversity during the Phanerozoic. *Science*, **177**, 1065–1071.
- RAUP, D. M. 1976. Species diversity in the Phanerozoic: an interpretation. *Paleobiology*, **2**, 289–297.
- SEPKOSKI, J. J. & KOCH, C. F. 1996. Evaluating paleontologic data relating to bioevents. In: WALLISER, H. (ed.) *Global Events and Event Stratigraphy*. Springer, Berlin, 21–34.
- SEPKOSKI, J. J., BAMBACH, R. K., RAUP, D. M. & VALENTINE, J. W. 1981. Phanerozoic marine diversity and the fossil record. *Nature*, **293**, 435–437.
- SMITH, A. B. 2001. Large-scale heterogeneity of the fossil record: implications for Phanerozoic biodiversity studies. *Philosophical Transactions of the Royal Society of London, B*, **356**, 351–367.

- SMITH, A. B. 2007. Marine diversity through the Phanerozoic: problems and prospects. *Journal of the Geological Society, London*, **164**, 731–745.
- SMITH, A. B. & 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.
- SPENCER-CERVATO, C. 1998. Changing depth distribution of hiatuses during the Cenozoic. *Paleoceanography*, **13**, 178–182.
- SUCHÉRAS-MARX, B., MATTIALI, E., PITTET, B., ESCARGUEL, G. & SUAN, G. 2010. Astronomically-paced coccolith size variations during the early Pliensbachian (Early Jurassic). *Palaeogeography, Palaeoclimatology, Palaeoecology*, **295**, 281–292.
- UPCHURCH, P. & BARRETT, P. M. 2005. Phylogenetic and taxic perspectives on sauropod diversity. In: CURRY ROGERS, K. A. & WILSON, J. A. (eds) *The Sauropods: Evolution and Paleobiology*. University of California Press, Berkeley, 104–124.
- UPCHURCH, P., MANNION, P. D., BENSON, R. B. J., BUTLER, R. J. & CARRANO, M. T. 2011. Geological and anthropogenic controls on the sampling of the terrestrial fossil record: a case study from the Dinosauria. In: MCGOWAN, A. J. & SMITH, A. B. (eds) *Comparing the Geological and Fossil Records: Implications for Biodiversity Studies*. Geological Society, London, Special Publications, **358**, 209–240.
- WANG, P.-X., ZHAO, Q.-H. ET AL. 2003. Thirty-million year deep sea records in the South China Sea. *Chinese Science Bulletin*, **48**, 2524–2535.
- WALL, P. D., IVANY, L. C. & 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.
- WINTER, A., JORDAN, R. W. & ROTH, P. H. 1994. Biogeography of living coccolithophorids in ocean waters. In: WINTER, A. & SIESSER, W. G. (eds) *Coccolithophores*. Cambridge University Press, Cambridge, 161–177.
- ZIVERI, P., BAUMANN, K.-H., BÖCKEL, B., BOLLMANN, J. & YOUNG, J. R. 2004. Present day coccolithophore biogeography of the Atlantic Ocean. In: THIERSTEIN, H. R. & YOUNG, J. R. (eds) *Coccolithophores: From Molecular Processes to Global Impact*. Springer, Berlin, 403–427.