Supporting Information

Mayhew et al. 10.1073/pnas.1200844109

SI Text

Supporting Methods. ***First dataset.*** Estimated low latitude shallow-sea temperatures came from the red curve of figure 4 in ref. 1, provided courtesy of Dana Royer. They derive from a measure of the ratio of ^{18}O: ^{16}O stable isotopes ($\delta^{18}O$) from mainly low latitude shallow-sea calcitic and phosphatic shells, mainly brachiopods but also conodont elements and belemnites (2). The temperature measures incorporate a correction for the effect of seawater pH on the $\delta^{18}O$ of carbonates (1), which were first detrended and then averaged in 10 Myr timesteps using a 50 Myr moving window (1). Atmospheric CO_2 concentrations (RCO_2 = ratio of mass of CO_2 at time t to that at present), as used in ref. 1, were also provided courtesy of Dana Royer. They derive from the GeoCarb-III model (3) and were used (1) as one of the driving variables to estimate changes in seawater pH (above). These 10 Myr interval data were based on an old time scale (4) not used by the other variables in the dataset hence were rescaled to new estimated dates based on a newer time scale (5) using linear interpolation across the interval boundaries.

Data on the richness, origination, and extinction of marine invertebrates per stage were taken from the stage level global compendium of Sepkoski (6), which records the first and last occurrences of genera and assumes that taxa occur in all intervening stages ("range-through"). The measures used ignore single interval taxa (singletons) and are therefore more robust than simply summing all taxa within a bin to variations in interval duration and preservation (7). The richness measure for each stage simply ignores single interval taxa, while the origination and extinction rates, p and q in ref. 7, are instantaneous per capita rates (per million years) that only use information from taxa that cross stage boundaries.

Sample standardized data on marine invertebrate richness, evenness, origination, and extinction rates were provided courtesy of John Alroy. Richness measures came from two methods: Item quota subsampling (IQS) (8) and the shareholder quorum subsampling (SQS) (9). The data derive from fossil collections data stored in the Paleobiology Database. IQS is based on the richness of taxa in a given sample quota of specimens drawn from a random subset of available collections. IQS is used here (figure 1 in ref. 8) not only to assess sensitivity of the results to different richness measures but also because it has been used to demonstrate some other important features of paleodiversity relevant to our analyses (10, 11). However IQS tends to underestimate richness in diverse communities because it is harder to randomly sample rare taxa with standard rarefaction. SQS, however, better tracks the true species pool size (12) by counting richness when the combined number of occurrences ("shares") of the resampled taxa ("shareholders") in a given time interval reaches a fixed, required proportion of all occurrences in that time interval ("shareholder quorum"). It therefore samples unevenly but more fairly across intervals (12). The evenness of local collections in each time bin (figure 2 in ref. 8) was derived from the number of genera per 100 specimens in local collections (8). The median number of genera was used to derive the slope of a log-log plot of genera against specimens, from which the evenness measure [Hurlbert's probability of interspecific encounter (12)] derives (8).

The extinction rate μ is the exponential decay rate of a cohort crossing the base of a time bin and continuing to its top, corrected for the fact that members of this cohort may be present but not sampled in the following, third, bin (10). The corresponding origination rate is λ (10). By concentrating on taxa actually sampled within a time bin and correcting for variation in sampling probability, these measures are less susceptible to so-called edge effects, such as the Signor–Lipps effect (which postulates that the chance that the observed last occurrence is the true one is vanishingly small, and thus extinction events are shifted backward in time). A similar process occurs with the forward shifting of the timing of the origination of taxa (10). The sample standardized fossil record data were compiled for 48 intervals averaging 11 Myr from Early Cambrian through to Neogene (8–10).

To apply modeling approaches to correct traditional range-through fossil data (i.e., those based on Sepkoski's compendium above) for sampling probability across geological stages, we used data on the estimated area of marine sedimentary rock as a proxy control variable, estimated from geological survey maps according to ref. 13. Data were compiled into the same 11 Myr intervals as for the above data. We used as control variables data from Europe alone, a well-sampled area that likely reflects the global fossil sampling effort, Australia alone, as a continent with a different sedimentary rock area profile (13), and the sum of Europe and Australia. The rock record measures were used in four ways. First, we detrended and transformed the measures (see below) and included them as explanatory variables in linear models of the richness of Sepkoski boundary crossers (see below). The other three approaches used the rock record measures to predict taxonomic richness in Sepkoski's data under three sets of assumptions (13). In the first set of assumptions, richness was assumed to be constant (Model I), while rock record varied. Richness and rock area values were logged, ranked individually, and the linear regression performed, giving an equation predicting richness from rock area. The observed values of rock area were then used to predict richness and the difference between the observed and predicted richness values taken as the corrected dataset for Model I. These corrected data were then modified in two further models (II and III). In Model II, a linear regression of the corrected data through all Phanerozoic time was made and the value of the slope of the line was added to each time bin to reflect the predicted long-term increase or decrease. The initial value was given by setting x to zero in the equation and this left the constant as the starting value and then the diversity value would increase or decrease by the value of the slope of the line for each time interval. This correction factor would be added to predicted diversity from Model I. In Model III, a three-phase diversification model was assumed by taking the residuals from three separate linear trends of Model I corrected richness through time: A Cambrian to mid-Devonian (Eifelian) rise, mid-Devonian (Giventian) to Triassic (Rhaetian) fall, and Early Jurassic (Hettangian) to Pliocene rise (13). Equations used were as given in figure 5*B* of ref. 14: Phase I: $y = 0.017x - 0.152$; Phase II: $y = -0.030x - 0.145$; Phase III: $y = -0.029x - 0.554$.

It is preferable, for time series analysis, to have data for all variables at equally spaced sampling intervals, and to achieve this in a way that most closely matched previous analyses of temperature and the fossil record (15), all data were interpolated with Akima interpolation splines (16) using the akima function R (17) to extract values for 10 Myr intervals.

Second dataset. Data on eustatic sea level (18–20) and the marine isotopic record of $\delta^{18}O$ (2), $\delta^{13}C$ (2), $^{87}Sr/^{86}Sr$ (21), and $\delta^{34}S$ (22) came from the compilation of ref. 23 at 11-Myr intervals corresponding closely to the time bins used in the sample standardized fossil record data mentioned above (9–11). The isotopic records of $\delta^{13}C$, $^{87}Sr/^{86}Sr$, and $\delta^{34}S$ came from a greater range of sources to $\delta^{18}O$ (mainly brachiopods and belemnites), including both biogenic (e.g., planktonic microfossils) and in the case of $\delta^{34}S$, abiogenic material (micritic carbonates). Long-term sea level changes are mostly inferred from stratigraphic sections of cratons indicating flooding events. The chief merit in using isotopic environmental variables in macroevolutionary analyses is that they are relatively unmodified datasets hence less prone to error than, for example, direct estimates of temperature. A drawback is that changes in these variables probably reflect not just changes in one feature of the environment but several (23), so we have less confidence that an individual change reflects the environmental feature we are interested in.

To these environmental variables we added all the sampled standardized fossil data (IQS, SQS, evenness, λ and μ) from the first dataset as well as the sea-water temperature estimates. To ensure that all variables had equally spaced sampling intervals, we used Akima splines again to interpolate values to the same 11-Myr intervals.

Analyses. Datasets were detrended using smoothing splines (24) to remove long term trends while retaining shorter term variability, using the function smooth.spline in R. The removal of long term trends is essential to isolate the time scale of interest (fluctuations over 10 s of Myr) and because inclusion of 100-Myr trends would cause spurious correlations between variables. As different datasets have different long term trends, a variety of different spline curvatures were applied to each dataset, set by the degrees of freedom (d. f.) of the spline, where greater d. f. allow the spline to track shorter term variability in the data (24). The best detrender was assessed following examination of autocorrelation and spectral plots of the original series and the detrended residuals to judge the efficiency with which the smoother was isolating patterns at the time scale of interest and the sensitivity of the residual variability to different detrenders. Typically a 5 d.f. spline effectively removed long-term patterns while retaining shorter term patterns (Table S1). The detrended series were analyzed using linear modeling (see below), so it is important to reduce bias in parameter estimates by ensuring that the detrended series were roughly symmetrical around the mean. Where necessary, a data transformation (typically log or square root) to reduce skew was applied to the original data prior to detrending to normalize the residuals, and occasionally a transformation was applied after detrending instead or as an additional step (Table S1). All detrended data were then standardized to a mean of zero and unit standard deviation. Standardizing the detrended series assists interpretation of the time series plots and subsequent analytical coefficients because all variables use the same scale.

Analyses included pairwise Pearson correlation, linear regression, and multiple regression using stepwise subtraction. For the latter, model simplification was performed by stepwise removal from a full model containing main effects (both datasets) and interactions (first dataset only), using the step function in R. Step removes parameters based on a comparison of Akaike Information Criterion (AIC) values (25) of all possible models with one less term, removing the term that leads to the greatest reduction in AIC at each step. The AIC (25) is a measure of the goodness of fit of a statistical model, describing the trade-off between model accuracy and model complexity, designed to discourage overfitting. Low AIC values represent a favorable trade-off (better accuracy for a given complexity).

Because the time series are serially auto-correlated, standard statistical tables will generally overestimate significance because they assume independence of the data. We therefore tested significance (experiment-wise) through bootstrapping the data to directly estimate confidence limits of the test statistics, using the function boot in R. Rows of data (x, y pairs in a correlation) are sampled with replacement from the true data to create a new pseudo-dataset of the same size as the original. The statistical test is applied and test statistic stored. This process is repeated many times (typically 1,000) to produce a distribution of the test statistic that illustrates the way the statistic may change with changing the sample, within the observed limits of the data. This distribution can then be used to calculate confidence intervals on the test statistic. We used the bias corrected and accelerated (bca) technique (26) for calculating confidence limits, which corrects for the bias (difference between the observed mean and bootstrap mean) and asymmetry of the bootstrap distribution.

Conducting many statistical tests increases the risk of making a Type I error somewhere within the test family. Given the large number of tests presented in this paper, readers should exercise caution before rejecting any particular null hypothesis. However, our overall conclusions are not substantially affected if any one test is erroneous (27), nor is it likely that the majority are: In the main text and Table S3 we report 42 hypothesis tests involving temperature or a temperature proxy; 24 of them are significant, >11 times the family-wise error rate of 2 at $P < 0.05$. In Table S3, 80 tests are reported, of which 25 are significant (i.e. >6.25 times the family-wise error rate). At least two of the tests, involving temperature in Table S3, are robust even to a strict Bonferroni correction at $P < 0.000625$ (see *Results*). The majority of tests were implemented to explore the robustness of our initial findings to alternative assumptions, so consistent findings should make them more robust, not less (28).

1. Royer DL, Berner RA, Montañez IP, Tibor NJ, Beerling DJ (2004) CO_2 as a primary driver of Phanerozoic climate. *GSA Today* 14:4–10.
2. Veizer J, et al. (1999) $^{87}Sr/^{86}Sr$, $\delta^{13}C$ and $\delta^{18}O$ evolution of Phanerozoic seawater. *Chem Geol* 161:59–88.
3. Berner RA, Kothavala Z (2001) GEOCARB III: A revised model of atmospheric CO_2 over Phanerozoic time. *Am J Sci* 301:182–204.
4. Harland WB, et al. (1990) *A Geologic Time Scale 1989*. (Cambridge Univ Press, Cambridge, UK).
5. Gradstein FM, Ogg JG (2004) *A Geologic Time Scale 2004*. (Cambridge Univ Press, Cambridge, UK).
6. Sepkoski JJ, Jr. (2002) A compendium of fossil marine animal genera. *Bull Am Paleontol* 363:1–560.
7. Foote M (2000) Origination and extinction components of taxonomic diversity: General problems. *Paleobiology* 26:74–102.
8. Alroy J, et al. (2008) Phanerozoic trends in the global diversity of marine invertebrates. *Science* 321:97–100.
9. Alroy, J (2010) Geographical, environmental and intrinsic biotic controls on Phanerozoic marine diversification. *Palaeontology* 53:1211–1235.
10. Alroy J (2008) Dynamics of origination and extinction in the marine fossil record. *Proc Natl Acad Sci USA* 105:11536–11542.
11. Melott AL (2008) Long-term cycles in the history of life: Periodic biodiversity in the Paleobiology database. *PLoS One* 3:e4044.
12. Alroy J (2010) The shifting balance of diversity among major marine animal groups. *Science* 329:1191–1194.
13. McGowan AJ, Smith AB (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.
14. Smith AB, McGowan AJ (2007) The shape of the marine palaeodiversity curve using the Phanerozoic sedimentary rock record of western Europe. *Palaeontology* 50:765–774.
15. Mayhew PJ, Jenkins GB, Benton, TG (2008) A long-term association between global temperature and biodiversity, origination and extinction in the fossil record. *Proc R Soc B* 275:47–53.
16. Akima H (1970) A new method of interpolation and smooth curve fitting based on local procedures. *J Assoc Comput Mach* 17:589–602.
17. R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria).
18. Haq BU, Schutter SR (2008) A chronology of Paleozoic sea-level changes. *Science* 322:64–68.
19. Haq BU, Hardenbol J, Vail PR (1987) Chronology of fluctuating sea levels since the Triassic. *Science* 235:1156–1167.
20. Miller KG et al. (2005) The Phanerozoic record of global sea-level change. *Science* 310:1293–1298.
21. McArthur JM, Howarth RJ, Bailey TR (2001) Strontium isotope stratigraphy: LOWESS version 3: Best fit to the marine Sr-isotope curve for 0–509 Ma and accompanying look-up table for deriving numerical age. *J Geol* 109:155–170.
22. Kampschulte A, Strauss H (2004) The sulfur isotopic evolution of Phanerozoic seawater based on the analysis of structurally substituted sulfate in carbonates. *Chem Geol* 204:255–286.

23. Cárdenas AL, Harries PJ (2010) Effect of nutrient availability on marine origination rates throughout the Phanerozoic eon. *Nat Geosci* 3:430–434.
24. Fewster RM, Buckland ST, Siriwardena GM, Bailie SR, Wilson JD (2000) Analysis of population trends for farmland birds using generalized additive models. *Ecology* 81:1970–1984.
25. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19:716–723.
26. Efron B (1987) Better bootstrap confidence intervals. *J Am Stat Assoc* 82:171–185.
27. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Statist Soc B* 57:289–300.
28. Rothman KJ (1995) No adjustments are needed for multiple comparisons. *Epidemiology* 1:43–46.

Table S1. Transformations and detrenders applied to each variable prior to mean-standardizing

Dataset	Variable	Transformation	Detrender spline (d.f.)
First dataset	Generic richness (Sepkoski boundary crossers)	Square root	5
	Generic richness (IQS)	Square root	5
	Generic richness (SQS)	$\log_e$	5
	Evenness	None	5
	Origination rates (Sepkoski *p*)	$\log_e$	5
	Origination rates (λ)	$\log_e$	5
	Extinction rates (Sepkoski *q*)	$\log_e$, residuals also $\log_e$	5
	Extinction rates (μ)	$\log_e$	5
	Temperature	None	5
	CO_2	Square root	5
	Rock area (Europe)	None	9
	Rock area (Australia)	Square root	9
	Rock area (Summed)	None	9
	O-P (Europe, Mod. I)	Square root	5
	O-P (Australia, Mod. I)	None	5
	O-P (Summed, Mod. I)	$\log_e$	7
	O-P (Europe, Mod. II)	None, residuals square rooted	5
	O-P (Australia, Mod. II)	None	5
	O-P (Summed, Mod. II)	$\log_e$, residuals logged	5
	O-P (Europe, Mod. III)	$\log_e$	3
	O-P (Australia, Mod. III)	None	5
	O-P (Summed, Mod. III)	None	None
Second dataset	Generic richness (IQS)	$\log_e$, residuals also $\log_e$	5
	Generic richness (SQS)	$\log_e$	5
	Evenness	$\log_e$	5
	Origination rates (λ)	$\log_e$	5
	Extinction rates (μ)	$\log_e$	5
	$\delta^{18}O$	None	5
	$\delta^{16}C$	$\log_e$	5
	$^{87}Sr/^{86}Sr$	None	5
	$\delta^{34}S$	$\log_e$	5
	Eustatic sea level	None	7
	Temperature	None	3

Table S2. Associations between traditional fossil record measures, or such measures that have been corrected for the rock record, and other variables (after transformation and detrending)

Response variable	Explanatory variables	Method	Coefficient	Upper 95%CI	Lower 95%CI
Generic richness (Sepkoski boundary crossers)	Temperature, 10 Myr previously	Correlation	−0.275	−0.424	−0.017
Generic richness (Sepkoski boundary crossers)	CO_2 10 Myr previously	Correlation	−0.363	−0.502	−0.195
Generic richness (Sepkoski boundary crossers)	Temperature, 10 Myr previously	Linear Model	−0.272	−0.487	−0.007
Generic richness (Sepkoski boundary crossers)	CO_2	Linear Model	−0.351	−0.569	−0.176
Generic richness (Sepkoski boundary crossers)	Temperature, European rock area as co-variate	Linear Model	−0.467	−0.757	−0.228
Generic richness (Sepkoski boundary crossers)	Temperature, Australian rock area as co-variate	Linear Model	−0.447	−0.742	−0.170
Generic richness (Sepkoski boundary crossers)	Temperature, summed rock area as co-variate	Linear Model	−0.471	−0.784	−0.221
O-P richness (Europe, Mod. I)	Temperature	Correlation	−0.103	−0.281	+0.141
O-P richness (Australia, Mod. I)	Temperature	Correlation	−0.265	−0.291	−0.057
O-P richness (summed, Mod. I)	Temperature	Correlation	−0.283	−0.286	−0.029
O-P richness (Europe, Mod. II)	Temperature	Correlation	−0.104	−0.223	+0.207
O-P richness (Australia, Mod. II)	Temperature	Correlation	−0.194	−0.250	+0.053
O-P richness (summed, Mod. II)	Temperature	Correlation	−0.098	−0.161	+0.191
O-P richness (Europe, Mod. III)	Temperature	Correlation	−0.329	−0.596	−0.256
O-P richness (Australia, Mod. III)	Temperature	Correlation	−0.158	−0.229	+0.129
O-P richness (summed, Mod. III)	Temperature	Correlation	−0.190	−0.425	−0.008
O-P richness (Europe, Mod. I)	Temperature, 10 Myr previously	Correlation	−0.329	−0.514	−0.163

Response variable	Explanatory variables	Method	Coefficient	Upper 95%CI	Lower 95%CI
O-P richness (Australia, Mod. I)	Temperature, 10 Myr previously	Correlation	−0.229	−0.201	+0.017
O-P richness (summed, Mod. I)	Temperature, 10 Myr previously	Correlation	−0.215	−0.232	+0.087
O-P richness (Europe, Mod. II)	Temperature, 10 Myr previously	Correlation	−0.341	−0.454	−0.069
O-P richness (Australia, Mod. II)	Temperature, 10 Myr previously	Correlation	−0.163	−0.254	+0.113
O-P richness (summed, Mod. II)	Temperature, 10 Myr previously	Correlation	−0.303	−0.356	+0.015
O-P richness (Europe, Mod. III)	Temperature, 10 Myr previously	Correlation	−0.401	−0.592	−0.284
O-P richness (Australia, Mod. III)	Temperature, 10 Myr previously	Correlation	−0.134	−0.230	+0.152
O-P richness (summed, Mod. III)	Temperature, 10 Ma previously	Correlation	−0.262	−0.445	−0.004
Origination rates (*p*)	Temperature	Correlation	+0.262	+0.469	−0.029
Extinction rates (*q*)	Temperature	Linear Model	+0.271	+0.550	−0.075

Table S3. Bivariate correlations (Pearson's *r*, **P* < 0.05 experiment-wise) between the standardized measures of marine invertebrate diversity, origination and extinction, and biotic and abiotic isotopic predictors

Variable	Standing richness (IQS)	Stand ing richness (SQS)	Even-ness	Origination rate (λ)	Extinction rate (μ)
Temperature	**0.466***	0.252	**0.414***	**0.419***	**0.361***
Previous temperature	0.221	0.007	0.285	0.220	**0.472***
$\delta^{18}O$	**−0.296***	−0.185	**−0.360***	−0.223	−0.139
Previous $\delta^{18}O$	−0.245	−0.041	**−0.520***	−0.241	−0.219
$\delta^{13}C$	+0.181	+0.192	−0.021	−0.096	+0.035
Previous $\delta^{13}C$	+0.127	**+0.388***	−0.216	+0.030	+0.031
$^{87}Sr/^{86}Sr$	−0.165	+0.170	**−0.408***	**+0.308***	+0.095
Previous $^{87}Sr/^{86}Sr$	−0.142	+0.111	**−0.363***	+0.008	−0.167
$\delta^{34}S$	−0.029	−0.139	+0.014	**+0.451***	+0.275
Previous $\delta^{34}S$	+0.059	+0.098	−0.163	**+0.395***	+0.203
Eustatic sea level	+0.082	+0.079	+0.071	−0.185	**−0.360***
Previous eustatic sea level	**+0.347***	+0.162	+0.159	−0.088	+0.237
Previous extinction rate, μ	**−0.461***	**−0.362***	−0.123	**+0.354***	
Previous standing richness (IQS)		**+0.304***	**+0.471***	−0.074	**+0.485***
Previous origination rate, λ	+0.202	+0.199	+0.085		−0.007
Previous standing richness (SQS)	**+0.501***		+0.145	−0.182	+0.157
Previous Evenness	**+0.356***	+0.139		+0.278	**+0.422***

IQS = item quota subsampling; SQS = shareholder quorum subsampling.

Other Supporting Information Files

Dataset S1 (XLSX)