Sea level, dinosaur diversity and sampling biases: investigating the 'common cause' hypothesis in the terrestrial realm

Richard J. Butler, Roger B. J. Benson, Matthew T. Carrano, Philip D. Mannion and Paul Upchurch

Proc. R. Soc. B published online 29 September 2010
doi: 10.1098/rspb.2010.1754

Supplementary data
"Data Supplement"

References
This article cites 35 articles, 22 of which can be accessed free
http://rspb.royalsocietypublishing.org/content/early/2010/09/24/rspb.2010.1754.full.html#ref-list-1

P<P
Published online 29 September 2010 in advance of the print journal.

Subject collections
Articles on similar topics can be found in the following collections
- palaeontology (132 articles)
- evolution (2213 articles)

Email alerting service
Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click here

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to Proc. R. Soc. B go to: http://rspb.royalsocietypublishing.org/subscriptions

This journal is © 2010 The Royal Society
Sea level, dinosaur diversity and sampling biases: investigating the ‘common cause’ hypothesis in the terrestrial realm

Richard J. Butler¹*, Roger B. J. Benson², Matthew T. Carrano³, Philip D. Mannion⁴ and Paul Upchurch⁴

¹ Bayerische Staatssammlung für Paläontologie und Geologie, Richard-Wagner-Straße 10, 80333 Munich, Germany
² Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK
³ Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, PO Box 37012, Washington, DC 20013-7012, USA
⁴ Department of Earth Sciences, University College London, Gower Street, London WC1E 6BT, UK

The fossil record is our primary window onto the diversification of ancient life, but there are widespread concerns that sampling biases may distort observed palaeodiversity counts. Such concerns have been reinforced by numerous studies that found correlations between measures of sampling intensity and observed diversity. However, correlation does not necessarily mean that sampling controls observed diversity: an alternative view is that both sampling and diversity may be driven by some common factor (e.g. variation in continental flooding driven by sea level). The latter is known as the ‘common cause’ hypothesis. Here, we present quantitative analyses of the relationships between dinosaur diversity, sampling of the dinosaur fossil record, and changes in continental flooding and sea level, providing new insights into terrestrial common cause. Although raw data show significant correlations between continental flooding/sea level and both observed diversity and sampling, these correlations do not survive detrending or removal of short-term autocorrelation. By contrast, the strong correlation between diversity and sampling is robust to various data transformations. Correlations between continental flooding/sea level and taxic diversity/sampling result from a shared upward trend in all data series, and short-term changes in continental flooding/sea level and diversity/sampling do not correlate. The hypothesis that global dinosaur diversity is tied to sea-level fluctuations is poorly supported, and terrestrial common cause is unsubstantiated as currently conceived. Instead, we consider variation in sampling to be the preferred null hypothesis for short-term diversity variation in the Mesozoic terrestrial realm.

Keywords: diversity; sea level; sampling; dinosaur; Mesozoic

1. INTRODUCTION

The fossil record offers our primary opportunity to quantify deep time evolutionary diversification. However, this record is unevenly sampled [1,2]: correlations between sampling metrics and observed palaeodiversity are frequently detected, leading many authors to suggest that diversity patterns cannot be read literally (e.g. [2–7]).

Palaeobiologists now frequently generate ‘sampling-corrected’ palaeodiversity curves, which may differ markedly from raw diversity curves [7–14]. However, corrections based on sampling must be applied cautiously. If genuine deep time diversity and our opportunities to sample that diversity are both driven by a common external factor, then the observed correlation between sampling metrics and diversity might not reflect causation. Valid concern over the existence of a third driving factor is termed the common cause hypothesis [3,5,6,12,15–17]; if it is true, then attempts to ‘correct’ palaeodiversity curves may actually distort genuine palaeodiversity signals. For example, sea-level changes are often proposed as an agent of common cause in the fossil record of shallow marine organisms: high sea level leads to the formation and the expansion of marine environments as a result of continental flooding, promoting increases in diversity as well as the accumulation of fossiliferous sediments and preservation of habitats [6,12,15,16]. By comparison, the mechanisms that might produce terrestrial common cause are poorly understood and little discussed [11,17]. Some workers have proposed that increased terrestrial surface area, resulting from the reduction in continental flooding associated with lower sea level, might lead to genuinely higher terrestrial biodiversity [17–19], as well as greater accumulation of terrestrial sediments [4,7], and thus more opportunities to sample palaeodiversity (electronic supplementary material, figure S1). Other authors have suggested that high sea level and increased continental flooding might lead to increased environmental heterogeneity and endemism in the terrestrial realm, generating higher biodiversity [20–22], with sampling also potentially increasing owing to the enhanced preservation of terrestrial fossils in shallow marine and coastal environments.

* Author for correspondence (r.butler@lrz.uni-muenchen.de).

2. MATERIAL AND METHODS

(a) Choice of taxonomic methods

Dinosaurs were the dominant elements of global terrestrial faunas for much of the Mesozoic [25]. Their fossil record is exceptionally well studied, encompasses a wide range of body sizes and is associated with accurate geographical and stratigraphic data ([25]; The Paleobiology Database, http://paleodb.org, hereafter PBDB). Thus, dinosaurs provide an excellent case study of sampling biases in the terrestrial realm [10,14,21,26–28]. Many authors have hypothesized correlations between dinosaur diversity and sea-level fluctuations [18–21,23,29,30], but quantitative tests have only been carried out for sauropodomorphs [14,28]. We do not consider dinosaur subclades (e.g. Ornithischia) separately in this contribution because our aim is to assess sampling biases and the terrestrial common cause hypothesis at the broadest scale currently possible. The evolutionary histories of taxonomically restricted groups are the subject of ongoing research.

(b) Time bins

Standard European stages and the absolute dates provided by Gradstein et al. [31] were used as the time bins for compiling data series. To examine the effect of variable time bin duration, we first assessed statistical correlation between bin length and taxic diversity, and bin length and geological sampling. Subsequently, for pairwise statistical comparisons between data series that are both potentially biased by unequal bin length (taxic diversity, sampling and proxy for continental flooding), we calculated first-order partial correlations in which the influence of bin length is removed. In general, uneven bin lengths do not impact substantially upon the results presented here (electronic supplementary material).

(c) Sources of data

Total non-marine surface area, which is inversely linked to continental flooding, was taken from the palaeogeographic maps of Smith et al. [32]. As a proxy for continental flooding, we used the data of Peters [6,16] and Peters & Heim [33] (electronic supplementary material), which record temporal variation in the number of marine gap-bound sedimentary packages within North America; although this compilation is a regional one, it undoubtedly has a global component. Mesozoic sea-level estimates were drawn from Miller et al. [34] (electronic supplementary material), who provided two data series: one for the curve of Haq et al. [35] covering the time period of 0–244 Ma and a novel one spanning 0–172 Ma. Because the points within these data series are not distributed evenly in time, we interpolated equally spaced data points at 0.1 Myr intervals. We then calculated the mean sea level for each of our time bins.

Stratigraphic occurrence data were collected for 749 valid non-avian dinosaur species (electronic supplementary material), representing the largest dataset of Mesozoic terrestrial animals yet compiled, and used to calculate a taxic diversity estimate (TDE). Sauropodomorph data were derived from Mannion et al. [14]. Data for ornithischians and theropods were taken from the PBDB (downloaded 2 February 2010; data compiled primarily by M.T.C.).

Data on temporal variation in sampling were taken from the PBDB (electronic supplementary material). Counts of distinct dinosaur-bearing collections (DBCs) or localities (bin counts range from 41 to 1405) and dinosaur-bearing formations (DBFs; bin counts range from 19 to 163) were based on all non-avian dinosaur records. We also compiled PBDB data on temporal variation in the proportion of total DBCs known from ‘marine’ environments (we did not compile an equivalent data series for DBFs because some formations contain both marginal marine and terrestrial horizons). Marine environments include both fully marine and marginal or coastal environments (e.g. deltaic, estuarine and lagoon).
using a residuals method [1,7]. This method calculates a modelled diversity estimate that represents the diversity expected if observed diversity variations result solely from the correcting factor. Diversity residuals (i.e. the differences between modelled diversity values and actual diversity values) following correction for sampling were subsequently compared statistically with sea level, while diversity residuals following correction for sea level were subsequently compared with sampling.

3. RESULTS
The runs tests indicate that most of the data series are non-random, and the presence of cyclicity (typically at lags of 5–9 time bins) and Late Triassic–Cretaceous trend (of increase in most data series, but decrease in non-marine surface area; figure 1) is confirmed by visual inspection of correlograms (electronic supplementary material) and of the raw data series plotted against time.

Raw taxic diversity counts for non-avian dinosaurs are significantly positively correlated with sea level and continental flooding, and negatively correlated with non-marine area in most cases (some correlations are rendered non-significant by corrections for multiple tests), regardless of which sea-level curve is considered and whether the effect of bin length is removed by partial correlations (table 1; electronic supplementary material). Diversity residuals following correction of taxic diversity counts for sampling are not significantly correlated with sea level (electronic supplementary material, table S1). By contrast, diversity residuals following correction of taxic diversity counts for sea level do show significant correlations with sampling in most cases, although only when transformations are used (electronic supplementary material, table S1).

4. DISCUSSION AND CONCLUSIONS
Our results are consistent with those of Barrett et al. [10] and Mannion et al. [14] in recovering tight correlations between observed dinosaur taxic diversity and proxies for geological sampling and collector effort (DBC and DBF) are significantly positively correlated with sea level and continental flooding, and negatively correlated with non-marine area in most cases, but these correlations are also rendered non-significant by transformations (table 1; electronic supplementary material, figure S2c). No significant results were obtained from cross-correlation of sea level against diversity or sampling (electronic supplementary material). Furthermore, no significant correlation was recognized between the proportion of the dinosaur fossil record collected from marine and/or coastal depositional environments and sea level or continental flooding (table 1). By contrast, a strongly significant positive correlation between sampling and taxic diversity is recovered in all cases, regardless of how the data are transformed (table 1; electronic supplementary material, figure S2b).

Figure 1. Raw data series plotted against geological time. (a) Number of marine gap-bound sedimentary packages within North America (black dashed line; left y-axis shows number of packages), Haq et al. (solid grey line) and Miller et al. (dotted-dashed line) sea-level curves (right y-axis shows metres below or above present day sea level), species level dinosaur taxic diversity (solid black line; right y-axis shows numbers of valid species), and non-marine surface area (grey dashed line; right y-axis shows area in 106 km2). (b) Species-level dinosaur taxic diversity (solid black line), number of dinosaur-bearing formations (dashed line) and number of dinosaur-bearing collections (solid grey line). Data series in (b) are log10 transformed.
Table 1. Summary of values for Pearson's product–moment coefficient for comparisons of key data series used in this study. Values above the diagonal represent correlations for raw data series; values below the diagonal represent correlations for generalized difference data series. Values in bold are statistically significant after correction for multiple comparisons; values marked with an asterisk were significant only prior to application of corrections for multiple statistical tests ($\alpha = 0.05$). $p$-values are not reported here for correlations between most raw data series because these are heavily biased by temporal autocorrelation (see text). Statistical results using first differenced and detrended data series, and using Spearman’s rank and Kendall’s $\tau$ coefficients, are presented in the electronic supplementary material. DBCs, dinosaur-bearing collections; DBFs, dinosaur-bearing formations; TDE, taxic diversity estimate.

<table>
<thead>
<tr>
<th></th>
<th>Haq sea level</th>
<th>Miller sea level</th>
<th>continental flooding</th>
<th>log TDE</th>
<th>log DBFs</th>
<th>log DBCs</th>
<th>continental area</th>
<th>% marine DBCs</th>
<th>log bin lengths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haq sea level</td>
<td></td>
<td></td>
<td>0.93</td>
<td>0.80</td>
<td>0.72</td>
<td>0.78</td>
<td>-0.82</td>
<td>-0.36</td>
<td>n.a.</td>
</tr>
<tr>
<td>Miller sea level</td>
<td>0.59 ($p = 0.013^*$)</td>
<td>0.81</td>
<td>0.60*</td>
<td>0.66*</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>continental flooding</td>
<td>0.02 ($p = 0.94$)</td>
<td>-0.09 ($p = 0.73$)</td>
<td>0.67</td>
<td>0.78</td>
<td>0.65</td>
<td>-0.77</td>
<td>-0.29</td>
<td>0.33 ($p = 0.1$)</td>
<td></td>
</tr>
<tr>
<td>log TDE</td>
<td>0.13 ($p = 0.55$)</td>
<td>0.22 ($p = 0.39$)</td>
<td>0.34 ($p = 0.1$)</td>
<td>0.87</td>
<td>0.89</td>
<td>-0.61</td>
<td>n.a.</td>
<td>0.21 ($p = 0.21$)</td>
<td></td>
</tr>
<tr>
<td>log DBFs</td>
<td>0.04 ($p = 0.86$)</td>
<td>-0.23 ($p = 0.38$)</td>
<td>0.35 ($p = 0.08$)</td>
<td>0.69 ($p = 0.0001$)</td>
<td>0.94</td>
<td>-0.73</td>
<td>n.a.</td>
<td>0.24 ($p = 0.25$)</td>
<td></td>
</tr>
<tr>
<td>log DBCs</td>
<td>-0.0002 ($p = 0.99$)</td>
<td>-0.01 ($p = 0.98$)</td>
<td>0.20 ($p = 0.34$)</td>
<td>0.82 ($p = 4.3 \times 10^{-7}$)</td>
<td>0.85 ($p = 6.6 \times 10^{-7}$)</td>
<td>-0.62</td>
<td>n.a.</td>
<td>0.28 ($p = 0.17$)</td>
<td></td>
</tr>
<tr>
<td>continental area</td>
<td>-0.10 ($p = 0.63$)</td>
<td>-0.36 ($p = 0.16$)</td>
<td>0.05 ($p = 0.81$)</td>
<td>0.14 ($p = 0.50$)</td>
<td>0.23 ($p = 0.26$)</td>
<td>0.14 ($p = 0.50$)</td>
<td>n.a.</td>
<td>n.a.</td>
<td></td>
</tr>
</tbody>
</table>
hypothesized mechanisms for terrestrial common cause. The most prominent suggests that both terrestrial diversity and sampling should be higher when sea level is low and non-marine surface area is greatest [17–19]. Although we have only considered a single taxonomic group, it is clear that observed dinosaur diversity does not conform to these predictions: observed diversity is highest during a time of relatively high sea level (Late Cretaceous), high continental flooding and low non-marine area. Indeed, when terrestrial tetrapods as a whole are considered (at coarse taxonomic levels), a similar trend towards higher observed diversity through the Mesozoic can be clearly recognized [39,40]. As a result, observations of trends in observed terrestrial diversity during the Mesozoic run counter to at least one hypothesis of how terrestrial common cause might work. An alternative version of common cause that fits the observed diversity patterns more closely would suggest that diversity should be highest at times of high sea level owing to increased habitat fragmentation and endemism [20–22], and that sampling should also increase owing to the enhanced preservation of terrestrial taxa in coastal and shallow marine settings [23]. However, the absence of significant correlation between the proportion of the dinosaur fossil record collected from coastal/marine deposits and sea-level fluctuations fails to support this hypothesis (for example, the proportion of the dinosaur fossil record known from coastal/marine deposits is much lower in the Late Cretaceous than in the Middle–Late Jurassic). Moreover, we note that dinosaur fossils are generally scarce in marine depositional environments (less than 20% of PBDB collections in most time bins; see electronic supplementary material), and that high sea level could actually decrease the proportion of sediment deposited in many coastal terrestrial settings, because sediments borne by rivers may be more likely to be carried out into the marine realm owing to overall shortening of the depositional system (e.g. [41,42]).

Substantial recent advances have produced global curves of marine invertebrate diversity that are standardized for uneven sampling through time [9,12]. Similar standardization work for terrestrial vertebrates is in its infancy, has been restricted to studies of individual clades [8,14,26,27], and is hampered by the current absence of comprehensive global databases of terrestrial vertebrate diversity and sampling at the genus or species level. As a result, the broad picture of diversity patterns for terrestrial vertebrates through the Phanerozoic continues to be read at face value by many authors (e.g. [17,43]), despite mounting evidence that sampling biases may play a profound role in influencing observed terrestrial diversity patterns [10,26,27,44]. Such evidence cannot be adequately explained away as a product of common cause, because current theoretical hypotheses for how terrestrial common cause might work are poorly constrained and not supported by the empirical data presented here.

Considerable future work is required to establish how sampling biases may affect proposed long-term diversity trends and mass extinction events in the terrestrial realm. Our results emphasize the extremely tight link between short-term fluctuations in terrestrial diversity and sampling, and fail to support current sea-level-driven hypotheses of terrestrial common cause. As a result, we hold that sampling biases should be regarded as the null hypothesis for explaining short-term fluctuations in observed diversity in the Mesozoic terrestrial realm. Finally, we recommend that macroevolutionary work that considers such short-term fluctuations in sampling, and correct for their effects on raw diversity counts, whenever possible.

This study is Palaeobiology Database official publication 121. We thank Paul Barrett for useful discussion, and Stephen Brusatte, Graeme Lloyd and Alistair McGowan for comments on an earlier version of the manuscript. Shanan Peters and an anonymous referee provided helpful reviews. R.J.B. is supported by an Alexander von Humboldt Research Fellowship.

REFERENCES