The fossil record and spatial structuring of environments and biodiversity in the Cenozoic of New Zealand

James S. Crampton, Michael Foote, Roger A. Cooper, et al.

Geological Society, London, Special Publications 2011; v. 358; p. 105-122
doi: 10.1144/SP358.8

The fossil record and spatial structuring of environments and biodiversity in the Cenozoic of New Zealand

JAMES S. CRAMPTON[1]*, MICHAEL FOOTE[2], ROGER A. COOPER[1], ALAN G. BEU[1] & SHANAN E. PETERS[3]

[1]*GNS Science, PO Box 30368, Lower Hutt, New Zealand*

[2]*Department of the Geophysical Sciences, The University of Chicago, 5734 South Ellis Avenue, Chicago, Illinois 60637, USA*

[3]*Department of Geology & Geophysics, University of Wisconsin-Madison, 1215 W. Dayton St., Madison, Wisconsin 53706, USA*

**Corresponding author (e-mail: j.crampton@gns.cri.nz)*

Abstract: There is increasing evidence to suggest that drivers of bias in the fossil record have also affected actual biodiversity history, so that controls of artefact and true pattern are confounded. Here we examine the role of spatial structuring of the environment as one component of this common cause hypothesis. Our results are based on sampling standardized analyses of the post-Middle Eocene record of shelf molluscs from New Zealand. We find that spatial structuring of the environment directly influenced the quality of the fossil record. Contrary to our expectations, however, we find no evidence to suggest that spatial structuring of the environment was a strong or direct driver of taxic rates, net diversity, or spatial structuring in mollusc faunas at the scale of analysis. Stage-to-stage variation in sampling standardized diversity over the past 40 Ma exhibits two superficially independent dynamics: (a) changes in net diversity were associated primarily with changes in origination rate; and (b) an unknown common cause related extinction rate to the quality of the fossil record and, indirectly, to spatial structuring of the environment. We suggest that tectonic drivers, manifest as second-order sequence stratigraphic cycles, are likely to have been a key element of this common cause.

Our perceptions of biodiversity history are influenced strongly by secular biases in the quality of the fossil record (e.g. Raup 1976; Peters & Foote 2001). Many authors have argued against a literal reading of the record, and various techniques to minimize the impacts of sampling biases have been developed (e.g. Connolly & Miller 2001; Foote 2003; Alroy 2010). In addition, however, a number of studies have proposed that the relationship between sampling biases and diversity dynamics may be complex. These studies have suggested that there are latent common-cause factors, such as plate tectonics and sea-level change, which have affected simultaneously the quality of the fossil record and *true* macroevolutionary history (e.g. Newell 1952; Sepkoski 1976; Smith 2001; Peters & Foote 2002; Foote 2003; Peters 2005, 2006*a*; Wall *et al.* 2009). Demonstrating potential common causes, separating their effects from simple preservation and sampling biases, and interpreting them mechanistically, are challenging problems that must be solved if palaeobiologists are to resolve true patterns and processes of macroevolutionary change through geological time.

The aim of the present paper is to examine the role of spatial structuring – of the environment and of biodiversity – in driving both macroevolutionary dynamics and biases in the fossil record (Fig. 1). Specifically, we are interested in the extent to which partitioning of marine environments may have influenced biodiversity and the preservation of that diversity. In part, the study is founded on the expectation that the geographical range and evolution–extinction dynamics of organisms occupying a particular habitat will be related to the area and fragmentation of that habitat (e.g. Rosenzweig 1995; Maurer & Nott 1998). At the same time, we anticipate that the quality of the marine fossil record will be influenced by, for example, the areal extent and distribution of depositional v. non-depositional or erosional environments, factors that are also likely to affect habitat fragmentation and diversity. To test the role of spatial structuring in the common cause hypothesis, we use the exemplary, well-studied and relatively complete fossil record of post-Middle Eocene, level-bottom, shelf molluscs from New Zealand (e.g. Crampton *et al.* 2003, 2006*a*). This record is approximately

From: McGowan, A. J. & Smith, A. B. (eds) *Comparing the Geological and Fossil Records: Implications for Biodiversity Studies*. Geological Society, London, Special Publications, **358**, 105–122.
DOI: 10.1144/SP358.8 0305-8719/11/$15.00

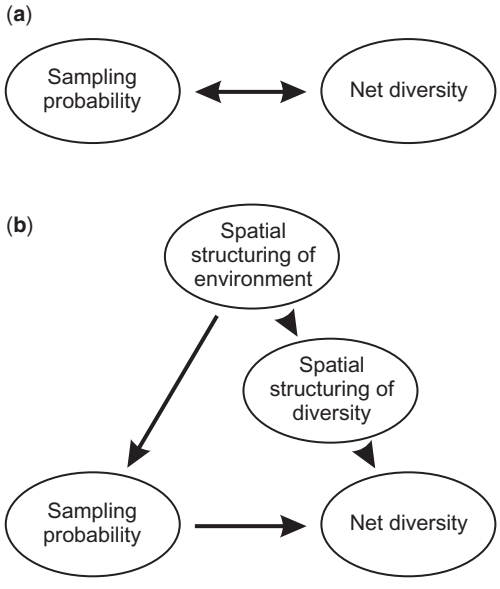

Fig. 1. Schematic showing the common-cause hypothesis and the relationships being tested here. (**a**) The apparent relationship between quality of the fossil record – here sampling probability – and diversity. A positive association is observed when diversity data are not corrected for uneven sampling through time. The common-cause hypothesis posits that this relationship results from both sampling related bias and also from latent factors that simultaneously influenced true diversity and the quality of the fossil record. (**b**) The potential common-causal agent that is tested here: spatial structuring of the environment. Arrows indicate direction of causality in the observed positive association.

40% complete at the species level for the Neogene, as estimated using two very different approaches (Cooper *et al.* 2006; Crampton *et al.* 2006*a*).

In this paper, therefore, we derive an index for spatial heterogeneity of the environment and adapt established indices of spatial differentiation in the New Zealand Cenozoic mollusc fauna. We then compare time series of these measures to time series of sampling probability (a proxy for the quality of the fossil record) and sampling standardized estimates of overall species richness, origination rate and extinction rate. All analyses and computations reported here were undertaken using R (R Development Core Team 2010).

Materials and methods

Material

This study is based on the rich fossil record of Cenozoic molluscs from New Zealand (including the Chatham Islands). Our analyses are restricted to gastropods, bivalves and scaphopods that are inferred to have inhabited level bottom environments at shelf depths (0–200 m). We have excluded pelagic, littoral, estuarine, bathyal and abyssal taxa because they are relatively poorly represented in the fossil record. The study was undertaken at the species level. Results are based primarily on two datasets that have been described elsewhere. The first, the synoptic dataset, comprises stratigraphic ranges and palaeoecological data for 5241 species, 1949 of which are undescribed. This dataset is taxonomically highly vetted and consistent, and was compiled by just two palaeontologists (A. G. Beu, P. A. Maxwell) over a period of many years (e.g. see Crampton *et al.* 2006*a*).

The second dataset is derived from the Fossil Record File Electronic Database (FRED), comprises individual occurrence records from over 8000 collection lists of Cenozoic molluscs, and contains data of varying vintage and quality. The FRED data have been subjected to several iterations of cleaning and vetting. In particular:

(1) All records from identifiers of unknown or doubtful expertise were eliminated.
(2) All uncertain identifications were eliminated.
(3) Collections that could not be identified to a single time bin (see below) were eliminated.
(4) All taxonomic names were checked against a list of synonyms and, where necessary, updated.
(5) Geographically and stratigraphically restricted, and biogeographically distinctive, records from northernmost New Zealand were excluded in order to avoid possible biasing effects, although inclusion of these data appears to make no significant difference to interpretations (results not presented here).

Following these adjustments, and including only level bottom shelf taxa, the dataset contains 30 114 occurrences of 1954 species from 6106 collections.

In order to use both these data compilations, and to derive maximum benefit from the taxonomically less reliable but otherwise valuable occurrence data in FRED, all our analyses are based on 1557 shelf taxa that are shared between the two datasets. These shared taxa are represented by 19 946 occurrences in 3482 collections in the FRED dataset. Because pre-Middle Eocene data are sparse, interpretations are restricted to post-40 Ma (late Middle Eocene) faunas (Fig. 2 and see below).

Measuring relationships between time series

Throughout this study, comparisons between different time series employ first differences between successive time bins, denoted with the Δ symbol, and

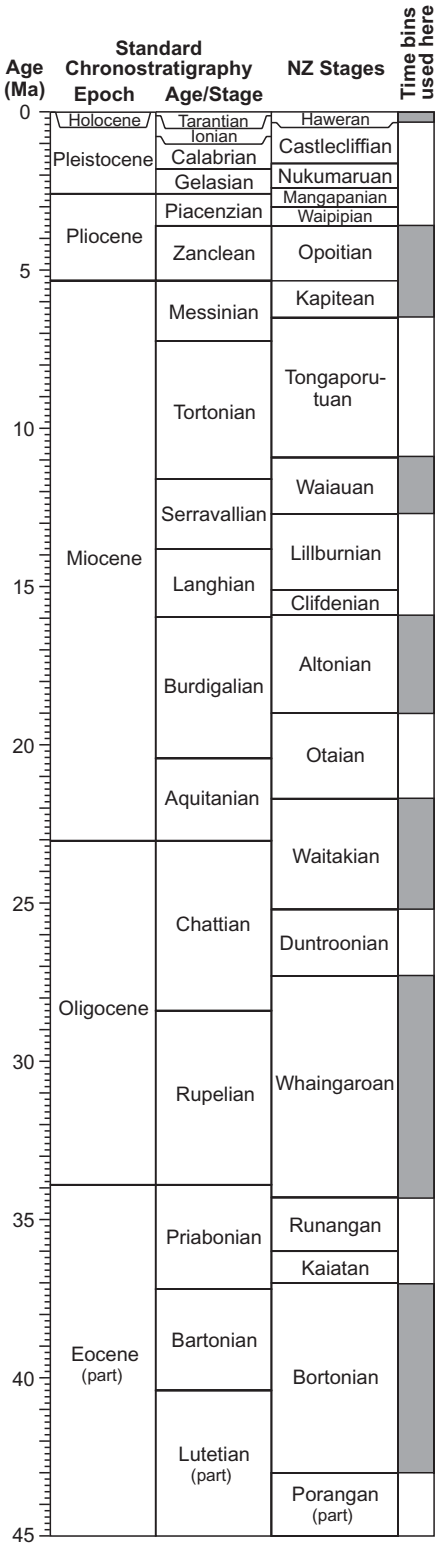

Fig. 2. New Zealand Eocene to Recent time-scale (after Cooper 2004; Hollis *et al.* 2010), showing the time bin groupings of stages used here.

Spearman's rank-order correlation coefficient (r_s). First-differencing is a simple treatment for minimizing autocorrelation in the data (e.g. McKinney 1990) and it means that comparisons between time series focus on short-term, time bin-to-time bin patterns of variation. The first difference is calculated simply as the value for one time bin minus the value for the preceding bin. Use of the non-parametric Spearman's correlation coefficient is a statistically conservative approach that makes few assumptions regarding underlying distributions of data. In cases where we consider pair-wise comparisons of several time series, we have corrected for inflation of type I error rates by adjusting individual significance levels using the false discovery rate procedure (Curran-Everett 2000).

Where relevant, we have explored relationships between different time series further using bivariate and multiple linear regression (again, based on first differences). In particular, we investigated models containing up to two predictor variables that were fitted individually and as additive or multiplicative (interactive) terms. We assessed the goodness-of-fit of simple and more complex regression models using *F*-tests and analyses of variance based on the residual sums of squares of competing nested models. In cases where assignment of predictor and response variables is not clear, we have tested all possible options for consistency of interpretation.

Estimation of diversity

Raw and sampling standardized estimates of mollusc diversity were generated using species occurrence records from the FRED dataset and the sampled within-bin counting protocol (e.g. Alroy *et al.* 2001, reasons for preferring this protocol are outlined in Crampton *et al.* 2006*b*; Alroy *et al.* 2008). The need for sampling standardization, and methods for achieving this, have been discussed at length elsewhere (e.g. Raup 1975; Alroy 1996, 2000, 2010; Miller & Foote 1996; Alroy *et al.* 2001, 2008). Here we have employed three different approaches to sampling standardization that make rather different assumptions about underlying diversity structure (e.g. Bush *et al.* 2004): by-lists unweighted standardization (LUW, Smith *et al.* 1985), by-lists occurrences weighted standardization (OW, Alroy 1996), and shareholder quorum standardization (SQ, Alroy 2010) (see Appendix A for details). Temporal patterns of diversity change are very similar across the family of analyses and

correlations are high (Appendix A), although the relative magnitudes of diversity peaks, the strength of a long-term trend, and error bars do vary. Given overall consistency between methods, for simplicity we report results based on the well-tested OW method, but note that none of our key interpretations change if we use results from either the LUW or SQ methods. Throughout the text, all comparisons between time series involve sampling standardized diversity unless specified with the prefix 'raw'.

Estimation of taxic rates

We estimated origination and extinction rates from biostratigraphic ranges as recorded in the synoptic dataset. To circumvent problems caused by incomplete sampling, we used the inverse survivorship modelling approach of Foote (2003, 2005). (The synoptic data, which we think provide the best representation of species' ranges, consist only of first and last appearances. Sampling standardization is therefore not an option, cf. Alroy 2008.) For living species, in order to avoid pull-of-the-Recent problems, we used the youngest occurrences as recorded in the FRED dataset (see Appendix A for further details). Rates given here are per-time bin, per-capita rates and assume pulsed turnover. Again, throughout the text, all comparisons between time series involve sampling standardized rates unless specified with the prefix 'raw'.

Estimation of spatial structuring of biodiversity

In order to quantify New Zealand-wide biogeographical structuring of molluscan diversity, we need a measure that is analogous to beta diversity (β) at the ecological scale. Beta diversity is the 'extent of differentiation of communities along habitat gradients' (Whittaker 1972, p. 214). In effect, we seek a measure of overall beta diversity averaged across all shelf habitat gradients; henceforth we refer to this as 'spatial turnover' of faunas. Unlike gamma (regional, γ) and alpha (community, α) diversities that are measured quantities, beta diversity is a derived metric. A plethora of concepts have been proposed for beta and its calculation has been the subject of extensive and vigorous debate (e.g. Ellison 2010 and associated forum papers). In particular, debate has revolved around the use of multiplicative v. additive partitions of diversity (e.g. Layou 2007; Holland 2010*a*, for discussion in a palaeontological context), and the conversion of some common beta indices to 'numbers equivalents'. In the present study, we lack species abundance proportions and are limited to measures that are appropriate for presence–absence data (Koleff *et al.* 2003 list 24 such measures). Our aim is not to contribute to debate about the merits of different approaches, but merely to derive some robust estimate of spatial turnover.

To this end, we have experimented with two measures of spatial turnover based on rather different, multiplicative formulations of beta diversity, namely Whittaker's (1960) original index, β_w, and Simpson's (1943) β_{sim} index (see Appendix A for details). Both are expressed as measures of dissimilarity, so that values of 0 correspond to perfect similarity between faunas and high positive values indicate maximum dissimilarity. Although the two indices yield time series that are significantly correlated, β_{sim} has the distinct advantage of being insensitive to sample size bias and variations in richness; henceforth we base all our interpretations on results from this method. We also calculated map sheet occupancy, an estimate of the proportion of potentially available geographical range actually inhabited by a species (Foote *et al.* 2007), as an independent proxy for spatial structuring of the biota; as expected, we find this to be inversely correlated with spatial turnover even though the two metrics target somewhat different attributes of faunal distribution (see Appendix A). We do not consider map sheet occupancy further here.

Estimation of sampling probability

Per-time bin sampling probability (R) is the probability that any species ranging through an entire bin has been sampled and recorded at least once from that bin. We use this as a metric of the quality of the fossil record. We estimated sampling probability using the standard gap statistic of Paul (1982), as modified by Foote & Raup (1996), calculated using the FRED occurrence data (see also Crampton *et al.* 2006*a*). Uncertainties were calculated as binomial standard errors.

Estimation of spatial structuring of the environment

In order to examine the relationship between spatial structuring of biodiversity and spatial structuring of the environment, we need to quantify temporal changes in the degree of gross fragmentation or partitioning of the environment. (Facies variation *per se*, although of interest, is likely to be expressed at much finer spatial and temporal scales than the partitioning of interest here; see Appendix A.) To measure environmental partitioning, we used data from the lithostratigraphic sections presented in King *et al.* (1999), as compiled into a digital database by Peters (see http://macrostrat.geology.wisc.edu and Peters 2006*b*). In this database, each of the 403 stratigraphic sections, arranged within 39 transects, has been digitized and coded according

to lithology, age and lithostratigraphic unit. The depositional environment of each unit was inferred separately using the compilation of Crampton *et al.* (2003), corrected for omissions and updated with new interpretations. For consistency with the molluscan diversity data, we considered only units that were deposited entirely or partly in shelfal marine environments.

To quantify environmental partitioning, we computed an index of spatial dispersion for deposition of shelfal facies (d_s), deposition of non-shelfal facies (dominantly deep marine, not considered here), and non-deposition or erosion (unconformities, d_u) (see Appendix A for details). For d_s, high values of this index indicate that shelfal units were deposited over much of the available and preserved area, and low values indicate patchy deposition relative to available and preserved area – that is, high environmental partitioning (Fig. 3). For d_u, high values indicate that non-deposition or erosion in shelfal environments was widespread, and low values correspond to geographical localization of non-deposition/erosion. We stress that, to the extent possible, d_u was computed in such a way as to encode changes that occurred in the shelfal environment and/or to reflect only short-term excursions out of this setting. Thus, for example, relatively brief subaerial exposure and minor erosion of shelfal habitats, lasting less than one time bin, have been captured. In contrast, periods of erosion/non-deposition that lasted through several time bins, or removed strata deposited during several preceding time bins, or were restricted apparently to deep marine environments, have been ignored. We also stress that our measures of spatial dispersion were calculated by comparing the observed spatial extent of facies to a random expectation derived from the actual spatial distribution of preserved sections in each time bin (see Appendix A); these measures are therefore insensitive to uneven representation of the rock record through time. Given the comparatively small number of sections available to us, this last point is an important reason why we could not simply use established coefficients of spatial structuring that have been developed in population ecology (e.g. Green 1966), which assume more comprehensive landscape coverage than available with our stratigraphic data.

In addition to our measures of (non-)depositional dispersion, following Peters (2005, 2006*a*) we also calculated per-time bin rates of initiation and truncation of shelfal sedimentary packages (macrostratigraphic 'origination' and 'extinction' rates, p_{strat} and q_{strat}, respectively). For these calculations, boundaries of shelfal sedimentary packages were defined by the limits of unconformities or intervals of non-shelf units. In other words, moving up a section, any transition from non-deposition, erosion, or non-shelf facies into shelf facies was coded as an initiation, and any transition out of shelf facies was counted as a truncation. Conformable transitions between different lithologies within the shelf, for example limestone to mudstone, were not counted as initiations or truncations.

Time-scale

Geological ages are given in terms of the New Zealand Cenozoic stages (Cooper 2004; Hollis *et al.* 2010). In order to minimize potential biases caused by unequal stage durations, short stages were combined to yield 13 Late Eocene to Pleistocene time bins (Fig. 2). The youngest of these, the Haweran Stage, is short and was retained simply as an edge-bounding bin. Ignoring the Haweran, the time bins have a mean duration of 3.5 Ma and a standard deviation of 1.4 Ma.

All the quantities discussed above are expressed per-time bin. Whereas some of these measures might be expected to correlate with time bin duration (e.g. sampling probability, see Crampton *et al.* 2006*a*), others should be independent of duration (e.g. origination and extinction rates under the assumption of pulsed turnover and our measures of environmental heterogeneity). In order to test for possible confounding effects of unequal time bin duration, we calculated correlation coefficients between all our measures and bin duration (excluding values for the short Haweran Stage). In addition, for all relationships of interest, we used multiple regression to test for sensitivity to the inclusion of time bin duration as a second predictor variable. In all but two cases, correlations between time bin duration and our key measures are low ($r_s < 0.17$) and non-significant, and inclusion of bin duration does not significantly improve regression fits between predictor and response variables. The two exceptions are β_{sim} and macrostratigraphic initiation rate (correlations against time bin duration, $r_s = -0.570$, $P = 0.067$ and $r_s = 0.557$, $P = 0.034$, respectively). Implications of these time bin duration effects are discussed in the Results.

We note that Crampton *et al.* (2006*a*) found a significant, positive correlation between sampling probability and time bin duration ($r_s = 0.442$, $P < 0.05$) for the New Zealand mollusc data, whereas here we do not ($r_s = 0.068$, $P = 0.842$). This difference results from the fact that in the present study we have grouped the seven shortest stages into time bins of relatively uniform duration, whereas Crampton *et al.* (2006*a*) did not.

Results

The key time series used in this study are shown in Figure 4. Bivariate scatter plots for important pairs

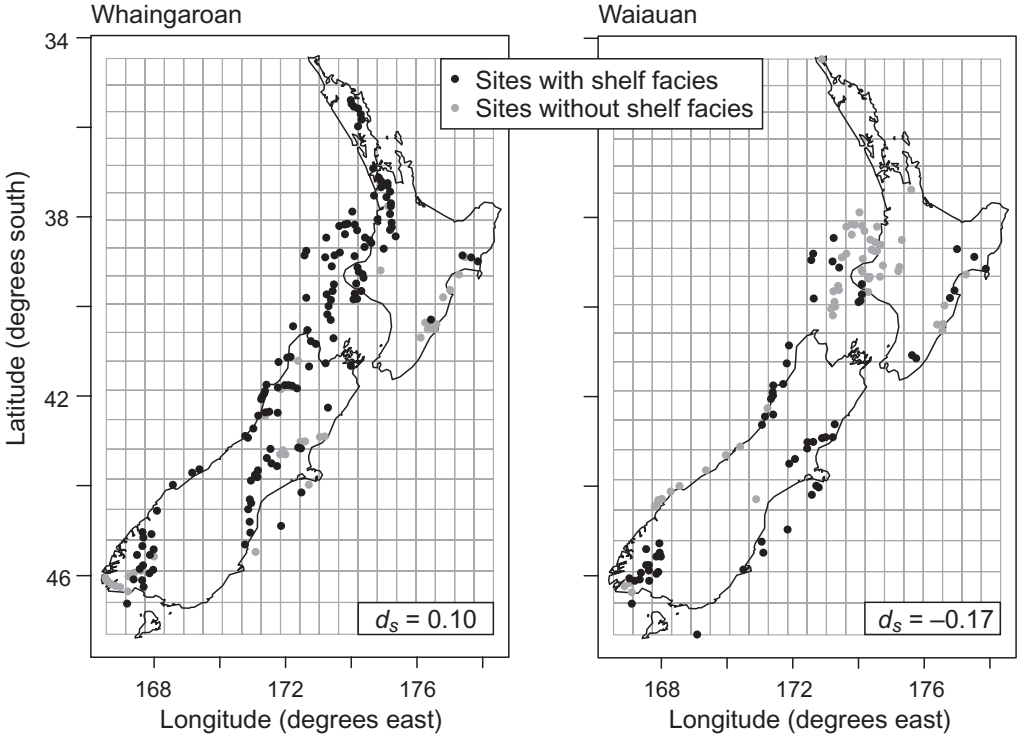

Fig. 3. Maps illustrating our measures of spatial structuring of the environment, in this case dispersion of deposition of shelfal facies (d_s) for two time bins with contrasting values. For clarity in these plots, sites of deposition of non-shelf facies and sites of non-deposition or erosion (unconformities) are combined; in our numerical analyses we separated these two categories. For the Whaingaroan, shelf facies are more dispersed relative to random expectation for the given distribution of measured sections (i.e. d_s is positive), whereas for the Waiauan, shelf facies are more clumped (i.e. d_s is negative). The 20 × 20 grid cells are equal-area but appear unequal in the latitude–longitude projection shown here. Note that some non-shelf facies or unconformities are present also at sites with shelf facies, but are not indicated here because such occurrences do not impact on the calculation or perception of d_s (see Appendix A for details and for a description of how we defined unconformities).

of metrics are shown in Figure 5. In the following, we consider three families of comparisons: (a) relationships between sampling probability and raw and sampling standardized diversity parameters; (b) relationships between taxic rates and other diversity parameters; and (c) relationships between spatial structuring of the environment and all other parameters. In passing, we also comment briefly on relationships between taxic rates and rates of macrostratigraphic initiation and truncation. Partly because of sparse data and partly because of poor constraints that result from being at the beginning of the time series (Foote 2001), the oldest interval considered here, the Bortonian stage, has a highly uncertain origination rate, with a standard error of 7.7 compared with a per-capita rate of 15.3. Therefore, this point is omitted from comparisons involving origination rate.

In general, palaeontological occurrence data are known to be affected by a lithification bias related to the combined effects of enhanced recovery of taxa from unlithified rocks and uneven temporal distribution of such rocks (e.g. Hendy 2009; Sessa *et al.* 2009). Ignoring the Haweran Stage (see above), 50% of the FRED collection lists used here have lithification recorded, 1% of these are unlithified, and most of the unlithified collections (35) are from the second youngest time bin, for which they represent 4% of collections with known lithification state. Whereas this situation might be expected to impart some bias, we find that removing these collections from analyses based on the FRED data makes no measurable difference (results not shown). We infer, therefore, that our key conclusions are unlikely to be influenced by a lithification bias.

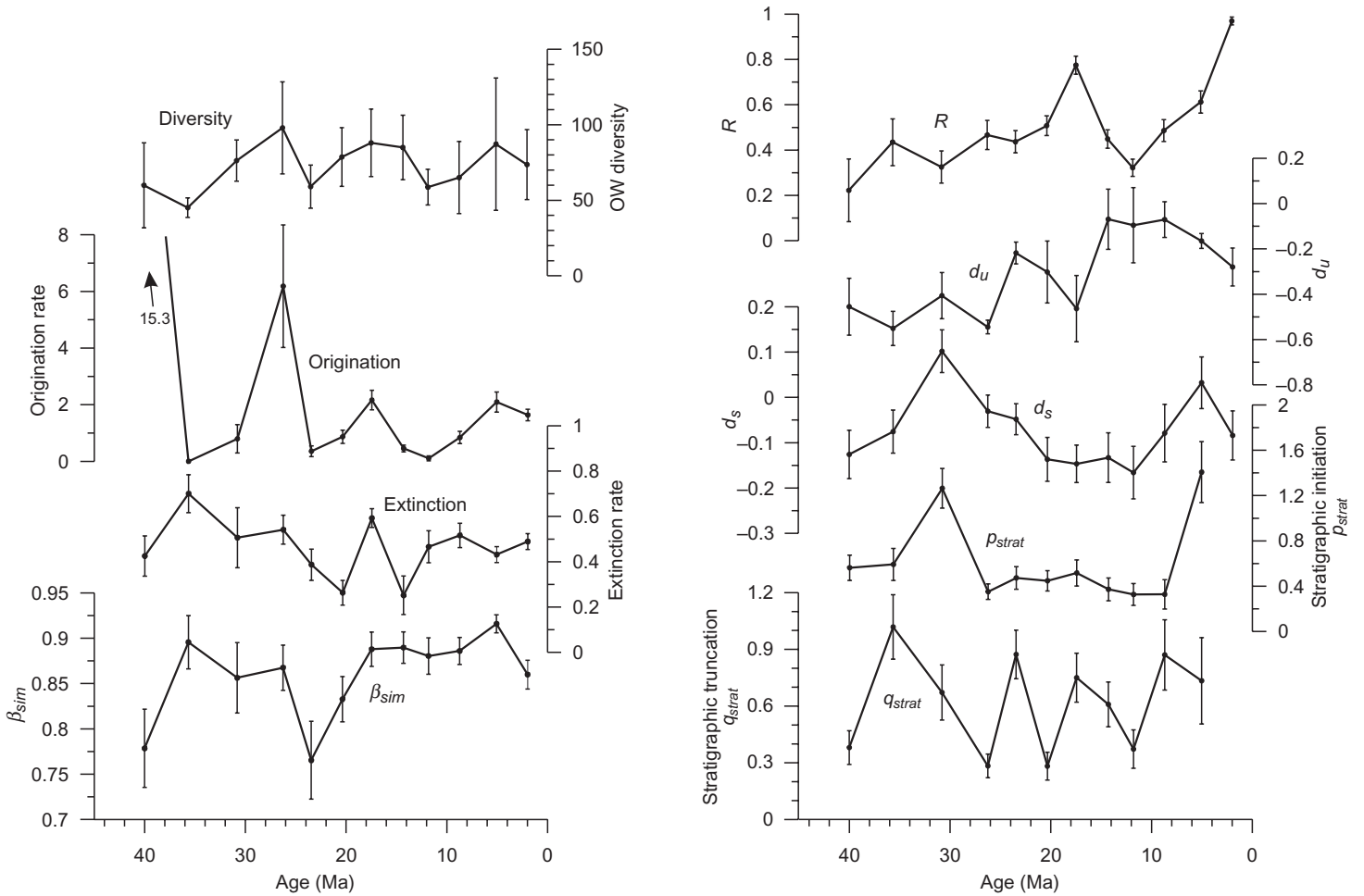

Fig. 4. Time series of key diversity and environmental metrics used in this study. All error bars are ±1 standard error. For explanations of the various measures, see Methods and Appendix A.

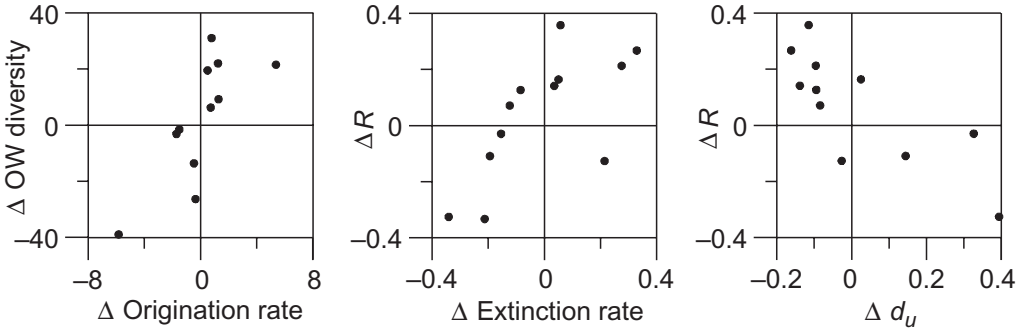

Fig. 5. Bivariate scatter plots for key pairs of metrics discussed in the text; the plots are based on first differences.

Relationships between sampling probability and raw sampling standardized diversity parameters

There are strong and significant positive correlations between raw diversity and sampling probability, and between raw extinction rate and sampling probability (Table 1). Sampling standardization removes the correlation between diversity and sampling probability, but the extinction rate-sampling probability correlation remains. Multiple regression suggests that the association between extinction and sampling probability is not affected by additive or interactive effects with origination or spatial turnover of faunas (results not shown).

Relationships between sampling standardized diversity, β_{sim}, and taxic rates

Temporal variations in diversity are strongly positively correlated with variations in origination rate (Table 2). The time series of extinction rate, by contrast, is uncorrelated with either origination rate or diversity. These results are robust irrespective of the measure of diversity used (results not shown). Furthermore, correlations between lagged time series (either first differences or raw data) are all apparently low and non-significant, although we acknowledge that our time series may be too short to test this rigorously. For example, origination rate is not correlated positively with extinction rate in the previous time bin and extinction rate is not correlated with diversity in the previous bin (results not shown), as might be expected if simple diversity-dependent dynamics were operating at the temporal scale of the time bins used here (cf. Alroy 2008).

There is a marginally significant and positive association between β_{sim} and origination rate, suggesting that origination rate may be higher at times when there is more spatial differentiation among faunas. This correlation is not significant once the effects of multiple comparisons are taken into account (Table 2). In addition, β_{sim} is possibly

Table 1. *Correlations between diversity parameters and sampling probability*

	r_s	P
Δ Raw diversity v. Δ R	0.839*	0.001
Δ OW diversity v. Δ R	0.056	0.869
Δ Raw origination v. Δ R	0.264	0.435
Δ Origination v. Δ R	0.518	0.107
Δ Raw extinction v. Δ R	0.782*	0.007
Δ Extinction v. Δ R	0.762*	0.006
Δ β_{sim} v. Δ R	0.455	0.140

Spearman's rank-order correlation coefficients between raw and sampling standardized diversity parameters on the one hand (sampling standardized unless stated), and sampling probability (R) on the other. An asterisk indicates correlations that remain significant at $P < 0.05$ after correction for the effects of multiple comparisons.

Table 2. *Correlations between diversity parameters and taxic rates*

	r_s	P
Δ Origination v. Δ extinction	0.162	0.549
Δ OW diversity v. Δ origination	0.791*	0.006
Δ OW diversity v. Δ extinction	−0.196	0.543
Δ OW diversity v. Δ β_{sim}	0.343	0.276
Δ β_{sim} v. Δ origination	0.691	0.023
Δ β_{sim} v. Δ extinction	0.483	0.115

Spearman's rank-order correlation coefficients between sampling standardized diversity, spatial turnover of faunas (β_{sim}) and taxic rates. An asterisk indicates correlations that remain significant at $P < 0.05$ after correction for the effects of multiple comparisons.

sensitive to variation in time bin duration which may, therefore, confound any apparent relationship with origination rate (see above). This possibility has been tested in two ways. First, bin duration is not identified as a significant second predictor in a multiple regression including β_{sim} and origination rate (results not shown here). Secondly, however, the partial correlation coefficient between origination and β_{sim}, based on residuals from linear regressions of both against time bin duration, is low and non-significant ($r_s = 0.378$, $P = 0.227$). We therefore conclude that the apparent positive association between β_{sim} and origination rate in our data is statistically non-significant and may be an artefact of time bin duration, in part at least, although this question should be re-examined with longer time series and larger datasets.

Relationships between spatial structuring of the environment and other parameters

Correlations between taxic rates and spatial turnover of faunas, on the one hand, and spatial structuring in the environment on the other, are all relatively weak and non-significant (Table 3). Sampling probability (R) is, however, significantly and negatively correlated with spatial dispersion of unconformities, d_u – species, as one would expect, seem to have a smaller chance of being sampled when unconformities are more widely dispersed. This correlation is robust to the grid size used in computation of d_u (results not shown). It is also supported by regression analysis ($R = 0.081–0.82d_u$, $se_{slope} = 0.21$, $t_{slope} = -3.86$, $p_{slope} = 0.004$, model $r^2 = 0.62$), a result that does not change substantively when other factors are included as additive or interactive terms.

Table 3. *Correlations between spatial structuring of the environment and other parameters*

	r_s	P
Δ Origination v. Δ d_u	−0.379	0.165
Δ Extinction v. Δ d_u	−0.468	0.070
Δ β_{sim} v. Δ d_u	−0.500	0.121
Δ R v. Δ d_u	−0.791*	0.006
Δ Origination v. Δ d_s	−0.229	0.411
Δ Extinction v. Δ d_s	−0.124	0.649
Δ β_{sim} v. Δ d_s	0.091	0.797
Δ R v. Δ d_s	−0.182	0.595

Spearman's rank-order correlation coefficients between measures of spatial structuring of the environment (d_u and d_s) on the one hand, and taxic rates, spatial turnover of faunas (β_{sim}) and sampling probability (R) on the other. An asterisk indicates correlations that remain significant at $P < 0.05$ after correction for the effects of multiple comparisons.

Table 4. *Correlations between stratigraphic rates, sampling probability and taxic rates*

	r_s	P
Δ p_{strat} v. Δ R	0.188	0.608
Δ q_{strat} v. Δ R	0.418	0.232
Δ Raw origination v. Δ p_{strat}	0.130	0.660
Δ Origination v. Δ p_{strat}	−0.059	0.844
Δ Raw extinction v. Δ q_{strat}	0.424	0.132
Δ Extinction v. Δ q_{strat}	0.479	0.073

Spearman's rank-order correlation coefficients between taxic rates, both raw and sampling standardized, and stratigraphic initiation and truncation rates (p_{strat} and q_{strat}, respectively).

Relationships between macrostratigraphic rates, sampling probability and taxic rates

We find no significant correlations between rates of macrostratigraphic initiation and truncation and sampling probability (Table 4). Likewise, we see no association between macrostratigraphic rates and taxic rate of origination, either raw or sampling standardized (Table 4). Although the relationship between extinction and macrostratigraphic truncation is not statistically significant, the two are positively correlated, in agreement with previous observations (Peters 2006*a*). Given the influence of time bin duration on macrostratigraphic initiation rate (see above), we tested these correlations using residuals from linear regressions of all quantities against time bin duration. The resultant partial correlations coefficients are little different from the raw coefficients shown in Table 4 (results not shown here).

Discussion

Key findings from our analyses are summarized in Figure 6. In essence, our results suggest that there might be at least two rather different dynamics affecting the real and apparent diversity history of New Zealand Cenozoic molluscs. First, sampling standardized origination rate is positively related to sampling standardized diversity. The causal direction in this association is unknown: increased origination rate might simply result in increased diversity, or increased diversity might promote origination via positive ecological feedbacks relating to, say, enhanced niche diversity.

Secondly, we find good evidence to suggest that extinction rate, both apparent and sampling standardized, and *apparent* diversity fluctuations vary in concert with sampling probability (R). The causal pathway between extinction and sampling must be indirect: if a species is extinct, then it cannot be

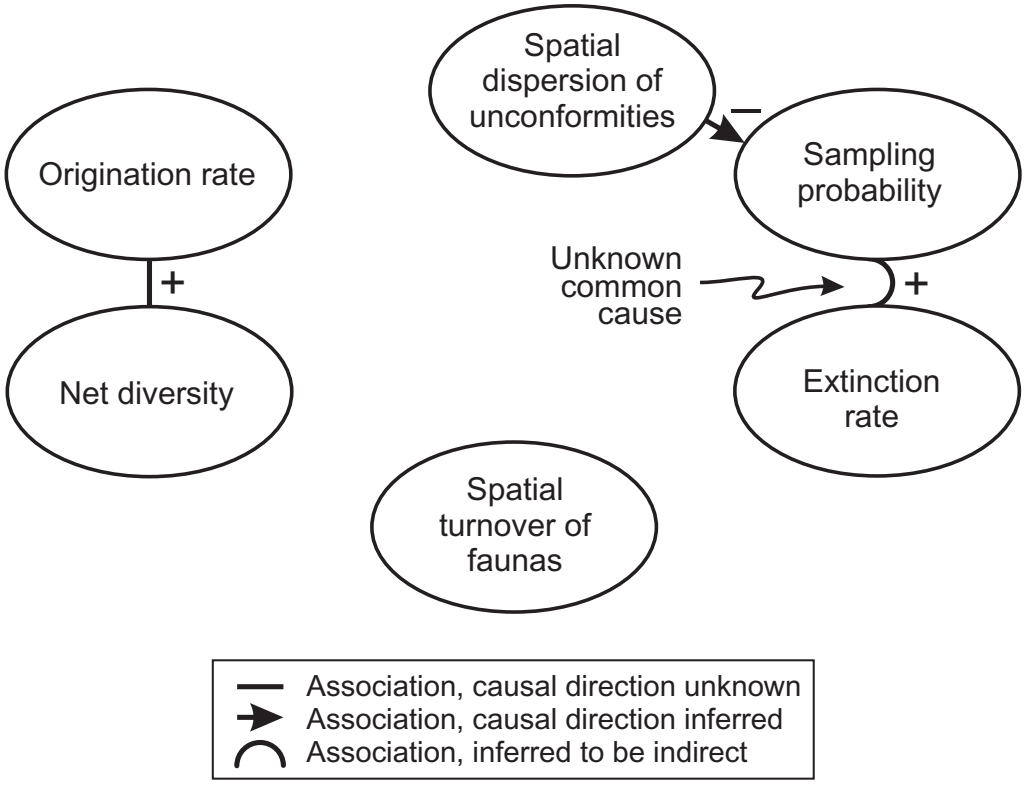

Fig. 6. Summary of key relationships identified here. ' + ' and ' − ' signs indicate positive and negative correlations, respectively. These relationships pertain to sampling standardized measures and short-term patterns of variation. Compare with Figure 1.

sampled, and if it is extant, then sampling must be controlled by some other factor. Sampling probability is itself related to spatial structuring of the environment, as measured using dispersion of unconformities (d_u), and the direction of causation is inferred to be from environmental structuring → sampling probability (Fig. 6). Specifically, for intervals when non-deposition or erosion was localized in discrete areas of uplift, sea-floor scour or sediment bypass – that is, by inference, the environment was highly structured – then sampling probability is high. Conversely, when non-deposition or erosion affected broad swaths of the shelf, then (unsurprisingly) sampling probability is low.

Unexpectedly, we find no convincing evidence for a link between spatial turnover of faunas (β_{sim}) and other diversity metrics or spatial structuring of the environment. We acknowledge, however, that such effects may have operated at spatial and temporal scales much finer than we can detect here.

From these results, therefore, and to answer the question posed in the Introduction, we would argue that there may be a common-cause effect linking spatial structuring of the environment, biases in the fossil record and diversity dynamics. This common cause, however, is manifest most strongly in the control of extinction rate and is not linked significantly to our proxies for spatial structuring of faunas, origination rate, or short-term changes in standing diversity. We explain and explore these findings in more detail below.

The strong positive correlation that exists between sampling probability and raw diversity in our data is, perhaps unsurprisingly, removed entirely by sampling standardization. In contrast, the positive correlation between extinction and sampling probability remains unchanged following sampling standardization. A similar association between times of high sampling standardized extinction and times of high sampling probability has been observed in an analysis of global marine diversity (Foote 2003). These observations are the basis for our claim that any common-cause effect on sampling and the diversity dynamic is likely to have operated through controls on extinction rate and not on net diversity directly (see further discussion below). (It

should be stressed that we rule out a simple and direct sampling effect on our sampling standardized extinction rate, which was constrained during computation to honour per-time bin sampling probability estimated from the FRED dataset. The lack of significant correlation between sampling standardized origination rate and sampling supports our inference that direct sampling effects do not drive our sampling standardized rates estimates.)

That said, we find that short-term changes in sampling standardized net diversity were in fact controlled primarily by changes in origination rate and *not* extinction rate. This result is consistent with observations for global, post-Palaeozoic marine genera (Foote 2000). Similarly, low correlations between coeval origination and extinction rates and between coeval extinction rate and diversity, as observed in our data, are consistent with patterns seen in global marine invertebrate genera (Alroy 2008). In contrast, however, we do not find significant correlations between origination and lagged extinction, or extinction and lagged diversity that have been observed in the global fauna (cf. Alroy 2008). This may suggest that, over the past 40 Ma, diversity-dependent dynamics either exerted relatively weak control of New Zealand molluscan diversity, or were manifest on a finer time-scale than we can resolve (i.e. less than the mean time bin duration of *c*. 3.5 Ma). Exploration of this question is beyond the scope of the present paper.

We find no statistically significant relationship between rates of initiation and truncation of stratigraphic packages and either taxic rates or sampling probability, even after correcting for time bin duration effects on macrostratigraphic initiation. This result contrasts with the findings of Peters (2005) for global Phanerozoic marine genera and the North American rock record, where there are significant associations between these measures. We note, however, that as in Peters' study, the strongest apparent relationship in the New Zealand molluscan data is between extinction and macrostratigraphic truncation; lack of significance here may reflect, in part, low statistical power due to the short time series available to us.

What then could be the common cause that links extinction rate and sampling probability in the New Zealand record? As noted by others, there is an expectation and supporting evidence to suggest that changes in relative sea-level have played a key role in driving both sampling probability (Smith 2001; Peters 2005, 2006*a*) and true extinction rate (Newell 1952; Sepkoski 1976; Hallam 1989, but see Hallam & Wignall 1999). Crampton *et al.* (2006*a*) inferred that there is indeed an association between second-order cycles of relative sea-level change, tied to tectonic activity, and the quality of the Cenozoic marine fossil record in New Zealand. Specifically, they posited that enhanced preservation of the fossil record occurred in mid-cycle positions and destruction of the record and/or low preservation potential occurred close to sequence boundaries and base-level falls. Such effects have been predicted by modelling experiments (Holland 2000), observed in the geological record of Europe (Smith & McGowan 2007), and are supported here by the association between dispersion of unconformities and sampling probability.

At the same time, we expect that tectonic cycles of uplift and erosion will have affected habitat area, partitioning, variety, and interconnectedness, and also ocean circulation, water mass properties, and nutrient availability. All of these factors could have affected local extinction rate. Importantly, analyses of faunal diversification, both for global Phanerozoic marine faunas and New Zealand Cenozoic molluscs, support models of pulsed turnover and, in particular, pulsed extinction (Foote 2005; Crampton *et al.* 2006*b*). The pulsed model indicates that turnover was focused, on average, at discrete points in time, rather than being distributed in a stochastically uniform way. Pulses of high extinction, it seems, may have been separated by periods of very low or negligible extinction (Foote 2007), although this pattern is not evident at the resolution of the time bins used here. These observations are consistent with the inference that episodic abiotic environmental change may, to a substantial extent, have paced extinction through the Phanerozoic. (This hypothesis does not, however, preclude a role for density-dependent ecological interactions in regulating diversity (e.g. Alroy 2008; Finnegan *et al.* 2008)).

At the resolution of our data, we note that the two times of lowest extinction rate, at *c*. 20 and *c*. 14 Ma (Fig. 4), both lie close to second-order sequence boundaries (King *et al.* 1999, see also Crampton *et al.* 2006*a*). These two sequence boundaries correspond to very major tectonic events and overall base-level falls in the New Zealand region, related to reorganizations of the plate boundary geometry. First, starting at *c*. 25 Ma and culminating at *c*. 20 Ma, the earliest significant manifestation of the modern plate boundary coincided with onset of subduction, widespread uplift, allochthon emplacement and, in the north, extensive volcanism (Rait *et al.* 1991; Herzer 1995; Wood & Stagpoole 2007; Uruski 2010). Secondly, between *c*. 15 and *c*. 12 Ma there was a major crustal closure event and increase in deformation rate, extensive uplift, local subsidence, and reorientation of the volcanic arc (Herzer 1995; Wood & Stagpoole 2007). Lastly, although less conspicuous than these two events, a similar decrease in extinction rate at *c*. 5 Ma coincided with the most recent second-order sequence boundary and what was a major

acceleration of uplift, erosion, and sedimentation (King *et al.* 1999; Wood & Stagpoole 2007).

The key point is that the intervals of low extinction rate apparently coincide with base-level falls; in other words, extinction rate is *low* when relative sea-level is low. This pattern is the opposite of that expected if diversity dynamics were being driven primarily by species-area affects, such that continental flooding increased habitat area, relaxed extinction pressures, and caused diversification (e.g. Newell 1967; Johnson 1974; Sepkoski 1976; Smith 2001; Wall *et al.* 2009). Although we acknowledge that the relationship between sea-level and habitable area is in fact likely to be highly complex and non-monotonic (Holland 2010*b*), our findings are consistent with the lack of a species-area effect observed both in this fauna (Crampton *et al.* 2006*b*) and in temperate marine bivalves (Harnik *et al.* 2010), and with the negative correlation between origination rate and sea-level observed in global Phanerozoic marine faunas (Cárdenas & Harries 2010).

More generally, the lack of significant associations between spatial turnover of faunas, spatial structuring of the environment, and any of our other diversity metrics supports the inference that simple habitat area and partitioning effects have not clearly been dominant and direct controls of large-scale Cenozoic molluscan diversity dynamics in New Zealand. This conclusion is unexpected. Again, however, we caution that these results may, to some extent, reflect low statistical power.

In any case, taken at face value, these findings suggest that other factors – such as nutrient availability or oceanographic effects (e.g. Cárdenas & Harries 2010) – have controlled extinction rate and, thereby, the relationship between environment, sampling, and diversity. We cannot, as yet, identify key drivers of extinction rate (or origination rate, for that matter), but we note that relationships between environmental variables and biotic consequences may be complex and are likely to be magnitude and scale-dependent. For example, a recent study (Algeo & Twitchett 2010) argued that greatly increased sedimentation rates during the end-Permian mass extinction may have contributed to the biotic crisis because of the harmful effects of siltation and eutrophication. In contrast, our data suggest that increased sediment fluxes during tectonic pulses coincided with times of lowered extinction rate.

The decoupling of sampling biases and environmental structuring from short-term variations in origination rate and diversity that is inferred here mirrors, in some ways, the finding of Wall *et al.* (2009) that outcrop area (i.e. sampling) operates largely independently of original habitat area to shape our view of apparent diversity history. Unlike Wall *et al.*, however, we conclude that there *is* a common-cause that simultaneously affected diversity, through an influence on extinction rate, and preservation of the fossil record.

Conclusions

Short-term variations in the diversity of New Zealand shelf molluscs over the past 40 Ma, and our perceptions of diversity change, apparently have been controlled by two rather different dynamics:

- Variations in sampling standardized diversity were related primarily and positively to changes in origination rate and *not* to changes in extinction rate.
- A common cause that affected simultaneously the quality of the fossil record and diversity dynamics operated primarily through its effect on extinction rate; times of increasing extinction rate tend to correspond to times of increasing sampling probability.

The quality of the fossil record is correlated inversely with spatial dispersion of unconformities that is, in turn, likely to have been controlled (at the temporal scale studied here) by tectonic processes manifest as second-order sequence stratigraphic cycles. Likewise, there is evidence to suggest that the extinction record is pulsed and has been paced by episodic environmental change, including second-order cycles of base-level change. The exact nature of the common cause linking sampling probability and extinction rate, however, remains unknown, although it is *unlikely* to have been related strongly or directly to gross scale environmental partitioning *per se*.

Contrary to our expectations, spatial structuring of the environment is not significantly correlated with spatial turnover (structuring) of faunas and does not seem to have been a strong or direct driver of changes in origination rate, extinction rate, or net diversity. Whereas these findings could reflect, to some extent, low statistical power resulting from the short time series available for analysis, the key point is that these effects, significant or not, are relatively small.

To answer the question posed at the start of this paper – to what extent did spatial partitioning of the environment influence molluscan diversity and the preservation of that diversity – we suggest that partitioning *did* influence the quality of the fossil record, but did *not* exert strong effect on sampling standardized diversity dynamics at the spatial scale of New Zealand and the temporal scale of stages.

For advice and discussion on the nature of the tectonic and stratigraphic records in New Zealand, and on statistical methods, we thank K. Bland, G. Browne, D. Harte, P. King, T. Mitchell, D. Rhoades, V. Stagpoole,

R. Sutherland and R. Wood. Thanks also to A. Smith and W. Kiessling for constructive and insightful comments that improved the manuscript. We acknowledge use of information contained in the New Zealand Fossil Record File (http://www.fred.org.nz/). Early stages of this work were supported by the Marsden Fund (contract GNS0404). JSC thanks organisers of the Lyell Meeting 2010 'Comparing the geological and fossil records: implications for biodiversity studies' for both the invite to participate in, and support to attend, the meeting.

Appendix A

Details of methods

Estimation of diversity

First, we employed the by-lists unweighted standardization (LUW, Smith *et al.* 1985). In this method, for each time bin, a fixed quota of 39 collection lists was drawn at random, without replacement, and the number of unique species was tabulated. This process was repeated 500 times and the average number of taxa in each time bin was calculated (Fig. A1). The standard error was computed as the standard deviation over all replicates. The quota was determined by the number of collections present in the most poorly sampled time bin. Secondly, we used the by-lists occurrences weighted standardization (OW, Alroy 1996). In this method, collection lists were drawn at random, without replacement, until a fixed quota of 109 species occurrences was accumulated, the number of taxa was tabulated, and the process was repeated 500 times (Fig. A1). Lastly, we used the recently developed shareholder quorum standardization (SQ, Alroy 2010). Although rather more complex than the other methods, in essence this approach samples to a quota based on 'coverage' – the summed frequencies of taxa that have

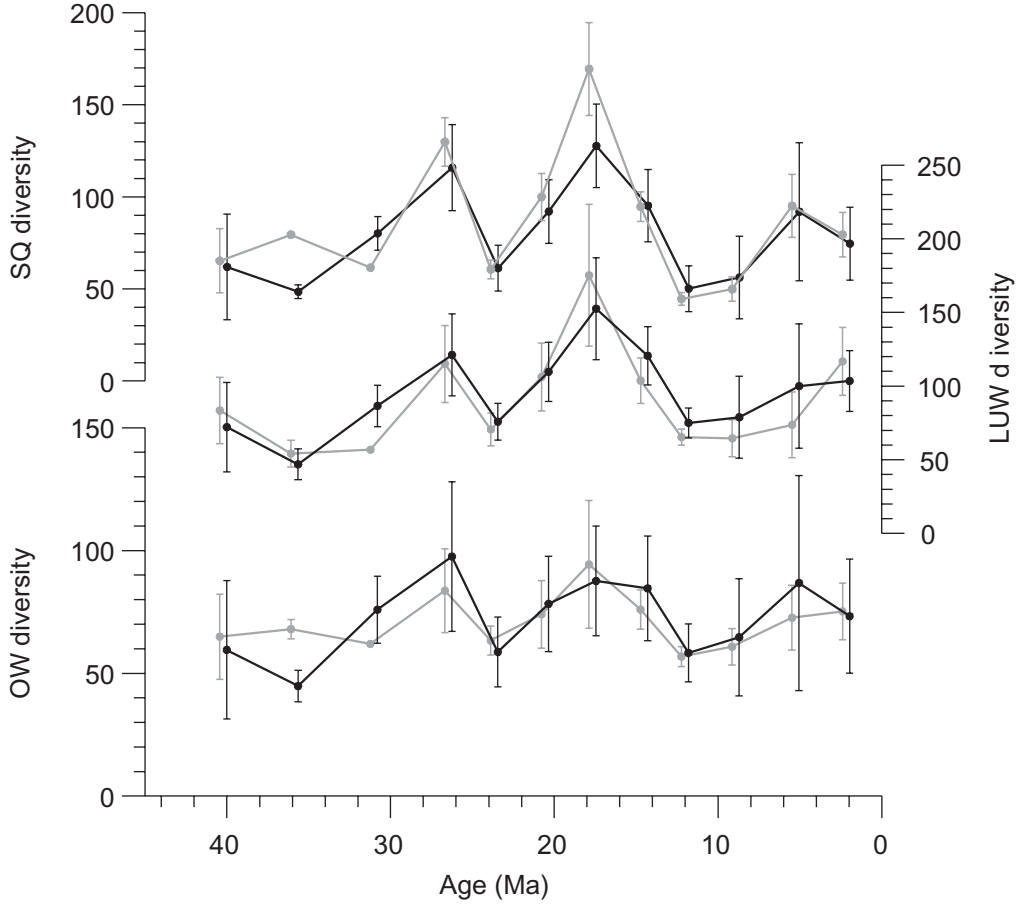

Fig. A1. Time series of sampling standardized diversity metrics examined here. Black curves employ the three-timer sampling probability correction; grey curves do not (see text for explanation). All error bars are ± 1 standard error.

been sampled at least once, where each frequency is the proportion of all occurrences accounted for by the species in question. The first time a taxon is encountered during random sampling, its full frequency is counted. In this way the method gives due recognition to uncommon taxa. Importantly, it does not sample to a fixed number of taxa, proportion of taxa, number of collections, or proportion of collections. During computation, the coverage quota q was adjusted separately for each time bin such that $q_{\text{adjusted}} = q/u$, where

$$u = 1 - \frac{s_1}{N - n_{\text{d}}},$$

s_1 = number of singleton species, n_{d} = the number of occurrences of the dominant species, and N is the total number of occurrences in a given time bin (see Alroy 2010 for a full explanation). As for the other sampling standardizations, results given here are based on 500 resampling trials (Fig. A1).

For all three sampling standardization methods and following Alroy *et al.* (2008), we adjusted the standardized diversity counts using the three-timer sampling probability in order to minimize otherwise uncorrected short-term variation in sampling. The three-timer sampling probability is

$$R_3 = \frac{3T}{(3T + \text{PT})},$$

where $3T$ = number of species sampled immediately before, during and after a time bin, and PT = number of species sampled immediately before and after the time bin, but not during. Diversity counts were multiplied by $R_{3\text{tot}}/R_{3\text{i}}$, where $R_{3\text{tot}}$ is the aggregate three-timer sampling probability calculated over all time bins for a given sampling trial, and $R_{3\text{i}}$ is the three-timer probability calculated for a given time bin and sampling trial. Use of the three-timer correction, although desirable, substantially inflates error bars on diversity estimates. Temporal patterns of estimates with and without the three-timer correction are, however, very similar (Fig. A1). Given this, and consistency between the three different sampling standardization methods (Table A1), we have confidence in the major patterns of diversity change inferred.

Table A1. *Correlations between sampling standardized measures of diversity*

	r_s	P
Δ SQ diversity v. Δ OW diversity	0.895	$\ll 0.001$
Δ SQ diversity v. Δ LUW diversity	0.867	< 0.001
Δ OW diversity v. Δ LUW diversity	0.874	< 0.001

Spearman's rank-order correlation coefficients between the different measures of diversity calculated here (all three adjusted using the three-timer correction, see Appendix A text for details).

Estimation of taxic rates

Inverse survivorship modelling of taxic rates uses numerical maximization of a likelihood function to find origination, extinction and sampling rates that yield the best fit between predicted forward and backward survivorship probabilities and the corresponding observed values (Foote 2003). Uncertainties in the parameter estimates, due to variance in the data together with imprecision in the optimization, were estimated by bootstrap resampling of biostratigraphic ranges (with replacement) 500 times. The model employed here assumes pulsed turnover of faunas (Foote 2005) – that is, that turnover occurred mainly at time bin boundaries – an assumption that is supported for the New Zealand mollusc data (Crampton *et al.* 2006*b*). Although inverse survivorship modelling allows for joint estimation of origination, extinction and sampling rates, here we constrained per-time bin sampling probability using values estimated directly from the FRED dataset using standard gap analysis (e.g. Paul 1982, as modified by Foote & Raup 1996; see main text). Constraining sampling rates in this way reduced uncertainties in the estimation of taxic rates.

Measures of spatial turnover of faunas

The first measure of spatial turnover of faunas tested here is based on β_{w} (Whittaker 1960), the simplest and most widely used beta diversity index (Koleff *et al.* 2003):

$$\beta_{\text{w}} = \frac{\gamma}{\bar{\alpha}},$$

where γ = total regional diversity and $\bar{\alpha}$ = average community diversity. β_{w} is a 'broad sense' measure *sensu* Koleff *et al.* (2003) that confounds effects of changes in species richness and species composition. For each time bin, we calculated average alpha diversity as the geometric mean of individual collection diversities, treating palaeontological collections as approximations of community assemblages. Use of the geometric mean is desirable given the strongly right-skewed distributions of collection diversities that vary over two orders of magnitude (although results do not change significantly if the median is used instead). Because the value of β_{w} is proportional to total diversity, it is highly sensitive to sample size. Hence, we used the OW sampling standardization protocol described above. Values of β_{w} used in our analyses are means based on a quota of 100 occurrences and 500 resampling trials; the standard error was computed as the standard deviation over all replicates. Because the resampling quota was determined by the most poorly sampled time bin and is comparatively low, error bars are correspondingly large. Increasing the quota to 275 occurrences means that error bars are substantially reduced but two time bins are excluded from the analysis. The time series based on quotas of 100 and 275 are, however, almost identical ($r_{\text{s}} = 0.997$, $P < 0.000$) and we therefore regard the major structure in the full time

series, based on a quota of 100 occurrences, as reliable despite the large error bars (Fig. A2).

The second and preferred measure of spatial turnover used here is based on the β_{sim} formulation of Simpson (1943), as re-expressed by (Koleff *et al.* 2003). It is a 'narrow sense' measure that focuses on compositional differences rather than variations in richness (cf. β_w, see above). Of the many formulations evaluated by Koleff *et al.* (2003), it is probably the one that performs best overall according to their criteria. For two 'quadrats', one the focal and one the neighbour:

$$\beta_{sim} = \frac{\min(b, c)}{\min(b, c) + a},$$

where the quadrats are the units of sampling, a = number of species shared between the quadrats, b = number of species present in the neighbouring but not the focal quadrat, and c = number of species present in the focal but not the neighbouring quadrat. We experimented with sampling quadrats based on individual collections and on 1:50 000 map sheets. Time series derived from these two approaches are very similar and, for consistency with our collection-based standardizations of diversity and β_w, all results reported here are based on collections as quadrat units. For each time bin, we computed β_{sim} for all pairwise combinations of focal and neighbouring collection and calculated the average. Trials using the complete dataset and the OW subsampling protocol, with a range of quota sizes, confirmed that the estimation of β_{sim} is essentially insensitive to sample size. For this reason, the final results used for interpretation (Figs 4 & A2) are based on the entire dataset and standard errors were estimated using bootstrap resampling, with replacement. Because calculation of β_{sim}

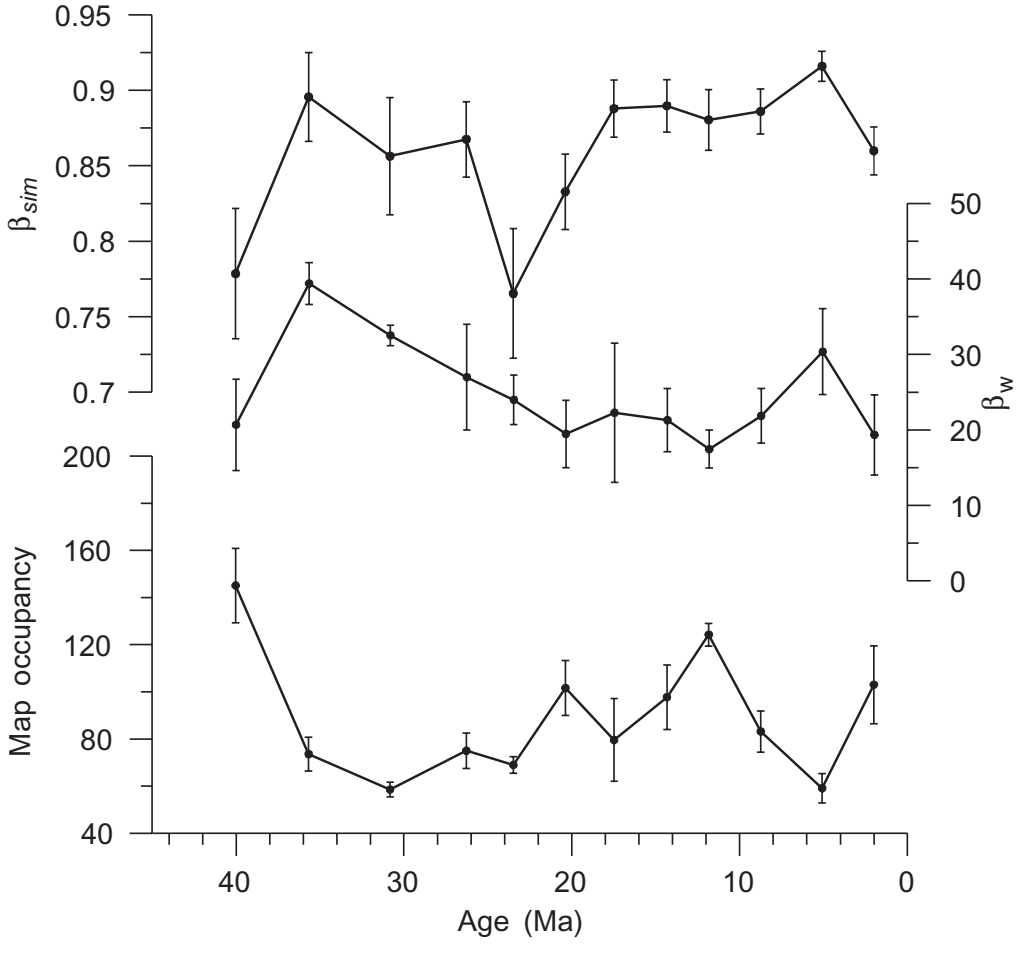

Fig. A2. Time series of faunal spatial turnover and map occupancy metrics examined here. Note that map occupancy is expressed as percentage deviation from the mean (see text for explanation). All error bars are ±1 standard error.

is time consuming for bins with many collections, the bootstrap resampling was limited to 100 trials and a maximum of 100 collections per time bin.

Lastly, we calculated average map sheet occupancy as an independent proxy for spatial structuring of faunas. For a given species, map sheet occupancy is the number of 1:50 000 topographical map sheets occupied as a proportion of the total number of sheets represented in a given time bin. Logically, map sheet occupancy and spatial turnover are expected to be inversely correlated, and this is the pattern we observe (Table A2, Fig. A2). Again, the measure is somewhat sensitive to sample size and therefore results were sampling standardized using the OW protocol and 500 resampling trials; the standard error was computed as the standard deviation over all replicates. As in the case of β_w, error bars are large for a resampling quota of 100 occurrences, but two time bins are excluded if the quota is increased. Given that patterns of temporal variation are almost identical if one uses quotas of 100 or 275 occurrences ($r_s = 0.967$, $P < 0.001$), for interpretation we simply expressed the two time series on a common scale by dividing by their respective means, combined them, and used values for the larger quota except for the two time bins that failed to meet this quota (Fig. A2).

We tested the impact of both 'under-' and 'oversampled' collections on the measurement of spatial turnover and map occupancy. Undersampled collections are those containing just a small number of species, and we might assume that these are either incompletely reported or taphonomically highly culled. Oversampled collections, on the other hand, are those containing relatively very high diversities that can be interpreted as atypical 'lagerstätte'. Both these types of collections might bias estimates of spatial turnover of faunas, although they are not expected to affect measures of map occupancy. In fact, experiments in which we removed collections with between 1, 2 or 3 species, and the upper 5%, 1% or 0.5% of most diverse collections, show that major patterns of temporal variation in our measures of spatial turnover are not unduly biased by these potentially problematic samples (e.g. $r_s = 0.874$, $P < 0.000$, for β_{sim} with and without single species collections and 0.5% of the most species-rich collections). As predicted, map occupancy is essentially insensitive to the presence of under- or oversampled collections ($r_s = 0.923$, $P < 0.000$, for a quota of 100 occurrences, with and without single species collections and 0.5% of the most speciose collections). Results reported throughout the text, therefore, are for analyses in which all collections have been retained (Fig. A2).

Table A2. *Correlations between different measures of spatial turnover of faunas*

	r_s	P
$\Delta\ \beta_{sim}$ v. $\Delta\ \beta_w$	0.608	0.040
Δ Map occupancy v. $\Delta\ \beta_{sim}$	−0.504	0.099
Δ Map occupancy v. $\Delta\ \beta_w$	−0.825*	0.002

Spearman's rank-order correlation coefficients between the different measures of spatial turnover of the fauna calculated here (β_{sim} and β_w) and map occupancy (see Appendix A text for details). An asterisk indicates correlations that remain significant at $P < 0.05$ after correction for the effects of multiple comparisons.

Measures of spatial structuring of the environment

Our indices of spatial dispersion of shelfal deposition (d_s) and non-deposition/erosion (d_u) were calculated as follows. Shelfal marine units are those inferred to have been deposited, wholly or in part, at shelf depths in open marine environments (excluding units restricted to estuarine environments). New Zealand was divided into an arbitrary grid of equal-area cells following projection of data using a Lambert cylindrical equal-area projection; we experimented with grids of 20 cells by 20 cells (each cell 3531 km^2), 30 cells by 30 cells (each cell 1569 km^2), and 40 cells by 40 cells (each cell 883 km^2); reported results are based on the 20 × 20 grid and we note instances where interpretations are sensitive to the grid used.

For each of the 403 stratigraphic sections, we considered only the interval lying between the base of the oldest stratigraphic unit and the top of the youngest stratigraphic unit and, for each time bin, we counted three quantities. First, we counted the number of sections, s_s, containing shelfal units within the time bin and the number of cells, c_s occupied by these sections. For this measure, different lithologies and facies were not discriminated (see below). Secondly, we counted the number of sections and cells containing non-shelfal units within the time bin (these quantities are not used further here). Lastly, we counted the number of sections, s_u, containing intervals of unconformity and/or non-deposition and the number of cells, c_u, occupied by these sections, according to the following protocol. Unconformities were counted only if they lie entirely within the time bin in question or at one of its boundaries, and if they are bounded above and/or below by shelfal units. These protocols were designed to ensure that counted unconformities, if erosional, are localized in time to reflect environmental events that occurred during the span of time bin, and to ensure that they reflect environmental changes that affected shelfal environments.

We then picked at random s_s sections from all those present within the time bin (sampled without replacement from the set of all sections, including those with non-shelf deposition or unconformities) and counted the number of cells occupied by this sample, c_{sr}. For each of j such randomizations, we calculated the proportional difference between the number of cells containing shelfal deposits and the corresponding number for the random array:

$$d_{s,j} = \ln(c_s/c_{sr,j}).$$

The random array was always drawn from the sections actually present in the time bin; therefore, spatial structure in the distribution of sections was accounted for. The reported values of d_s for each time bin are the averages of 100 such randomizations and the standard error was computed as the standard deviation over all replicates. If d_s is equal to zero, the distribution of shelfal deposits is indistinguishable from random. Positive and negative values of d_s correspond to shelfal deposits that are more dispersed and more clustered, respectively, than expected by chance. Most of our average measures are negative (see Fig. 4), but we refer to them in a relative sense as more dispersed if they are less negative. Values of d_u were calculated in the same way, using s_u, c_u and $c_{ur.j}$.

We also calculated an analogous set of measures of spatial dispersion of shelf-only lithofacies, distinguishing coarse (conglomerate and breccia), medium (sandstone) and fine (siltstone and mudstone) siliciclastics, carbonate, volcanic rocks, and unconformities. The resulting temporal patterns of dispersion of shelfal lithofacies show uniformly low correlations with our measures of diversity and spatial turnover of faunas (results not shown here). This is surprising but, as noted in the main text, we suggest that structuring of lithofacies may be expressed, and may influence biodiversity dynamics, at finer spatial and temporal scales than resolved here. This question will be examined elsewhere.

References

Algeo, T. J. & Twitchett, R. J. 2010. Anomalous Early Triassic sediment fluxes due to elevated weathering rates and their biological consequences. *Geology*, **38**, 1023–1026.

Alroy, J. 1996. Constant extinction, constrained diversification, and uncoordinated stasis in North American mammals. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **127**, 285–311.

Alroy, J. 2000. New methods for quantifying macroevolutionary patterns and processes. *Paleobiology*, **26**, 707–733.

Alroy, J. 2008. Dynamics of origination and extinction in the marine fossil record. *Proceedings of the National Academy of Sciences of the USA*, **105**, 11 536–11 542.

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 USA*, **98**, 6261–6266.

Alroy, J., Aberhan, M. *et al.* 2008. Phanerozoic trends in the global diversity of marine invertebrates. *Science*, **321**, 97–100.

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.

Cárdenas, A. L. & Harries, P. J. 2010. Effect of nutrient availability on marine origination rates throughout the Phanerozoic eon. *Nature Geoscience*, **3**, 430–434.

Connolly, S. R. & Miller, A. I. 2001. Joint estimation of sampling and turnover rates from fossil databases: capture-mark-recapture methods revisited. *Paleobiology*, **27**, 751–767.

Cooper, R. A. (ed.) 2004. *The New Zealand geological timescale*. Institute of Geological and Nuclear Sciences, Monograph, **22**, 1–284. Lower Hutt, New Zealand.

Cooper, R. A., Maxwell, P. A., Crampton, J. S., Beu, A. G., Jones, C. M. & Marshall, B. A. 2006. Completeness of the fossil record: estimating losses due to small body size. *Geology*, **34**, 241–244.

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.

Crampton, J. S., Foote, M. *et al.* 2006*a*. Second-order sequence stratigraphic controls on the quality of the fossil record at an active margin: New Zealand Eocene to Recent shelf molluscs. *Palaios*, **21**, 86–105.

Crampton, J. S., Foote, M. *et al.* 2006*b*. The ark was full! Constant to declining Cenozoic shallow marine biodiversity on an isolated mid-latitude continent. *Paleobiology*, **32**, 509–532.

Curran-Everett, D. 2000. Multiple comparisons: philosophies and illustrations. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology*, **279**, R1–R8.

Ellison, A. M. 2010. Partitioning diversity. *Ecology*, **91**, 1962–1963.

Finnegan, S., Payne, J. L. & Wang, S. C. 2008. The Red Queen revisited: reevaluating the age selectivity of Phanerozoic marine genus extinctions. *Paleobiology*, **34**, 318–341.

Foote, M. 2000. Origination and extinction components of taxonomic diversity: Paleozoic and post-Paleozoic dynamics. *Paleobiology*, **26**, 578–605.

Foote, M. 2001. Inferring temporal patterns of preservation, origination, and extinction from taxonomic survivorship analysis. *Paleobiology*, **27**, 602–630.

Foote, M. 2003. Origination and extinction through the Phanerozoic: a new approach. *The Journal of Geology*, **111**, 125–148.

Foote, M. 2005. Pulsed origination and extinction in the marine realm. *Paleobiology*, **31**, 6–20.

Foote, M. 2007. Extinction and quiescence in marine animal genera. *Paleobiology*, **33**, 261–272.

Foote, M. & Raup, D. M. 1996. Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology*, **22**, 121–140.

Foote, M., Crampton, J. S., Beu, A. G., Marshall, B. A., Cooper, R. A., Maxwell, P. A. & Matcham, I. 2007. Rise and fall of species occupancy in Cenozoic fossil molluscs. *Science*, **318**, 1131–1134.

Green, R. H. 1966. Measurement of non-randomness in spatial distributions. *Researches on Population Ecology*, **8**, 1–7.

Hallam, A. 1989. The case for sea-level change as a dominant causal factor in mass extinction of marine invertebrates. *Philosophical Transactions of the Royal Society of London, B*, **325**, 437–455.

Hallam, A. & Wignall, P. B. 1999. Mass extinctions and sea-level changes. *Earth-Science Reviews*, **48**, 217–250.

HARNIK, P. G., JABLONSKI, D., KRUG, A. Z. & VALENTINE, J. W. 2010. Genus age, provincial area and the taxonomic structure of marine faunas. *Proceedings of the Royal Society, B*, **277**, 3427–3435.

HENDY, A. J. W. 2009. The influence of lithification on Cenozoic marine biodiversity trends. *Paleobiology*, **35**, 51–62.

HERZER, R. H. 1995. Seismic stratigraphy of a buried volcanic arc, Northland, New Zealand and implications for Neogene subduction. *Marine and Petroleum Geology*, **12**, 511–531.

HOLLAND, S. M. 2000. The quality of the fossil record: a sequence stratigraphic perspective. *Paleobiology; Supplement*, **26**, 148–168.

HOLLAND, S. M. 2010*a*. Additive diversity partitioning in palaeobiology: revisiting Sepkoski's question. *Palaeontology*, **53**, 1237–1254.

HOLLAND, S. M. 2010*b*. The common-cause hypothesis: initial predictions of habitable area from simple geometric models. *Geological Society of America Abstracts with Programs*, **42**, 138. Geological Society of America, Boulder, Colorado, USA.

HOLLIS, C. J., BEU, A. G. *ET AL*. 2010. Calibration of the New Zealand Cretaceous–Cenozoic timescale to GTS2004. *GNS Science Report*, **2010/43**, 1–20.

JOHNSON, J. G. 1974. Extinction of perched faunas. *Geology*, **1**, 479–482.

KING, P. R., NAISH, T. R., BROWNE, G. H., FIELD, B. D. & EDBROOKE, S. W. 1999. Cretaceous to Recent sedimentary patterns in New Zealand. *Institute of Geological and Nuclear Sciences, Folio Series*, **1**, 1–35. Lower Hutt, New Zealand.

KOLEFF, P., GASTON, K. J. & LENNON, J. J. 2003. Measuring beta diversity for presence–absence data. *Journal of Animal Ecology*, **72**, 367–382.

LAYOU, K. M. 2007. A quantitative null model of additive diversity partitioning: examining the response of beta diversity to extinction. *Paleobiology*, **33**, 116–124.

MAURER, B. A. & NOTT, M. P. 1998. Geographic range fragmentation and the evolution of biological diversity. *In*: MCKINNEY, M. L. & DRAKE, J. A. (eds) *Biodiversity Dynamics; Turnover of Populations, Taxa, and Communities*. Columbia University Press, New York, 31–50.

MCKINNEY, M. L. 1990. Classifying and analysing evolutionary trends. *In*: MCNAMARA, K. J. (ed.) *Evolutionary Trends*. Belhaven Press, London, 28–58.

MILLER, A. I. & FOOTE, M. 1996. Calibrating the Ordovician Radiation of marine life: implications for Phanerozoic diversity trends. *Paleobiology*, **22**, 304–309.

NEWELL, N. D. 1952. Periodicity in invertebrate evolution. *Journal of Paleontology*, **26**, 371–385.

NEWELL, N. D. 1967. Revolutions in the history of life. *Special Papers of the Geological Society of America*, **89**, 63–91.

PAUL, C. R. C. 1982. The adequacy of the fossil record. *In*: JOYSEY, K. A. & FRIDAY, A. E. (eds) *Problems of Phylogenetic Reconstruction*. Academic Press, London, 75–117.

PETERS, S. E. 2005. Geologic constraints on the macroevolutionary history of marine animals. *Proceedings of the National Academy of Sciences of the USA*, **102**, 12326–12331.

PETERS, S. E. 2006*a*. Genus extinction, origination, and the durations of sedimentary hiatuses. *Paleobiology*, **32**, 387–407.

PETERS, S. E. 2006*b*. Macrostratigraphy of North America. *The Journal of Geology*, **114**, 391–412.

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.

RAIT, G., CHANIER, F. & WATERS, D. W. 1991. Landward- and seaward-directed thrusting accompanying the onset of subduction beneath New Zealand. *Geology*, **19**, 230–233.

RAUP, D. M. 1975. Taxonomic diversity estimation using rarefaction. *Paleobiology*, **1**, 333–342.

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

ROSENZWEIG, M. L. 1995. *Species Diversity in Space and Time*. Cambridge University Press, Cambridge.

SEPKOSKI, J. J. 1976. Species diversity in the Phanerozoic: species-area effects. *Paleobiology*, **2**, 298–303.

SESSA, J. A., PATZKOWSKY, M. E. & BRALOWER, T. J. 2009. The impact of lithification on the diversity, size distribution, and recovery dynamics of marine invertebrate assemblages. *Geology*, **37**, 115–118.

SIMPSON, G. G. 1943. Mammals and the nature of continents. *American Journal of Science*, **241**, 1–31.

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. & 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.

SMITH, E. P., STEWART, P. M. & CAIRNS, J. J. 1985. Similarities between rarefaction methods. *Hydrobiologia*, **120**, 167–170.

URUSKI, C. 2010. New Zealand's deepwater frontier. *Marine and Petroleum Geology*, **27**, 2005–2026.

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.

WHITTAKER, R. H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, **30**, 279–338.

WHITTAKER, R. H. 1972. Evolution and measurement of species diversity. *Taxon*, **21**, 213–251.

WOOD, R. A. & STAGPOOLE, V. M. 2007. Validation of tectonic reconstructions by crustal volume balance: New Zealand through the Cenozoic. *Geological Society of America Bulletin*, **119**, 933–943.