

ARTICLE

Received 28 Mar 2014 | Accepted 28 Jul 2014 | Published 4 Sep 2014

DOI: 10.1038/ncomms5818

Disentangling rock record bias and common-cause from redundancy in the British fossil record

Alexander M. Dunhill^{1,2,†}, Bjarte Hannisdal³ & Michael J. Benton²

The fossil record documents the history of life, but the reliability of that record has often been questioned. Spatiotemporal variability in sedimentary rock volume, sampling and research effort especially frustrates global-scale diversity reconstructions. Various proposals have been made to rectify palaeodiversity estimates using proxy measures for the availability and sampling of the rock record, but the validity of these approaches remains controversial. Targeting the rich fossil record of Great Britain as a highly detailed regional exemplar, our statistical analysis shows that marine outcrop area contains a signal useful for predicting changes in diversity, collections and formations, whereas terrestrial outcrop area contains a signal useful for predicting formations. In contrast, collection and formation counts are information redundant with fossil richness, characterized by symmetric, bidirectional information flow. If this is true, the widespread use of collection and formation counts as sampling proxies to correct the raw palaeodiversity data may be unwarranted.

¹Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK. ²School of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol BS8 1RJ, UK. ³Centre for Geobiology, Department of Earth Science, University of Bergen, Allegaten 41, N-5007 Bergen, Norway. † Present address: School of Earth and Environment, University of Leeds, LS2 9JT, UK. Correspondence and requests for materials should be addressed to A.M.D. (email: a.dunhill@leeds.ac.uk).

Understanding biotic evolution through deep time is a key research agenda in palaeobiology and Earth system science^{1–3}. The evidence resides in the fossil record, and yet this record is compromised by incompleteness and bias, and it has been debated whether bias dominates the data^{1,4–9}, or not^{10–13}. Further, in the search for methods to provide a bias-free, or corrected, palaeodiversity signal, proposals have been made to use rock outcrop areas and collection and formation counts as sampling proxies to correct the raw palaeodiversity data^{8,14–18}, but the usefulness of such approaches has been queried^{19,20}.

There are further implications for interpreting how life diversified from a single species billions of years ago to 5–10 million eukaryotic species today²¹, and the consequent impact on global climate, and chemistry of oceans and atmosphere: do palaeodiversity records support a heavily damped and perturbed exponential model of diversification^{22–24}, did biodiversity reach an equilibrium level equivalent to today's biodiversity 500 Myr ago^{4,6}, or has the diversity of individual clades fluctuated idiosyncratically in response to adaptive radiations and extinctions over the past 500 Myr¹? The latter two hypotheses imply that the apparent rise in palaeodiversity through the past 500 Myr must be explained as sampling bias. It has not been clear how this debate could be resolved^{1,4,6,10,12} except by the input of alternative data sets or methods.

The most commonly implemented techniques to account for bias are sampling standardization, which seeks to equalize or make sampling fair at the level of collections^{1,25–27} and, model fitting using sampling proxies to identify times of poor and good sampling and to apply *post hoc* corrections^{8,17,27}. Sampling proxies include measures of rock outcrop area, geological formations and collections.

Palaeodiversity (fossil taxonomic richness) curves and their covariation with sampling proxies can be explained by any of three hypotheses, alone or in combination: (1) the rock record bias (RRB) hypothesis^{4–7,27} that variability in the amount of available rock determines fossil diversity; (2) the common-cause (CC) hypothesis^{11,13,28–30} that much of the covariation of fossil and rock records is because both are driven by a third factor such as sea level change; or (3) the redundancy (RED) hypothesis^{20,31} that rock and fossil record proxies covary because of operational redundancy (that is, more collecting may result in greater richness, but high richness may also result in more collecting) and statistical redundancy (that is, various time series are different versions of the same signal), reflecting the mutual reinforcement of the sampling proxy and fossil richness. It is important to note that the CC and RED hypotheses are distinct in that the CC requires that two non-redundant variables are driven by a third variable, whereas the RED requires a two-way causality between variables without a third factor driving the dependence. The outstanding question is how to quantify the relative importance of these three, likely non-mutually exclusive, hypotheses. What is needed is some statistical means of indicating the directionality of potential drive–response relationships, in lieu of mechanistic process models.

Regional scale, rather than global, studies may provide a useful approach^{32–35}, and if well chosen, they can offer the advantage of comprehensive, evenly reported data and focus on a single geological history. Here, we present an investigation of the rock and fossil records of Great Britain, which, despite being recorded from a relatively limited geographic region, encompass an almost continuous range of stratigraphic intervals that are probably the most intensely geologically sampled and documented area in the world^{35–37}. Geological data have been recorded for over 200 years and the British Geological Survey has established detailed, fine-scale stratigraphies, and made its rich data stores available in electronic, georeferenced format (Fig. 1).

Here, we evaluate the RRB, CC and RED hypotheses using information transfer (IT), a non-parametric technique for quantifying the relative strength and directionality of predictive information flow between components of a coupled system (see Supplementary Methods and Supplementary Figs 1–5)^{13,38,39}. We apply pairwise correlation and IT to detailed Phanerozoic marine and terrestrial stratigraphical data sets from Great Britain (Fig. 2), global environmental proxy (Supplementary Fig. 6), and palaeodiversity data (downloaded from the Paleobiology Database (PaleoDB; <http://paleobiodb.org/>; Fig. 2). Sampling proxies are assessed at the epoch level to determine how well they correlate with, and predict, palaeodiversity.

Under the RRB hypothesis, we expect rock quantity proxies (that is, formation counts and outcrop area) to have the strongest influence on palaeodiversity, beyond mutual correlations with other environmental variables, and we expect IT to flow from rock quantity to palaeodiversity. For the CC hypothesis, other environmental proxies would have the strongest influence on palaeodiversity, beyond mutual correlations with rock quantity, and IT would flow from environmental proxies to palaeodiversity. For the RED hypothesis, we expect sampling proxies whose origin is partly controlled by the distribution of fossil richness in the field (for example, collections and formations) to show strong, bidirectional IT with palaeodiversity, to be redundant with palaeodiversity in conditional IT (CIT), and to respond to the same environmental drivers as palaeodiversity. Because the three hypotheses are non-exclusive, and because each of the observed records may capture multiple underlying processes, the predicted relationship between pairs of variables may not be uniquely specified under each hypothesis, but may vary depending on the relationship between other variables. Assessing the relative degree of support for the different hypotheses therefore requires the full combination of statistical results.

We find some common signals, and some differences between the marine and terrestrial data. Both data sets show that collection and formation counts are information redundant with fossil richness, characterized by symmetric, bidirectional information flow. However, whereas marine outcrop area contains a signal useful for predicting changes in genera, collections, and formations, terrestrial outcrop area contains a signal useful for predicting only terrestrial formations.

Results

Marine data. In the marine data, the strongest correlations are found between palaeodiversity (genera) and collections, between genera and formations and between formations and outcrop (see Supplementary Table 1 for correlation coefficients). The correlation between genera and outcrop area is not significant after false discovery rate correction, and compared with genera, collections show weaker correlations with both formations and outcrop. It is important to note that correlations are calculated after first differencing, which isolates short-term (bin-to-bin) changes and filters out longer-term variation. In contrast, IT takes into account both short- and long-term variation, relying instead on surrogate data to accommodate for autocorrelation. IT analysis thus modifies and expands our view of the statistical relationships in the marine data (Fig. 3). There is strong, bidirectional IT between genera and collections (Fig. 3a), and although there is a slight asymmetry favouring IT from collections to genera, this asymmetry is very rarely significant (Collections > Genera is rarely detected), suggesting that they predict each other equally well. The IT between formations and genera is somewhat weaker, but also bidirectional (Fig. 3b). IT from genera to formations is detected with slightly higher frequency than in the opposite direction, but again this asymmetry is not significant. Collections

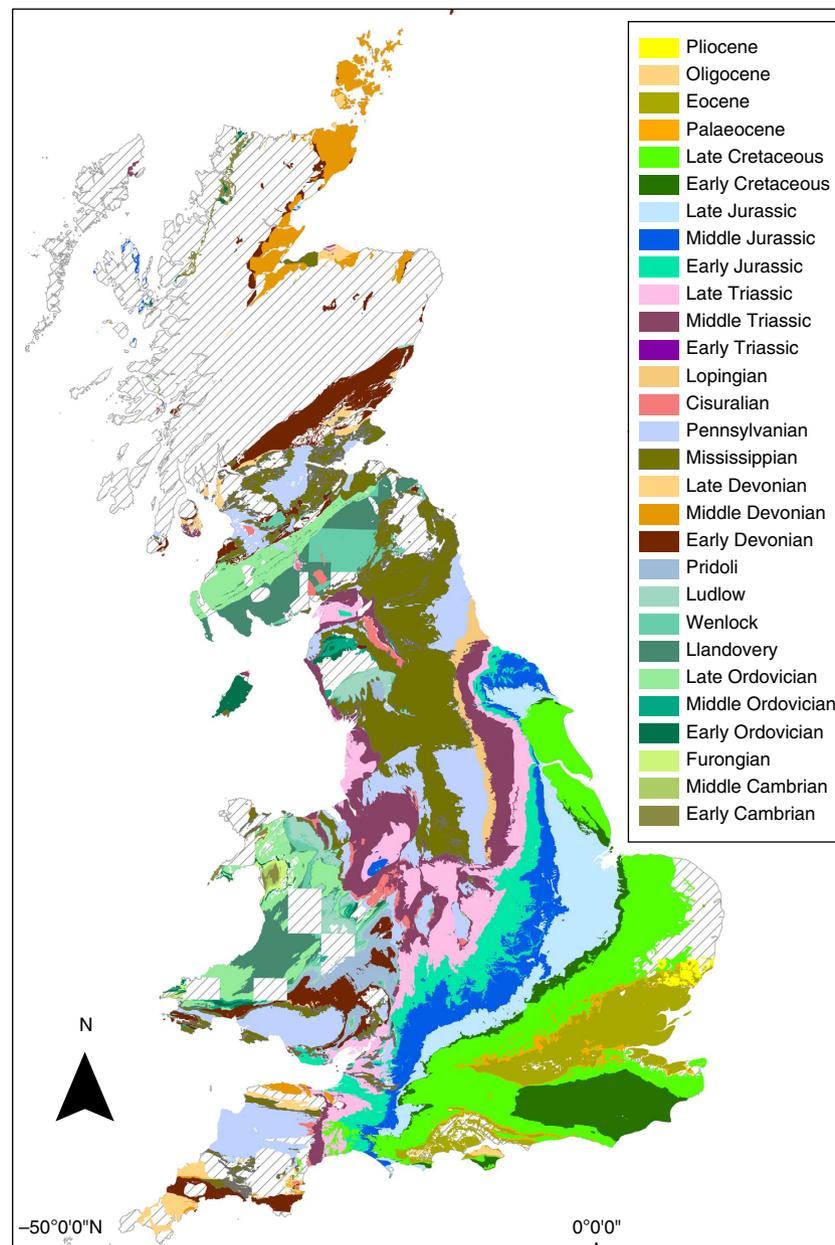


Figure 1 | The Phanerozoic record of Great Britain. Geological map of British Phanerozoic sedimentary rock outcrop at the Epoch level (British Geological Survey DiGMapgb-50 1:50,000). Hashed areas represent extent of igneous and metamorphic rocks or areas that have not been digitally mapped. There are no data recorded from the Guadalupian or Miocene.

and formations (Fig. 3c) show essentially the same pairwise IT association as genera and formations, the only notable difference being that (contrary to the correlations) the relationship between the former appears slightly stronger than between the latter. The IT between genera and outcrop area also converges on significance in both directions, but in this case there is significant asymmetry in favour of IT from outcrop to genera (Outcrop > Genera), suggesting that changes in outcrop area can be used to predict changes in palaeodiversity more than vice versa (Fig. 3d). IT from outcrop to collections shows the same significant directionality (Outcrop > Collections), albeit with a slightly lower detection rate (Fig. 3e). Finally, the IT between formations and outcrop area is the most clearly asymmetric relationship (Outcrop > Formations), verging on unidirectional, suggesting that outcrop is much more useful for predicting formations than vice versa (Fig. 3f).

These findings together imply that in the marine data, genera and collections are tightly linked to each other, and to a lesser extent to formations, through symmetric, bidirectional IT, while outcrop area contains a signal useful for predicting changes in genera, collections and formations (Fig. 4).

CIT analyses of the marine data suggest that formations contribute no information on genera not already provided by collections, and little information on collections not already found in genera (Fig. 5a). CIT from formations to collections is slightly stronger than from formations to genera (Fig. 5a), attributable to a slight asymmetry (typically insignificant) in the pairwise IT from collections to genera (Fig. 3a). Genera and collections have the strongest mutual relationship, but outcrop nonetheless provides significant CIT (Fig. 5b). Although pairwise IT from outcrop to collections is no greater than from outcrop to genera (Fig. 3d), CIT from outcrop to collections given genera

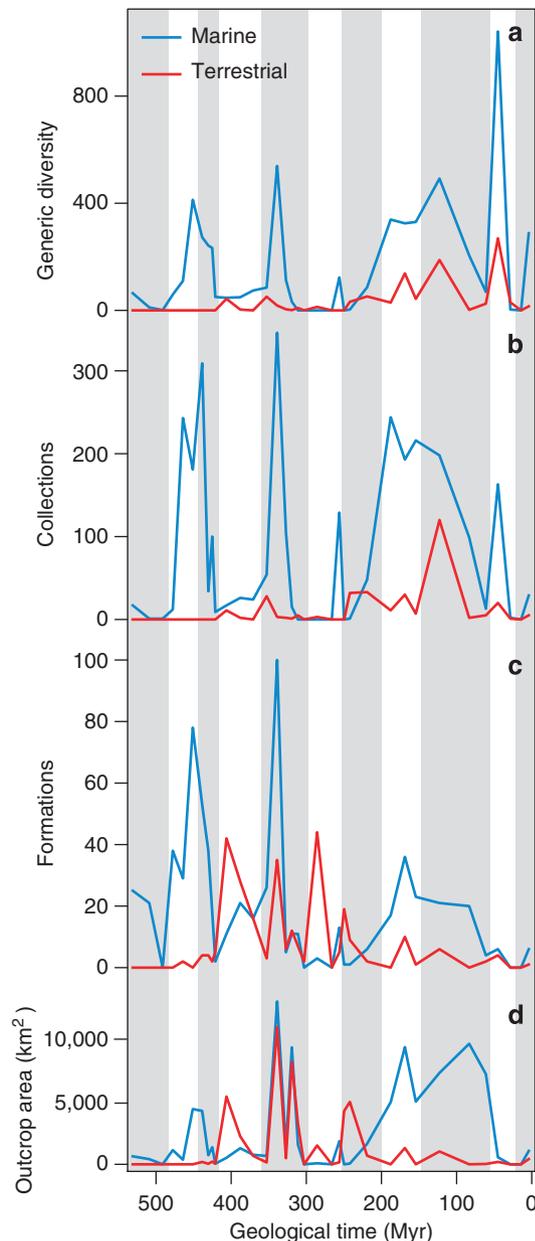


Figure 2 | Palaeodiversity and sampling proxies. Time series of (a) generic palaeodiversity, (b) collections, (c) formations and (d) outcrop area through the British marine and terrestrial Phanerozoic at Epoch level.

exceeds CIT from outcrop to genera given collections (Fig. 5b), again attributed to pairwise asymmetry between collections and genera (Fig. 3a). Outcrop contains non-redundant CIT on both genera and formations (Fig. 5c). Pairwise IT from outcrop to formations (Fig. 3f) roughly equals that from outcrop to genera (Fig. 3d), but because pairwise IT from genera to formations is slightly (if insignificantly) greater than from formations to genera (Fig. 3b), conditioning on genera results in weaker CIT from outcrop to formations (Fig. 5c). Replacing genera with collections involve the same interrelationships, but pairwise IT from collections to formations (Fig. 3c) is slightly stronger than from genera to formations (Fig. 3b), resulting in even weaker CIT from outcrop to formations conditioned on collections (Fig. 5d) relative to conditioning on genera. Although both genera and collections show significant pairwise IT to outcrop (Fig. 3d,e), both are typically insignificant when conditioned on formations

(Fig. 5e). The latter is even weaker, both in pairwise IT (Fig. 3f) and CIT (Fig. 5e). Any IT from genera and collections to formations or outcrop disappears when genera and collections are conditioned on each other (Fig. 5f).

Terrestrial data. In the terrestrial data, the strongest correlations are found between genera and collections, and between formations and outcrop area (Supplementary Table 1), and these are also the only significant IT relationships. IT between genera and collections is strong and symmetric (Fig. 3g), whereas IT between formations and outcrop is weaker and approaches significant asymmetry in favour of outcrop (Outcrop > Formations; Fig. 5l). In the conditional analyses, neither formations nor outcrop contain any significant information on genera or collections, while the latter show strong, symmetric CIT (Fig. 5g,h). Outcrop and formations show significant CIT, with CIT from outcrop to formations being stronger than in the opposite directions (Fig. 5i-k). Genera and collections show no CIT to outcrop or formations (Fig. 5l). The terrestrial CIT results thus reiterate the pairwise IT results, suggesting that terrestrial outcrop area contains a signal useful for predicting terrestrial formations, more than vice versa (Fig. 4).

Palaeoenvironmental proxy data. Although correlations suggest significant relationships between marine formations and $^{87/86}\text{Sr}$, and between terrestrial formations and $\delta^{18}\text{O}$ (Supplementary Table 1), we found no significant IT between the UK Phanerozoic records and global palaeoenvironmental proxy records, including sea level. These results should be regarded as tentative until regional proxy records (for example, regional flooding) are tested.

Discussion

In combination, three of these findings, namely (i) strong symmetric IT, (ii) similar responses to outcrop area and (iii) cancelling out in conditional analyses, suggest that genera and collections are information redundant in the marine data. In the terrestrial data, only (i) can be demonstrated. The causal relationship between palaeodiversity and collections can in principle go both ways (Fig. 3), because palaeontologists' collecting effort is to some extent guided by fossil richness in the field^{19,33,34,40}. Therefore, caution is needed if collections are employed to 'correct' the palaeodiversity record. We did not assess occurrences (=localities) because of difficulties of definition and data compilation. However, we expect they would show the same patterns of potential RED as between collections and palaeodiversity.

Our results also suggest that marine formations can be considered information redundant with respect to palaeodiversity and collections. Although changes in fossil richness are not part of the formal criteria for defining formations, both may be confounded by changes in the primary depositional environment, and greater environmental/faunal turnover may enable a finer partitioning of formations. If formation boundaries were independent of fossil diversity, then changes in average richness from one formation to the next should not differ significantly from the changes obtained after randomly shuffling the formations over a fixed distribution of genera. Here, we use UK Triassic–Jurassic data¹⁹ to show that cross-formation changes in average generic richness are greater than would be expected under independence (see Methods for randomization test description; Fig. 6). The differences are less significant in the regional subsets, partly due to smaller sample sizes, and are generally less significant in the terrestrial than in the marine formation sets (Supplementary Figs 7–9). This covariation suggests that formations should not be considered a measure of

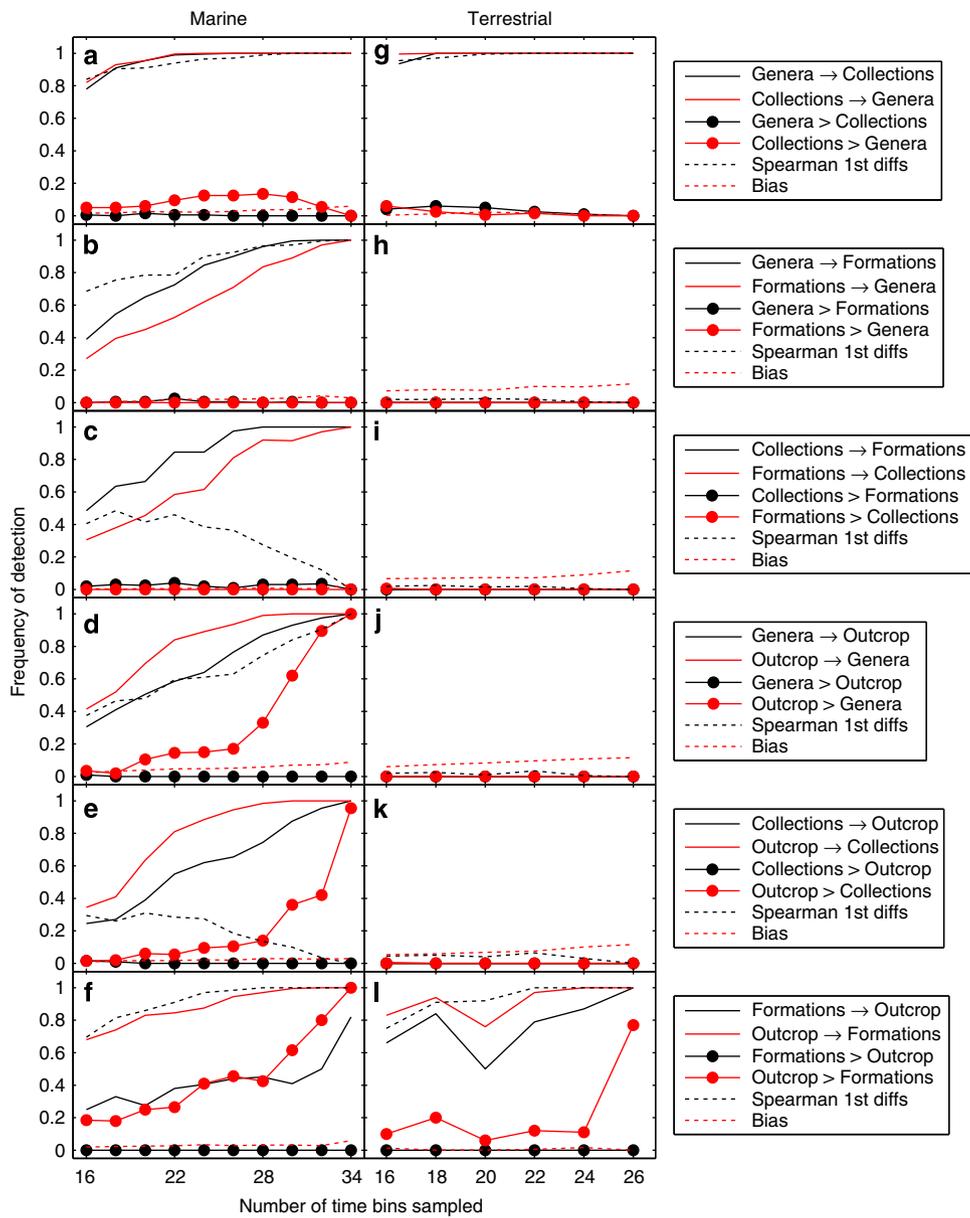


Figure 3 | Directional IT and correlations between sampling proxies and palaeodiversity in the marine and terrestrial data. Frequency of detection represents the proportion of significant results ($\alpha=0.01$) in 500 analyses by repeated random sampling of the original time series. Detection frequency (statistical ‘sensitivity’) is evaluated as a function of the number of time bins (epochs) sampled in each iteration. For each pair of variables X and Y (indicated in the plot legend for each row of panels), solid lines correspond to significant IT in each direction ($X \rightarrow Y$ and $Y \rightarrow X$). Filled circles correspond to a detected difference in magnitude between the two (where $X > Y$ denotes that $X \rightarrow Y$ is significantly greater than $Y \rightarrow X$), which represents significant asymmetry or directionality in the IT. Stippled black line corresponds to significant Spearman rank-order correlations on first differences. Stippled red line represents an index on [0,1] of potential bias due to differences in non-stationarity, where bias values of 0.1 or less are considered negligible. See text for description of the results in each panel.

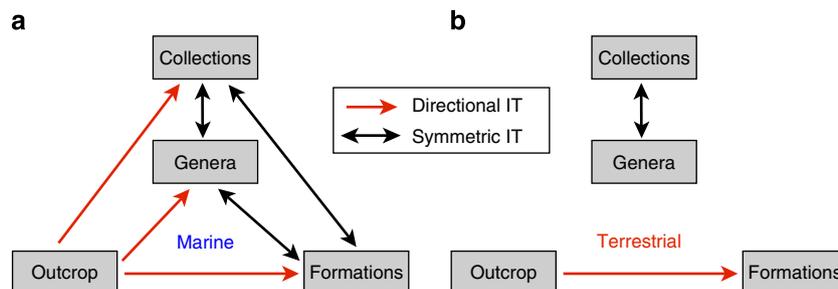


Figure 4 | IT in the marine and terrestrial realms. Flowcharts summarizing IT results in **a**, marine data and **b**, terrestrial data, indicating symmetric (black arrows) and directional (red arrows) information flow.

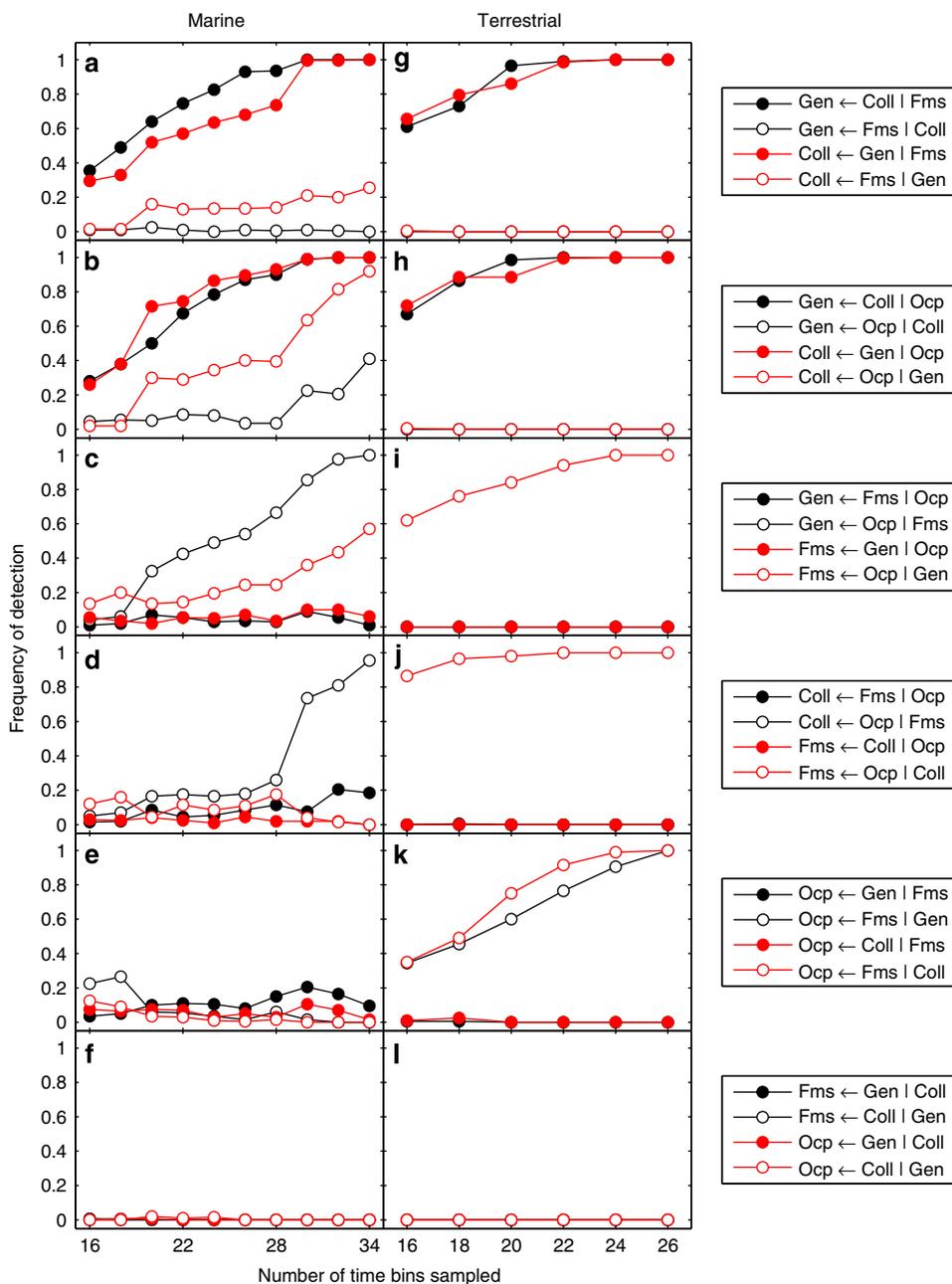


Figure 5 | CIT amongst sampling proxies and palaeodiversity in the marine and terrestrial data. Frequency of detection represents the proportion of significant results ($\alpha = 0.01$) in 500 analyses by repeated random sampling of the original time series. Detection frequency (statistical ‘sensitivity’) is evaluated as a function of the number of time steps sampled in each iteration. For each set of three variables X, Y and Z (indicated in the plot legend for each row of panels), significant $X \leftarrow Y|Z$ implies that Y contains information useful for predicting changes in X beyond their mutual interaction with Z. Black and red filled circles correspond to switching the ‘response’ (X) and ‘driver’ (Y) variables for a fixed conditioning variable Z (that is, opposite directions $X \leftarrow Y|Z$ and $Y \leftarrow X|Z$). For a given colour, filled and open circles correspond to switching the ‘driver’ and conditioning variable for a fixed ‘response’ variable. See text for description of the results in each panel. Gen, Genera; Coll, Collections; Fms, Formations; Ocp, Outcrop.

rock quantity or sampling that is derived independently of changes in observed fossil diversity, and thus strengthens the argument of the RED. However, in line with the IT results, formations are less redundant with palaeodiversity in the terrestrial data than in the marine data.

Outcrop area, on the other hand, is a potential sampling proxy that may drive, but is not driven by, palaeontological research effort or fossil richness in the field. Our analysis suggests that outcrop area does contain a signal useful for predicting richness, collections and formations in the marine data (Fig. 4). Although

this result may seem to favour the CC, it is not strictly a test of the CC hypothesis. Richness, collections, formations and outcrop area are all to some extent confounding factors, and thus bound by common-cause relationships in the general, statistical sense. However, the CC hypothesis as used in the palaeodiversity literature states more specifically that covariation between palaeodiversity and the amount of sedimentary rock is observed because both respond to similar environmental drivers, such as the degree of continental flooding¹¹. To test this, we ideally need palaeoenvironmental proxies that are separate from the rock

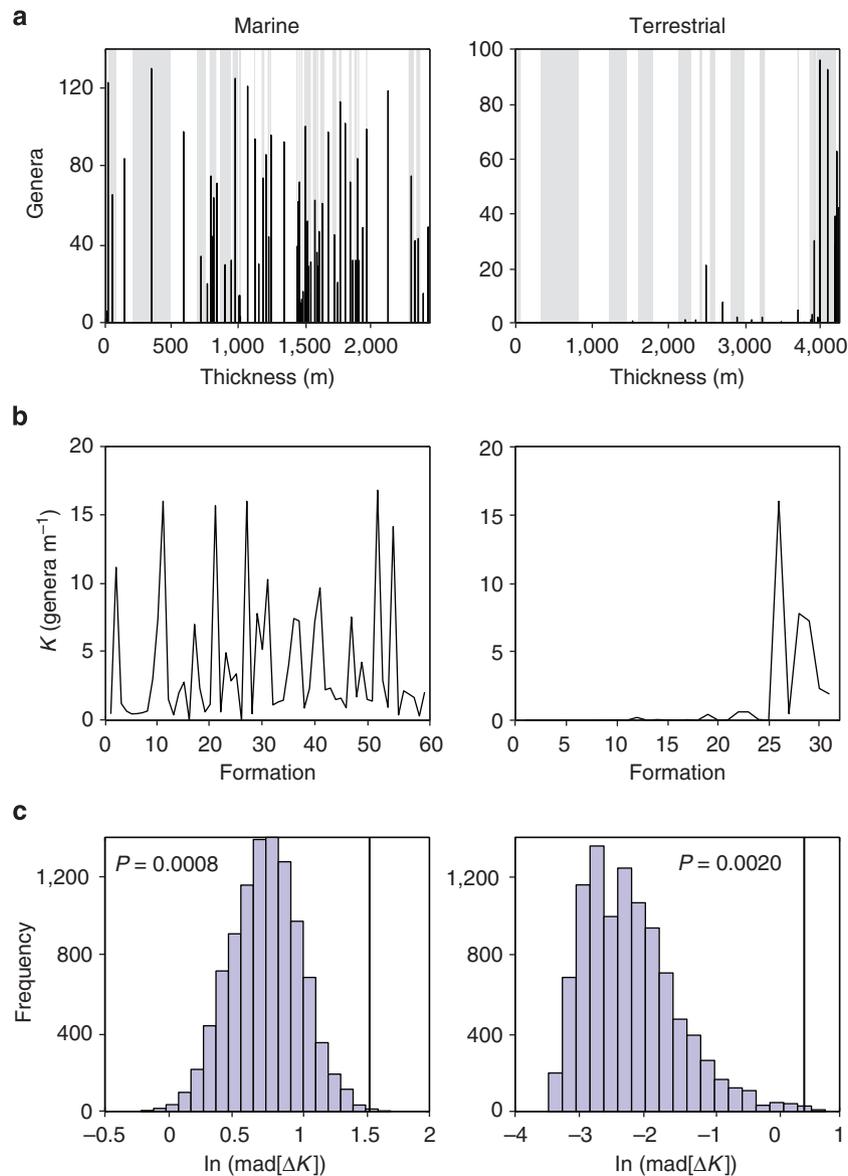


Figure 6 | Testing for fossil formation independence in Triassic-Jurassic rocks. (a) Cumulative thickness of stacked formations (alternating white/grey-shaded), with the number of reported genera (black bars) indicated at the centre of each formation. (b) Average fossil richness K (genera per m) in each formation. (c) Comparing the observed volatility of K (black line), measured as the mean absolute deviation (mad) of first differences (ΔK), against null distributions for 10,000 shuffles of K (randomly reordering the formation stack while keeping the distribution of genera fixed). P -values for the permutation test represent the proportion of values in the null distribution (histogram) that exceeded the observed volatility.

quantity proxy (for example, Hannisdal and Peters¹³). Outcrop area is a proxy that may reflect relative changes in the extent of continental flooding (primary depositional/shelf area), but also subsequent erosion and preservation (rocks available for sampling). Hence, in the absence of a regional flooding proxy, the outcrop area results are consistent with both RRB and CC hypotheses. Regardless of conditioning on outcrop area or formations, CIT between palaeodiversity and collections remains strong and symmetric, as expected under RED. If CC had acted to drive both richness and collections, but collections were the main cause of richness variation via anthropogenic sampling without RED, then we would expect asymmetric CIT from collections to palaeodiversity conditioned on outcrop area. If we assume that outcrop area represents CC mechanisms more than RRB, then our finding that CIT from outcrop to palaeodiversity is stronger than CIT from formations to palaeodiversity would suggest CC. However, formations seem to provide very

little information not already contained in the other variables, and irrespective of whether CC or RRB is acting, the bidirectional IT between formations and palaeodiversity, together with the formation randomization results, are more simply explained by RED. In the marine realm, a species-area effect could result from either a continental flooding effect, in line with CC, or a sampling effect, in line with RRB; there is greater spatiotemporal continuity of deposition, facies tend to intergrade, and formations may be indirectly influenced by changes in fossil diversity. Marine formations may therefore be more redundant with palaeodiversity, and changes in area/volume may be more important than changes in habitat-specific preservation potential, thus outcrop better predicts diversity. In the terrestrial realm, there is no obvious area effect, little spatiotemporal continuity of deposition, and fossil preservation is more linked to habitat.

Sedimentary processes may also account for differences between the marine and terrestrial realms, with consistent good preservation

in marine deposits and sporadic preservation in terrestrial sediments^{19,41}. Therefore, the amount of rock preserved and accessible to palaeontologists may partly control palaeodiversity in the marine realm, whilst any species-area effect (whether RRB or CC driven) is overwhelmed by sporadic preservation in the terrestrial realm¹⁹. It could also be that we see a CC mechanism operating in the marine realm^{11,13,42}, but not in the terrestrial realm⁴³ given the lack of hypothesized common drivers of sedimentation and diversity in the terrestrial realm. Either way, outcrop area is a generalized measure of rock availability and the degree of exposed bedrock can vary significantly across rocks at different altitudes, locations, and ages^{36,44}. In any case, a singular sampling metric will fail to capture all bias affecting palaeodiversity^{5,20,33}, and unless CC can be ruled out, outcrop area should not be used as a simple correction factor.

It would be ideal if a sampling proxy could be found that would allow palaeontologists to remove bias from their empirical palaeodiversity data. Several approaches to correcting the fossil record have been proposed, but our results suggest caution in applying these. Our results do not address the rarefaction and SQS sampling correction approaches adopted by Alroy^{1,25}, but they do have a strong bearing on methods that use residuals from comparisons of fossil record (palaeodiversity) and rock record (collection counts, formation counts, map areas) time series.

First, we argue that it is wrong to claim that outcrop areas calculated from geological maps are a meaningful measure of sampling^{5,7,17,27} because we find limited evidence that outcrop areas equate to rock accessibility^{20,33,34,36,44}. Of course, outcrop areas may include some aspect of sampling, but this is probably overwhelmed by a number of unpredictable factors, such as (1) outcrop areas incorporate tracts of unpredictable size that are devoid of fossiliferous rocks²⁹, (2) they include other tracts of unpredictable size that are concealed beneath soil^{29,36,44}, (3) these first two factors are heavily dependent on rock facies, so outcrop areas of different ages cannot be compared unless they comprise similar facies distributions^{19,33,35,41}, (4) fossil recovery depends on the maturity of sampling of individual localities, so equivalent outcrop areas of different ages or in different parts of the world may be sampled to a greater or lesser extent and so cannot be compared as metrics of sampling and (5) the species-area effect and its relationship with flooding may not be (log-) linear or even monotonic⁴⁵, and outcrop area is unlikely to be related to species counts in a simple way. However, this problem is generally avoided as most studies do not assume a linear relationship between outcrop area and species counts. As outcrop area represents the total amount of rock preserved more effectively than it represents the amount of accessible rock (that is, exposed bedrock), the IT between outcrop area and palaeodiversity may indicate a stronger role for CC relative to RRB. However, independent, regional environmental proxy records are required to further resolve this.

Further, our results from correlation, IT and formation randomization tests all indicate that formation and collection counts may be influenced by fossil diversity, and so neither can be used as an independent sampling proxy to correct the palaeodiversity record, as formerly recommended by some^{8,14,15}. Proceeding to use such methods to generate a 'bias-corrected' palaeodiversity curve risks obliterating true biological signal. The question is probably much more complex than has often been assumed. Ultimately, palaeodiversity reconstruction will have to address a multitude of biasing factors, by defining a space of hypotheses in the form of stochastic models⁴⁶, and confronting these models with available sources of palaeontological, stratigraphic and geochemical data, using formal inversion methods that more realistically account for the time-varying uncertainty in both models and data.

Methods

Data. Generic occurrence data, formation counts and collection counts were obtained from the PaleoDB (<http://paleodb.org/>) in January 2013. PaleoDB data was derived from 31,321 occurrences of 4,029 genera within 2,829 collections. As a whole, the PaleoDB varies in completeness through geological time. Therefore, it is inevitable that this British subset should suffer from this same issue. In particular, the time bins of the Late Carboniferous, the Permian and the Miocene appear particularly under-sampled, whereas the Lower Palaeozoic, Jurassic, Cretaceous and Eocene appear to be the most thoroughly sampled. However, as this study is primarily aimed at detecting sampling issues, this should not cause a problem regarding the interpretation of our results. Outcrop area measurements were obtained from the British Geological Survey digital bedrock geology DiGMapgb-50 of the UK (1:50,000; Fig. 1). Palaeoenvironmental time series were resampled from the original data sources (Hannisdal and Peters¹³) and bin-averaged in the time bins of the UK rock and fossil data.

Thickness and palaeodiversity for UK Triassic–Jurassic formations were obtained from Dunhill *et al.*¹⁹ for (i) sequence across the entire UK, (ii) the Wessex basin, (iii) the East Midlands basin and (iv) the Yorkshire basin. Formations were stacked in stratigraphic sequence (using mean thickness across different locations for each formation in the total sequence) with the number of fossil genera G_i assigned to a single point in the centre of the i th formation. Average fossil richness $K_i = G_i/Z_i$ (genera per m) was calculated for each formation i with thickness Z_i . Significance tests were carried out on the mean absolute deviation of first differences in the observed K against 10,000 shuffles of K calculated by keeping G fixed but randomly reordering Z , to test whether variation across formation boundaries is greater than would be expected by chance.

Statistical analysis. Spearman rank-order correlations were calculated on first differences. IT was calculated pairwise, using 500 amplitude-adjusted Fourier transform surrogates to establish significance of IT in each direction, $X \rightarrow Y$, and $Y \rightarrow X$. If one or both, then the difference between IT in opposite directions was compared to that of the surrogate distribution to test for significantly asymmetric (directional) information flow ($X > Y$ or $Y > X$, where $X > Y$ denotes that $X \rightarrow Y$ is significantly greater than $Y \rightarrow X$). CIT was calculated on sets of three variables to test if the IT between X and Y was still significant when taking into account their common interaction with a third variable, Z ³⁸. To evaluate whether or not differences in non-stationarity could bias the IT between two time series, a bias index was calculated from the KPSS test⁴⁷, such that a maximum value of one indicates different non-stationarity at all time lags, and a minimum value of zero indicates no differences. If needed to minimize bias, time series were detrended (linearly, or using a higher-order polynomial fit) and power transformed to stabilize the variance (Box–Cox). The Eocene was removed to avoid excessive data regularization prompted by the Eocene diversity 'spike'. All records were normalized to mean zero and unit standard deviation before analysis. Correlation/IT results are presented as sensitivity analyses, giving the proportion of significant findings (frequency of detection) in 500 analyses by iterative sampling of the original data, plotted as a function of the number of time bins sampled in each iteration (see Supplementary Methods for more details).

References

- Alroy, J. The shifting balance of diversity among major marine animal groups. *Science* **329**, 1191–1194 (2010).
- Benton, M. J. The red queen and the court jester: species diversity and the role of biotic and abiotic factors through time. *Science* **323**, 728–732 (2009).
- Kump, L. R., Kasting, J. F. & Crane, R. G. *The Earth System* 3rd edn, 420 (Prentice Hall, 2010).
- Raup, D. M. Taxonomic diversity during the Phanerozoic. *Science* **177**, 1065–1071 (1972).
- Smith, A. B. Marine diversity through the Phanerozoic: problems and prospects. *J. Geol. Soc. London* **164**, 731–745 (2007).
- Raup, D. M. Species diversity in the Phanerozoic: an interpretation. *Paleobiology* **2**, 289–297 (1976).
- Smith, A. B. Large-scale heterogeneity of the fossil record: implications for Phanerozoic biodiversity studies. *Philos. Trans. R. Soc. Lond. B Biol. Soc.* **356**, 351–367 (2001).
- Peters, S. E. & Foote, M. Biodiversity in the Phanerozoic: a reinterpretation. *Paleobiology* **27**, 583–601 (2001).
- Peters, S. E. & Foote, M. Determinants of extinction in the fossil record. *Nature* **416**, 420–424 (2002).
- Benton, M. J. Diversification and extinction in the history of life. *Science* **268**, 52–58 (1995).
- Peters, S. E. Geologic constraints on the macroevolutionary history of marine animals. *Proc. Natl Acad. Sci. USA* **102**, 12326–12331 (2005).
- Sepkoski, J. J., Bambach, R. K., Raup, D. M. & Valentine, J. W. Phanerozoic marine diversity and the fossil record. *Nature* **293**, 435–437 (1981).
- Hannisdal, B. & Peters, S. E. Phanerozoic earth system evolution and marine biodiversity. *Science* **334**, 1121–1124 (2011).

14. Benson, R. B. J., Butler, R. J., Lindgren, J. & Smith, A. S. Mesozoic marine tetrapod diversity: mass extinctions and temporal heterogeneity in geological megabiases affecting vertebrates. *Proc. R. Soc. B Biol. Sci.* **277**, 829–834 (2010).
15. Barrett, P. M., McGowan, A. J. & Page, V. Dinosaur diversity and the rock record. *Proc. R. Soc. B Biol. Sci.* **276**, 2667–2674 (2009).
16. Brocklehurst, N., Kammerer, C. F. & Fröbisch, J. The early evolution of synapsids, and the influence of sampling on their fossil record. *Paleobiology* **39**, 470–490 (2013).
17. Smith, A. B. & McGowan, A. J. The shape of the Phanerozoic marine palaeodiversity curve: how much can be predicted from the sedimentary rock record of Western Europe? *Palaentology* **50**, 765–774 (2007).
18. Wall, P. D., Ivany, L. C. & Wilkinson, B. H. Revisiting Raup: exploring the influence of outcrop area on diversity in light of modern sample-standardization techniques. *Paleobiology* **35**, 146–167 (2009).
19. Dunhill, A. M., Benton, M. J., Twitchett, R. J. & Newell, A. J. Testing the fossil record: sampling proxies and scaling in the British Triassic–Jurassic. *Palaeoogeogr. Palaeoclimatol. Palaeoecol.* **404**, 1–11 (2014).
20. Benton, M. J., Dunhill, A. M., Lloyd, G. T. & Marx, F. G. Assessing the quality of the fossil record: insights from vertebrates. *Geol. Soc. London Spec. Publ.* **358**, 63–94 (2011).
21. Costello, M. J., May, R. M. & Stork, N. E. Can we name earth's species before they go extinct? *Science* **339**, 413–416 (2013).
22. Kitchell, J. A. & Carr, T. R. in *Phanerozoic Diversity Patterns*. (ed. Valentine, J. W.) 277–309 (Princeton Univ. Press, 1985).
23. Benton, M. J. & Emerson, B. C. How did life become so diverse? The dynamics of diversification according to the fossil record and molecular phylogenetics. *Palaentology* **50**, 23–40 (2007).
24. Stanley, S. M. An analysis of the history of marine animal diversity. *Paleobiology* **33**, 1–55 (2007).
25. Alroy, J. *et al.* Effects of sampling standardization on estimates of Phanerozoic marine diversification. *Proc. Natl Acad. Sci. USA* **98**, 6261–6266 (2001).
26. Mayhew, P. J., Bell, M. A., Benton, T. G. & McGowan, A. J. Biodiversity tracks temperature over time. *Proc. Natl Acad. Sci. USA* **109**, 15141–15145 (2012).
27. Smith, A. B., Lloyd, G. T. & McGowan, A. J. Phanerozoic marine diversity: rock record modelling provides an independent test of large-scale trends. *Proc. R. Soc. B Biol. Sci.* **279**, 4489–4495 (2012).
28. Peters, S. E. Genus extinction, origination, and the durations of sedimentary hiatuses. *Paleobiology* **32**, 387–407 (2006).
29. Peters, S. E. & Heim, N. A. The geological completeness of paleontological sampling in North America. *Paleobiology* **36**, 61–79 (2010).
30. Peters, S. E. & Heim, N. A. Stratigraphic distribution of marine fossils in North America. *Geology* **39**, 259–262 (2012).
31. Benton, M. J., Ruta, M., Dunhill, A. M. & Sakamoto, M. The first half of tetrapod evolution, sampling proxies, and fossil record quality. *Palaeoogeogr. Palaeoclimatol. Palaeoecol.* **372**, 18–41 (2013).
32. Crampton, J. S. *et al.* Estimating the rock volume bias in paleobiodiversity studies. *Science* **301**, 358–360 (2003).
33. Dunhill, A. M., Benton, M. J., Newell, A. J. & Twitchett, R. J. Completeness of the fossil record and the validity of sampling proxies: a case study from the Triassic of England and Wales. *J. Geol. Soc. London.* **170**, 291–300 (2013).
34. Dunhill, A. M., Benton, M. J., Twitchett, R. J. & Newell, A. J. Completeness of the fossil record and the validity of sampling proxies at outcrop level. *Palaentology* **55**, 1155–1175 (2012).
35. Smith, A. B. & Benson, R. B. J. Marine diversity in the geological record and its relationship to surviving bedrock area, lithofacies diversity, and original marine shelf area. *Geology* **41**, 171–174 (2013).
36. Dunhill, A. M. Using remote sensing and a GIS to quantify rock exposure area in England and Wales: implications for paleodiversity studies. *Geology* **39**, 111–114 (2011).
37. Lloyd, G. T. & Friedman, M. A survey of palaeontological sampling biases in fishes based on the Phanerozoic record of Great Britain. *Palaeoogeogr. Palaeoclimatol. Palaeoecol.* **372**, 5–17 (2013).
38. Hannisdal, B. Non-parametric inference of causal interactions from geological records. *Am. J. Sci.* **311**, 315–334 (2011).
39. Hannisdal, B. Detecting common-cause relationships with directional information transfer. *Geol. Soc. London Spec. Publ.* **358**, 19–29 (2011).
40. Raup, D. M. Systematists follow the fossils. *Paleobiology* **3**, 328–329 (1977).
41. Rook, D. L., Heim, N. A. & Marcot, J. Contrasting patterns and connections of rock and biotic diversity in the marine and terrestrial fossil records of North America. *Palaeoogeogr. Palaeoclimatol. Palaeoecol.* **372**, 123–129 (2013).
42. Benton, R. B. J. & Butler, R. J. Uncovering the diversification history of marine tetrapods: ecology influences the effect of geological sampling biases. *Geol. Soc. London Spec. Publ.* **358**, 191–208 (2011).
43. Butler, R. J., Benson, R. B. J., Carrano, M. T., Mannion, P. D. & Upchurch, P. Sea level, dinosaur diversity and sampling biases: investigating the 'common cause' hypothesis in the terrestrial realm. *Proc. R. Soc. Lond. B Biol. Sci.* **278**, 1107–1113 (2011).
44. Dunhill, A. M. 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 (2012).
45. Holland, S. M. Sea level change and the area of shallow-marine habitat: implications for marine biodiversity. *Paleobiology* **38**, 205–217 (2012).
46. Erwin, D. H. A call to the custodians of deep time. *Nature* **462**, 282–283 (2009).
47. Kwiatkowski, D., Phillips, P. C. B. & Schmidt, P. Testing the null hypothesis of stationarity against the alternative of a unit root. *J. Econom.* **54**, 159–178 (1992).

Acknowledgements

This study benefitted from data compiled within The Paleobiology Database (www.paleodb.org) by numerous colleagues and is Paleobiology Database official publication 207. We thank Alistair McGowan and James Crampton for providing comments on a previous draft. A.M.D. is supported by a NERC doctoral training grant NE/H525111/1 and a Royal Commission for the Exhibition of 1851 Fellowship. B.H. is supported by the Bergen Research Foundation.

Author contributions

A.M.D. compiled the palaeodiversity data and compiled and analysed the spatial sampling proxy data using ArcGIS v10.1. B.H. performed the information transfer analysis and formation randomization tests. A.M.D., B.H. and M.J.B. designed the study and wrote the paper.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Dunhill, A. M. *et al.* Disentangling rock record bias and common-cause from redundancy in the British fossil record. *Nat. Commun.* **5**:4818 doi: 10.1038/ncomms5818 (2014).