Paleobiology, 32(4), 2006, pp. 509–532

The ark was full! Constant to declining Cenozoic shallow marine biodiversity on an isolated midlatitude continent

James S. Crampton, Michael Foote, Alan G. Beu, Phillip A. Maxwell, Roger A. Cooper, Iain Matcham, Bruce A. Marshall, and Craig M. Jones

Abstract.—In recent years several authors have questioned the reality of a widely accepted and apparently large increase in marine biodiversity through the Cenozoic. Here we use collection-level occurrence data from the rich and uniquely well documented New Zealand (NZ) shelfal marine mollusc fauna to test this question at a regional scale. Because the NZ data were generated by a small number of workers and have been databased over many decades, we have been able to either avoid or quantify many of the biases inherent in analyses of past biodiversity. In particular, our major conclusions are robust to several potential taphonomic and systematic biases and methodological uncertainties, namely non-uniform loss of aragonitic faunas, biostratigraphic range errors, taxonomic errors, choice of time bins, choice of analytical protocols, and taxonomic rank of analysis.

The number of taxa sampled increases through the Cenozoic. Once diversity estimates are standardized for sampling biases, however, we see no evidence for an increase in marine mollusc diversity in the NZ region through the middle and late Cenozoic. Instead, diversity has been approximately constant for much of the past 40 Myr and, at the species and genus levels, has declined over the past ~5 Myr. Assuming that the result for NZ shelfal molluscs is representative of other taxonomic groups and other temperate faunal provinces, then this suggests that the postulated global increase in diversity is either an artifact of sampling bias or analytical methods, resulted from increasing provinciality, or was driven by large increases in diversity in tropical regions. We see no evidence for a species-area effect on diversity. Likewise, we are unable to demonstrate a relationship between marine temperature and diversity, although this question should be re-examined once refined shallow marine temperature estimates become available.

James S. Crampton, Alan G. Beu, Roger A. Cooper, Craig M. Jones, and Iain Matcham. GNS Science, Post Office Box 30-368, Lower Hutt, New Zealand. E-mail: j.crampton@gns.cri.nz
Michael Foote. Department of the Geophysical Sciences, University of Chicago, 5734 South Ellis Avenue, Chicago, Illinois 60637
Bruce A Marshall. Museum of New Zealand Te Papa Tongarewa, Post Office Box 467, Wellington, New Zealand
Phillip A. Maxwell. 257 Otipua Road, Highfield, Timaru, New Zealand

Accepted: 2 June 2006

Introduction

Has marine biodiversity increased markedly over the past 500 million years, or not (Raup 1979)? For the past 25 years, since a widely cited study by Sepkoski and colleagues (1981), the consensus answer to this question has been affirmative. In particular, many analyses have suggested that marine biodiversity increased markedly during the Cenozoic and reached a Phanerozoic maximum during the Pliocene and Pleistocene (e.g., Sepkoski and Hulver 1985; Signor 1985, 1990; Sepkoski 1993; Bambach 1999; Jablonski et al. 2003; Bush et al. 2004). If correct, this implies that diversity may be indeterminate and self-augmenting (e.g., Vermeij 2005), or that saturation has not yet been achieved. Recently, however, several authors have revisited earlier uncertainties and have cautioned that such patterns may be in part or entirely the result of various biases in the fossil record (Alroy et al. 2001; Peters and Foote 2001). Earlier, Raup (1976: p. 289) suggested that ''the apparent number of [fossil] species is strongly dependent on sampling and that many of the changes in diversity seen in the Phanerozoic are artifactual. Consequently, there is no compelling evidence for a general increase in the number of [marine] invertebrate species from Paleozoic to Recent. . . . Diversity may have been in dynamic equilibrium throughout much of this time.''

This question is key to understanding past macroevolutionary and macroecological dynamics and is now a central research theme in paleobiology. Because of unsolved methodological problems, however, the question of

 0094-8373/06/3204-0001/$1.00

global Phanerozoic diversity increase remains open (these problems are discussed in more detail below). In addition, some authors have questioned the value of attempts to quantify paleobiodiversity at the global scale, suggesting instead that large scale patterns of diversity change should be studied at the level of biogeographically meaningful regions (Vermeij and Leighton 2003).

Here we use collection-level occurrence data from the exceptionally well known fossil record of New Zealand (NZ) Cenozoic marine molluscs to examine first-order patterns of regional-scale marine biodiversity change over the past 40 Myr on an isolated, southern mid-latitude continent. Taken at face value, the fossil record suggests a substantial increase in diversity from the Eocene to the Pliocene. Following sampling standardization, however, we show that marine shelfal molluscan biodiversity was approximately constant for much of the past 40 Myr and declined in the Pliocene and Pleistocene. We are unable to demonstrate either a species-area effect in shelfal molluscs or a significant correlation between diversity and marine temperature. We note that our data do not bear directly on the hypothesis that sample-level, alpha diversity has increased through time, a question that relates to spatial structuring of diversity within regions (e.g., Bambach 1977; Powell and Kowalewski 2002; Bush and Bambach 2004; Kowalewski et al. 2006).

Material

The NZ Cenozoic marine fossil record provides a valuable data set for the study of regional biodiversity dynamics for several reasons. (1) NZ has by far the richest Cenozoic fossil record of molluscs in the Southern Hemisphere, containing more than twice the total number of species known from all other mid- to high-latitude southern regions combined. Recent work has shown that the overall completeness of the NZ molluscan record—the proportion of species that has been sampled and recorded at least once—is 32% for the entire Cenozoic and 41% for the Neogene (Crampton et al. 2006; see also Cooper et al. 2006). (2) NZ has been geographically isolated for approximately 80 Myr, resulting in >95% endemism in the mollusc fauna at the species level, even allowing for some degree of regional taxonomic artifact (Beu unpublished data; see also Spencer et al. in press). Because of high endemism, biodiversity dynamics at the species level are likely to have been dominated by in situ processes and to have been insulated from the effects of wholesale immigration and emigration. (The figure for Cenozoic species endemism, although perhaps surprisingly high, is not grossly inconsistent with a recent estimate of 85% species-level endemism in the living mollusc fauna [Spencer et al. in press]. At the genus level, endemism in the Cenozoic mollusc fauna is likely to be <23% [Maxwell *in* Spencer et al. in press].) (3) The fossil record in NZ is uniquely well-documented by the locality- and occurrence-based Fossil Record File database (FRED data set). This has existed in paper form since 1946 and has been progressively digitized since 1970 (Crampton et al. 2003). The digital database contains 6154 Cenozoic mollusc collection lists that were used in the present analysis. (4) In addition to the FRED, stratigraphic ranges, shell composition and inferred paleoecological habits of the Cenozoic mollusc fauna have been synthesized independently in a taxonomic database that we call the synoptic data set (Crampton et al. 2006). This data set includes 5241 species, 1949 of which are undescribed, and 1272 genera.

In the context of the FRED data set, a collection list is a record of all taxa collected and identified by one or more paleontologists from a single locality on a single occasion. Given differences in sampling and recording practices among paleontologists and over time, a single locality might comprise a single bed or horizon, several beds, or an entire outcrop. The database does not at present contain individual species abundance information, although such data will be included in future as they become available. In addition to fossil lists, the FRED data set includes information on stratigraphic relationships and sedimentologic properties of fossiliferous strata. Of relevance here are data on matrix lithification, which is classified into four subjective categories: ''hard,'' ''moderately hard,'' ''moderately soft,'' and ''unconsolidated.''

Inferred paleoecological habits of each species are recorded in the synoptic data set and were based on the ecology of closest living relatives, functional interpretations, lithofacies and faunal associations (including microfossil data), and general stratigraphical considerations (Crampton et al. 2003). Shell composition, also recorded in the synoptic data set, was classified in a way that aimed to capture potential bias in our fossil data sets. Thus, taxa were classed as calcitic if they have a calcitic component that could, in the absence of aragonite, be identified to species level; in the event of wholesale aragonite loss and all else being equal, these taxa should dominate the fossil record. We note that our compositional categories differ somewhat from those adopted in other recent studies of taphonomic bias related to shell mineralogy and microstructure (e.g., Kidwell 2005).

Our study was restricted to level-bottom, benthonic molluscs that are inferred to have been confined to, or to have ranged into, shelf water depths (i.e., $< \sim$200 m water depth). Pelagic taxa and taxa that were restricted to estuarine, rocky substrate, bathyal, or abyssal habitats were excluded. The study was restricted in this way because non-shelf marine molluscs are greatly underrepresented in the NZ fossil record (Beu and Maxwell 1990; Crampton et al. 2003; Cooper et al. 2006), reflecting the relatively low abundance of fossiliferous estuarine, bathyal, and abyssal facies and/or the low abundance of mollusc fossils within those facies. By limiting the analysis to level-bottom shelf taxa, we have minimized potential bias associated with uneven sampling of different paleoenvironments through time (e.g., Bush et al. 2004).

In addition, 167 collections from northernmost NZ were also excluded from analysis because they represent a comparatively rich but biogeographically distinct component of the mollusc fauna that is restricted to just a brief interval of time (21.7–16 Ma [Beu and Maxwell 1990]); their inclusion introduces a small bias in diversity estimates for that interval (Crampton et al. 2006).

Prior to analysis, fossil lists in the FRED data set were subjected to several iterations of automated and manual checking and updating. These data cleaning procedures are described in Crampton et al. 2006. Most importantly, a large number of junior synonyms were replaced by their revised names: out of 5189 genus ± subgenus ± species ± subspecies name combinations in the original data, 2731 or 62% required one or more updates.

Methods

Throughout this paper, comparisons between different time series are based conservatively on the nonparametric Spearman's rank-order correlation coefficient, r_s. We avoid the more commonly used Pearson's product-moment coefficient because it is more sensitive to outliers in the data. In addition, comparisons are based on first differences between successive time intervals—a simple approach to eliminating autocorrelation within time series and resultant induced correlations between series (e.g., McKinney 1990). Correlations we report, therefore, aim to express detrended agreement between time series.

Taxonomic Rank Considerations.—Most previous studies of paleobiodiversity have assumed that, compared to and as proxies for species, genera and families are taxonomically more stable and highly sampled in the fossil record and, therefore, are more reliable estimators of diversity history (e.g., Patterson and Smith 1989; Alroy 1998; Smith 2001; Forey et al. 2004; but cf. Signor 1985; Roy et al. 1996; Rosenzweig 1998; Scotland and Sanderson 2004). Many of the problems associated with use of the species category are reduced here because most of our data have been generated by a small number of molluscan workers using relatively consistent taxonomic approaches—indeed, 76% of faunal lists were prepared by just three taxonomists who worked in close collaboration (Crampton et al. 2006). In addition and as noted above, we have used various data cleaning procedures to reduce species-level taxonomic noise in the FRED data set.

Our diversity analyses, therefore, were undertaken at three taxonomic levels—family+subfamily, genus+subgenus, and species+subspecies (henceforth, family, genus, and species level). Thus, families and subfamilies were effectively regarded as a single, family-group rank and, unless stated, the nomi-

nate subfamily was assumed; similarly for genera and species. This follows common practice and reflects the fact that taxa and their sub-taxa are essentially interchangeable and very much subject to the whims of individual taxonomists.

Sampling Standardization.—Numerous studies have discussed biases in the fossil record related to variations in sampling intensity and their likely impacts on perceived patterns of biodiversity history (e.g., Raup 1976, 1979; Signor 1985; Sepkoski 1994; Alroy et al. 2001; Peters and Foote 2001). Existing methods for sampling standardization of paleobiodiversity data are based on randomized resampling of faunal lists, an extension of the method of analytical rarefaction (e.g., Raup 1975; Alroy 1996, 2000; Miller and Foote 1996; Alroy et al. 2001). For the time intervals of interest, taxonomic occurrences or entire lists are sampled randomly (without replacement) and, to overcome the effects of stochastic resampling error, repeatedly. Each resampling trial includes a specified number of occurrences or lists—the quota—and the results from many trials yield expected relative taxonomic diversities for each time interval for a given quota. When setting this quota, there is an inevitable tradeoff between the desire to use a large quota and thus reduce stochastic uncertainty, and the desire to use a value that is low enough to be satisfied by poorly sampled time intervals—intervals that cannot satisfy the quota are excluded from analysis. (It should be remembered that rarefaction curves cannot be extrapolated beyond their endpoints without making unreasonable assumptions regarding underlying statistical distributions [Tipper 1979].)

Many problems of sampling standardization of paleobiodiversity data remain unresolved—in particular, there is no well-tested method that can tease apart the complex effects of variations in alpha and beta diversities, evenness, and sampling intensity (Alroy 2000; Alroy et al. 2001; Powell and Kowalewski 2002; Bush et al. 2004). Furthermore, any sampling standardization protocol that is optimal at one taxonomic rank is likely to be suboptimal at another rank because of variations in taxon-abundance distributions between ranks. To solve these problems requires large databases of collection-based taxonomic data that include individual taxon abundance counts. Although such databases are being compiled (e.g., Peters 2004; Finnegan and Droser 2005; Kosnik 2005; Kowalewski et al. 2006), no comprehensive solutions to the problems of sampling standardization have yet been devised.

In the present study, taxon abundance data are lacking and we rely on methods that will yield, at best, approximations to true sampling-standardized diversity curves. We stress, ours is not a study of alpha diversity such as reported in Bambach (1977), Powell and Kowalewski (2002), and Bush and Bambach (2004). Following Bush et al. (2004), we assume that ''true'' relative genus-level diversity patterns will be bracketed by three resampling approaches:

1. A fixed quota based on the number of lists (''by-lists unweighted,'' LUW) (e.g., Smith et al. 1985). This approach assumes that apparent changes in average alpha diversity are due entirely to true variations in alpha diversity through time.
2. A fixed quota based on the number of lists weighted by occurrences to the power of x (O^xW) (Alroy 2000; Bush et al. 2004). This approach assumes that apparent variations in alpha diversity are due entirely to sample-size biases. The optimal value of x, i.e., the value that properly corrects for the effects of sample-size variation on apparent alpha diversity, is determined largely by the amount of beta diversity. Beta diversity refers to between-habitat diversity (within a region) and is defined as the ratio of the total number of taxa in a set of collection lists to the average number of taxa in each list (Bush et al. 2004). In terms of beta diversity, the NZ mollusc data are very similar to a global Eocene marine macroinvertebrate data set for which x_{optimal} at the genus level was determined empirically to be 1.4–1.3 (Fig. 1) (Bush et al. 2004); here we use a factor of 1.4.
3. O^xW and variable quotas for each time interval (O^xW, varying quota). The quotas are varied to reflect presumed changes in true

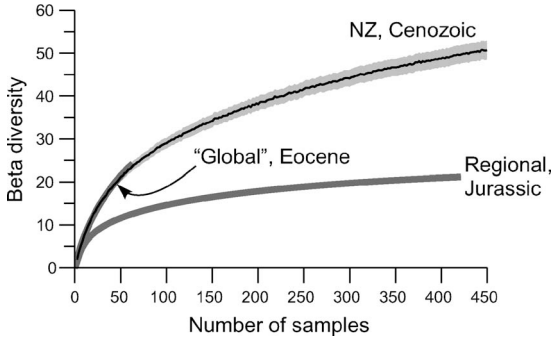

FIGURE 1. Beta diversity of the NZ Cenozoic mollusc data compared with global Eocene and Western European Late Jurassic marine macroinvertebrate data sets (from Bush et al. 2004). The NZ curve matches that of the global Eocene, suggesting that occurrences weighted to the power of 1.4 ($O^{1.4}W$) sampling-standardization is appropriate (Bush et al. 2004). The NZ curve is based on 1000 resamplings of the FRED data set; the error envelope is ±1 SE.

alpha diversity, using a running average of the ratio of O^xW to lists across a moving window of seven time bins (modified from Alroy 2000: pp. 720–721, eq. 9).

For any time interval, the number of taxa can be calculated in different ways. Two important counting protocols have been used elsewhere: within-bin taxa and boundary-crossers (e.g., Foote 2000; Alroy et al. 2001). The within-bin count sums all taxa that are actually sampled within a time bin; the boundary-crosser category sums taxa with stratigraphic ranges that cross a bin boundary, i.e., taxa that coexisted at an instant of time. The latter protocol has several important theoretical and practical advantages over the within-bin count, in particular relating to estimates of origination and extinction rates. Boundary-crossers are, however, subject to very pronounced edge-effect distortions that reduce apparent diversity within about two average taxon durations from the upper and lower boundaries of the study interval (Foote 2000: Fig. 6). Because the study interval here is only about ten times the average species duration, edge effects are significant and preclude use of the boundary-crosser count. For this reason, we focus here on within-bin taxa. This has another advantage, namely that the within-bin count is relatively insensitive to taxonomic noise (see below). This is because the within-bin count requires only that a taxon be distinguished in a given time bin to be counted; it does not matter whether that taxon is identified and named correctly or not. In contrast, the boundary-crosser protocol requires a taxon to be correctly and consistently identified and named in two or more bins before it will be counted.

Estimation of Taxic Rates.—The influence of edge-effect distortions, discussed above, also prevents calculation of per capita origination and extinction rates directly from the FRED data set (Foote 2000). Instead, we have used the inverse survivorship modeling approach of Foote (2003, 2005) applied to the synoptic data set of biostratigraphic ranges to calculate taxic rates. This method uses numerical maximization of a likelihood function to find the origination and extinction rates that yield the best agreement between predicted forward and backward survivorship probabilities and the corresponding observed values (Foote 2005). Uncertainty in the parameter estimates, due mainly to variance in the data but also to imprecision in the optimization procedure, was estimated by bootstrap resampling of biostratigraphic ranges (with replacement), the optimization procedure being repeated for each bootstrap sample. Because the approach is computationally expensive, the reported results are based on 100 bootstrap replicates. We verified that 100 replicates converged on a stable solution by comparing the means and standard deviations from the first 50 and second 50 solutions; results for these two subsets are only trivially different and thus we are confident that additional replicates would not significantly alter our parameter estimates.

In addition to estimates of origination and extinction rates, inverse survivorship modeling can also estimate sampling probability. In this study, however, we have used independently derived estimates of sampling probability from Crampton et al. (2006) to constrain our solutions for origination and extinction rates; these sampling probabilities were based on the FRED data set and calculated using the method of gap analysis by stage (CI_{bda} of Maas et al. 1995: p. 191).

The taxic rates reported below were calculated under the assumption of pulsed turn-

over, i.e., that origination and extinction were concentrated at a single episode per stage, rather than being dispersed continuously through each stage (Foote 2005). This assumption is justified by consideration of the deviations between the predicted and observed rates under the pulsed and continuous models: the pulsed deviations are consistently and significantly smaller than the continuous, indicating greater likelihood support for the pulsed model (results not presented). Pulsed turnover of NZ Cenozoic mollusc genera was predicted in a semiquantitative way by Beu (1983, 1990) and Beu and Maxwell (1990). Globally, the pulsed model is supported by inverse survivorship analysis of Sepkoski's (2002) compendium of marine animal genera (Foote 2005; see also Peters 2005).

Selection of Time Bins and Treatment of Collection Ages.—The time bins used in this study are derived from the local stages of the NZ geological timescale (e.g., Cooper 2004). Although correlations with the international timescale have changed substantially over the past 50 years, the system of local stages has remained relatively stable and, therefore, ages of fossil collections can be compared reliably, even when dealing with data of varying vintage.

Several workers have noted the importance of using uniform duration time bins in a study such as ours (e.g., Raup 1975; Alroy 1996, 1998). This follows because, all else being equal, one expects that a long interval of time would accrue greater diversity than a short interval of time. Given that the NZ stages are of uneven duration and represent the finest time resolution available to us, we have experimented with various stage groupings designed to reduce inequalities in time bin durations. Three arrangements of stages younger than 50 Ma are shown in Figure 2 and explained below:

1. Twenty-one ungrouped stages (mean duration 2.4 Myr, standard deviation 1.8 Myr).
2. Fourteen minimally grouped stages (mean duration 3.5 Myr, standard deviation 1.4 Myr).
3. Eight grouped stages (mean duration 6.2 Myr, standard deviation 1.2 Myr).

The impacts of different time bin arrangements on diversity estimates are discussed in the results section.

During estimation of diversity using the FRED data set, collection lists were assigned to time bins on the basis of their paleontologically determined age ranges (using both molluscs and any associated microfloras and faunas). Depending on the bins used and resolution of age-diagnostic fossils, age range constraints of some collections span more than one bin. At the level of New Zealand stages, 39% of lists fall into this category; at the level of eight grouped stages, 20% of lists have spanning ranges. For each resampling trial, such collections were assigned randomly to one or other of their possible time bins. Assignment probabilities for each bin were weighted by the relative duration of that bin as a proportion of the total possible stratigraphic range of the collection in question. This approach ensures that as many as possible of the data were retained in the analyses—an important consideration for some poorly sampled time bins that otherwise would have fallen below resampling quotas during standardization. The approach, however, also contributes additional variance to the results, could blur distinctions between adjacent time bins, and might bias diversity estimates. In fact, comparison of analyses in which bin-spanning collections were either omitted or included demonstrates that inclusion does not introduce bias into sampling-standardized diversity estimates: for all the time bin arrangements used here, standardized diversity curves with and without bin-spanning collections are almost identical and well within error (results not presented here).

Results

Effects of Time Bin Arrangement.—As average time bin duration increases, detail and structure are lost progressively from sampling-standardized diversity curves (Fig. 2). It is encouraging, however, to see that broad patterns are consistent between the three time bin arrangements in our trials here, suggesting that major features of biodiversity history are reasonably robust to unequal time bin durations. Furthermore, in the most highly resolved and

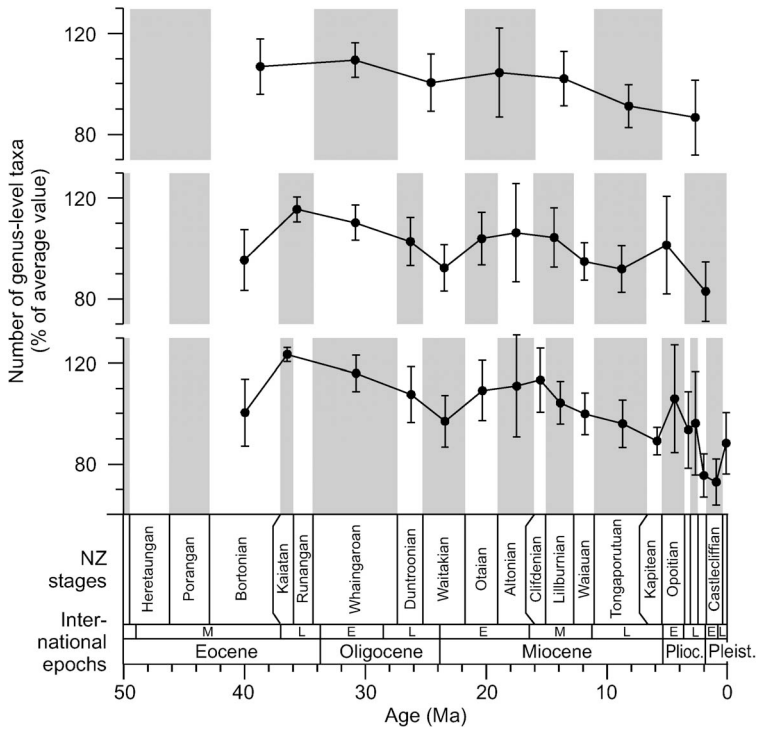

FIGURE 2. Sampling-standardized genus-level diversity curves for NZ shelf molluscs for the past 40 Myr, based on different time bins indicated (here and elsewhere) by the gray bars. Despite differences in bins, all curves show essentially the same long-term pattern. The curves are based on the FRED data set, the $O^{1.4}W$ sampling-standardization protocol, a quota of 1500, and 500 resampling trials; error bars indicate ± 1 SD. In this and subsequent figures, the local stages of the NZ timescale are shown against the international epochs (correlations from Cooper 2004); the unlabeled NZ late Pliocene stages are, from oldest to youngest, the Waipipian, Mangapanian, and Nukumaruan; the youngest NZ stage is the Haweran.

uneven arrangement of 21 ungrouped stages, there is only a weak, negative, and nonsignificant correlation between stage duration and detrended, sampling-standardized genus richness ($r_s = -0.084$, n.s.); this correlation is slightly stronger but still nonsignificant at the species level ($r_s = -0.123$, n.s.). This suggests that most of the structure in the sampling-standardized diversity data is due to factors other than unequal stage duration. Henceforth, most interpretation is based on the minimal grouping shown in the middle panel of Figure 2, although our major conclusions are insensitive to stage grouping arrangement. For comparisons of diversity with origination and extinction rates, however, we have used ungrouped stages that yield enough data points (compared to the number of parameters being estimated) for robust inverse survivorship modeling of taxic rates.

Evaluation of Taxonomic and Biostratigraphic Biases.—We evaluated the effect of taxonomic error on apparent biodiversity patterns by comparing analyses of the FRED data set based on the raw data and following taxonomic cleaning. Using the within-bin taxonomic counting protocol (an important caveat, see above), we find that taxonomic ''noise'' has negligible effect on apparent patterns of relative biodiversity change through time, although, unsurprisingly, the raw signals are elevated by the retention of junior synonyms, an error that increases slightly in magnitude toward the Recent (Fig. 3). This encouraging result accords well with previous studies that demonstrated comparatively little effect of taxonomic error on perceived biodiversity patterns (Sepkoski 1993; Adrain and Westrop 2000; but cf. Ausich and Peters 2005).

The question of appropriate taxonomic rank in paleobiodiversity studies is further complicated by possible effects of paraphyletic and

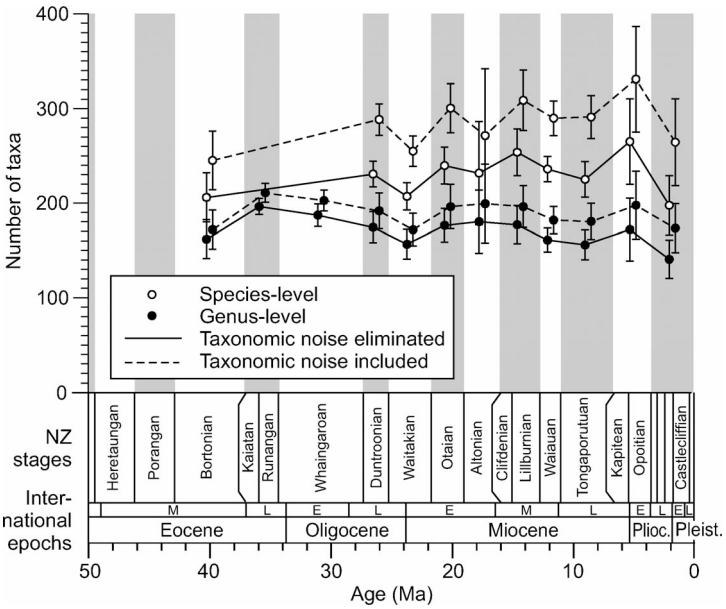

FIGURE 3. Sampling-standardized diversity curves for NZ shelf molluscs for the past 40 Myr, showing the effects of taxonomic noise in the FRED data set on perceived diversity patterns. The dashed curves are for data that include a large amount of known taxonomic error; the solid lines show the results once this error is eliminated. Taxonomic error has a negligible effect on the secular pattern of taxonomic diversity. Both analyses are based on the $O^{1.4}W$ sampling-standardization protocol, a quota of 1500, and 500 resampling trials; error bars indicate ± 1 SD.

monotypic groupings on perceptions of diversity (Smith and Patterson 1988; Sepkoski and Kendrick 1993). Recent work on these issues has shown that sampling rate and the number and size distribution of taxonomic groups are more important controls on the recovery of diversity information than taxonomic rank or paraphyly per se (Robeck et al. 2000). Thus, when sampling is good, the most reliable diversity estimates are achieved by using classifications with large numbers of small taxa. For the NZ Cenozoic molluscan record, genus and species sampling rates are indeed good—averaging 0.51 occurrences per taxon per Myr and 0.43 occurrences per taxon per Myr, respectively, over the 50 Myr time interval considered here[1] (Crampton et al. 2006: Appendix 1). In view of this, we infer that both species and genera have the potential to yield robust estimates of NZ Cenozoic shelfal molluscan diversity history, given appropriate sampling standardization.

We also wanted to identify possible biases in apparent biodiversity that are introduced by erroneous biostratigraphic ranges. An earlier study demonstrated that ranges of many species in the FRED data set are long compared to their ranges in the synoptic data set (Crampton et al. 2006: Fig. 4). In most cases, these range extensions in the FRED data set result simply from incorrect identifications of taxa in stages outside their true biostratigraphic distribution. (It is important to remember, however, that not all these range extensions in the FRED data set will be wrong and, in some unknown proportion of cases, the adopted ranges in the synoptic data set will in fact be too short.) Figure 4 shows, for each stage, the proportion of genus- and species-level taxonomic occurrences in the FRED data set that are within their biostratigraphic ranges according to the synoptic data set. For genera, this proportion is high and relatively constant through time, ranging between 79% and 98%. For species, the proportion is lower

[1] These averages are based on the results of gap analysis by stage, a method that detects long-term trends in sampling rate. The averages have been adjusted to take account of a maximum allowable, ''perfect'' sampling rate of 1.0 (Robeck et al. 2000; cf. Crampton et al. 2006).

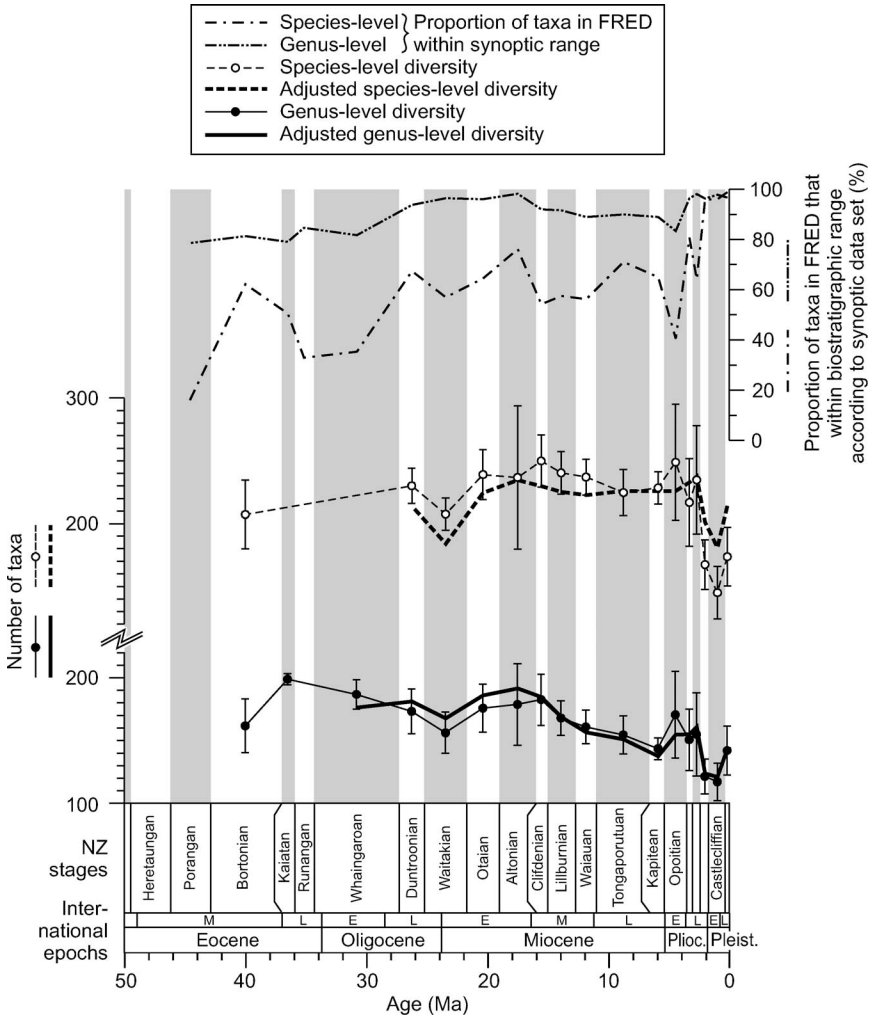

FIGURE 4. Plot showing the effects on apparent diversity of spurious biostratigraphic range extensions in the FRED data set. The top two curves show the proportions of genus- and species-level taxonomic occurrences in the FRED data set that are within biostratigraphic range according to the synoptic data set. Binomial errors on these curves are negligible and, for clarity, have been omitted. The lower curves show sampling-standardized diversity curves as calculated directly from the FRED data set and following adjustment for biostratigraphic range extensions. The plot reveals that spurious biostratigraphic range extensions in the FRED data set have only a minor impact on the major patterns of diversity change through time. The unadjusted curves are based on the $O^{1.4}W$ sampling-standardization protocol, a quota of 1500, and 500 resampling trials; error bars indicate ± 1 SD. The adjusted curves are based on the residuals of regressions of standardized diversity against proportion of taxa within range (both parameters detrended); these curves have been scaled to have the same mean as the unadjusted curves.

and more variable, and shows a conspicuous increasing trend through time, ranging from 41% to 99% (ignoring the stages that are too poorly sampled to yield sampling-standardized estimates of diversity). These proportions of FRED taxa that are within range are negatively and significantly correlated with stage-to-stage (i.e., detrended) changes in standardized diversity: $r_s = -0.4489$ ($p < 0.05$) and $r_s = -0.4769$ ($p < 0.05$) for genera and species, respectively. In other words, the presence of spurious range extensions in FRED explains some of the short-term structure that we see in our sampling-standardized diversity curves: as expected, where the proportion of taxa out of range is high, then apparent diversity is elevated, and vice versa.

To estimate the effect of this bias, for each taxonomic rank we calculated the residuals of a regression of standardized diversity against

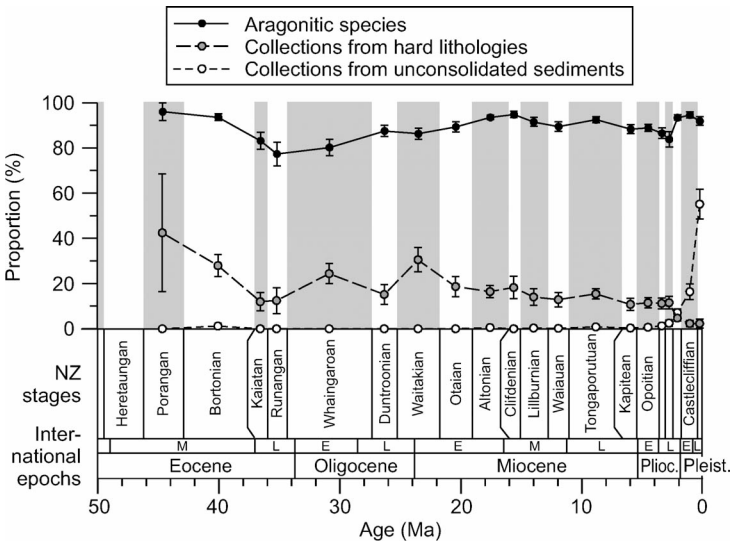

FIGURE 5. Potential taphonomic influences on apparent biodiversity. The solid line shows the proportion of aragonitic versus calcitic species in the FRED data set; for the purposes of this exercise, taxa were recorded as calcitic if they have a calcitic component that could be identified in the absence of aragonite. Because there is little temporal trend in the proportion of aragonitic taxa, shell mineralogy is unlikely to be a major determinant of sampling-standardized diversity in our data. The dashed lines show the proportions of collections from lithified sedimentary rocks versus unconsolidated sediments in the FRED data set. Note that for clarity, only the two end-members of the four hardness categories are plotted and, therefore, they do not sum to 100%; the ''moderately hard'' and ''moderately soft'' classes are relatively constant through time and are ignored here. The increase in hard lithologies and corresponding decrease in unconsolidated sediments with increasing age are likely to bias sampling-standardized diversity estimates downward in the older stages, although the magnitude of this bias has not been quantified here (see text for further discussion). Error bars indicate ± 1 SE (binomial), calculated as $SE(P) = \sqrt{P(1 - P)/n}$, where P is the proportion in question and n is the total number of observations.

proportion of taxa within range (both parameters detrended), and used these residuals to generate a new, adjusted, standardized diversity curve (Fig. 4). In most respects the adjusted and unadjusted curves differ little, particularly at the genus level. The two most important discrepancies between the curves are the removal of the diversity peak in the early Pliocene following adjustment, and reduced magnitude of the species-level diversity decrease in the late Pliocene and Pleistocene. We conclude, therefore, that, whereas spurious range extensions in FRED do have some effect on our sampling standardized diversity curves, they do not affect the major patterns of diversity change that we detect. Again, this result is consistent with earlier studies (Sepkoski 1993; Adrain and Westrop 2000; but cf. Ausich and Peters 2005). Henceforth, we use the unadjusted curves in our interpretations and avoid conclusions that might be significantly influenced by the biostratigraphic range extension bias.

Evaluation of Taphonomic Biases.—A much-discussed, potential large-scale taphonomic bias in the fossil record is the selective and non-uniform loss of taphonomically unstable, aragonitic taxa from the fossil record (e.g., contrast Bush and Bambach 2004; Kidwell 2005). Our data reveal that aragonitic forms dominate all stages over the past 45 Myr, constituting between 77% and 96% of raw species-level richness in the FRED data set (Fig. 5). (It should be noted here that very few identifications in the FRED data set are from molds or casts and that the calculated proportions of aragonitic species are based almost entirely on collections in which shell material is preserved; see discussion of this issue by Kowalewski et al. [2006].) Importantly, there is no secular trend apparent in our shell compositional data, the lowest values being in the late Eocene to early Oligocene, and in the late Pliocene. These two low points do not correspond to times of overall low mollusc diversity and, therefore, do not seem to mark times

of wholesale loss of diverse, aragonitic gastropod faunas. From these results we suggest that substantial loss of aragonitic taxa is unlikely and we see no evidence for large-scale bias of New Zealand's Cenozoic mollusc fossil record by aragonite dissolution. Our findings agree well with some global studies based on Phanerozoic molluscs that, similarly, have found little evidence for large-scale, systematic, taphonomic bias of the fossil record by aragonite loss (Kidwell 2005; Behrensmeyer et al. 2005). Our results are, however, in marked contrast to other studies of Paleozoic and Mesozoic faunas that have reported moderate- to large-scale dissolution of aragonitic molluscs (Cherns and Wright 2000; Wright et al. 2003; Bush and Bambach 2004).

Recently, using samples from the Miocene to Pleistocene of NZ, Hendy (2005) documented a substantial bias in the apparent record of fossil biodiversity related to lithification of sedimentary rocks. He suggested that unlithified sediments may yield up to twice the sampling-standardized genus-level diversity as lithified sedimentary rocks from comparable environmental settings, this difference being due to poor recovery of small and fragile specimens from lithified units (see also Behrensmeyer et al. 2005; Cooper et al. 2006; Kowalewski et al. 2006). Given that unlithified rocks are expected to be more common in younger successions, then biodiversity estimates may be biased upwards as one approaches the Recent or, more correctly, biased downward in older strata.

Information in the FRED data set reveals that unconsolidated sediments are, indeed, confined largely to units younger than about 3 Ma, and indurated lithologies show progressive increase in proportion back through time (Fig. 5). This apparent pattern is based on a subjective classification that will have been applied somewhat inconsistently by different workers and in strata of different age. Despite this, it undoubtedly reflects real features of NZ's stratigraphic record that are the result of basin evolution and diagenetic processes at an active margin. Because matrix hardness is recorded for only about one-third of collections (2324), and data are sparse for some stages, we are unable to perform separate sampling standardizations on collections from hard versus unconsolidated units (cf. Hendy 2005). It is likely, however, that a lithification bias of unknown magnitude affects our results and will have resulted in a relative over-estimation of sampling-standardized diversity in the younger stages, in particular in the youngest three stages. (At the species level, however, this bias will be contrary to the effects of spurious range extensions, discussed above.) The lithification effect may explain, in part, the trend of increasing per-stage sampling probability toward the Recent that was documented by Crampton et al. (2006), although this trend is also the result of increasing outcrop area through the Cenozoic.

Results of Sampling Standardization.—Figure 6 shows genus-level diversity curves based on the FRED data set and three different sampling standardization protocols explained above. Error bars are large compared with much of the structure in the plots, a reflection in part of the comparatively small quotas used in the analyses. We can determine, however, that the shapes of the curves do not change significantly when quotas are increased. This conclusion is based on examination of full resampling curves (e.g., Fig. 7), which reveal that diversity rankings for each of the time bins remain unchanged, even at high quotas. Hence, although error bars are relatively large, the curves are reliable representations of the inferred relative diversity history based on each of the three sampling standardization protocols.

As explained previously, the three curves shown in Figure 6 may be expected to bracket the ''true'' pattern of relative diversity change. Clearly, the curves are rather different, in particular for time bins older than 27 Ma. Despite these differences, the plots reveal several consistent and important patterns of diversity change through time:

1. To a first order, diversity has been approximately constant through much of the Cenozoic.
2. Diversity was relatively low around the Oligocene/Miocene boundary, in the late Miocene, and in the late Pliocene–Pleistocene. According to two of the plots, Neo-

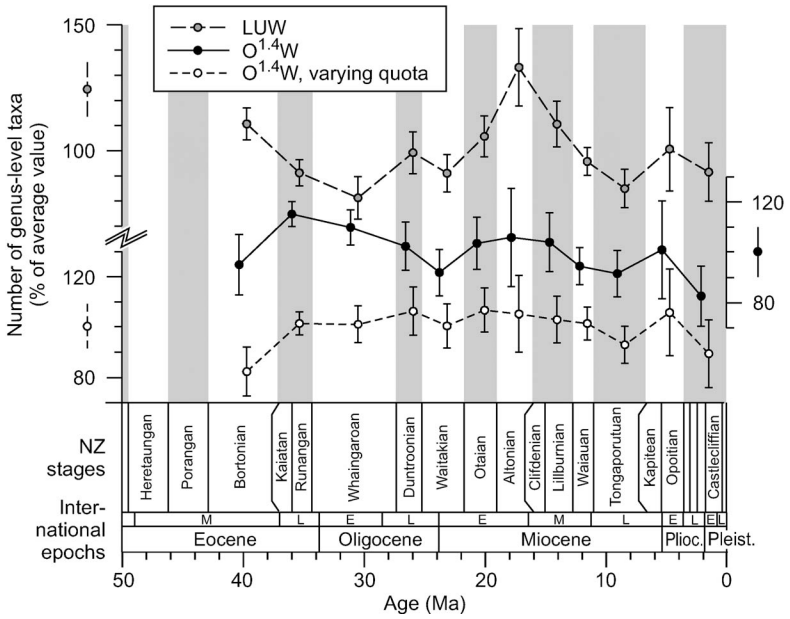

FIGURE 6. Genus-level diversity curves for NZ shelf molluscs for the past 40 Myr, based on the FRED data set and on different sampling standardization protocols. The top curve is based on the by-lists unweighted protocol (LUW) with a quota of 100 lists; the middle curve uses the $O^{1.4}W$ protocol with a quota of 1500; the bottom curve uses the $O^{1.4}W$ protocol and a variable quota based on presumed changes in true alpha diversity ($O^{1.4}W$ varying quota; see text for further explanation). Because the absolute values of standardized diversity are dependent on the quotas used, and to make the curves comparable, the data are shown as percentages of average within-curve diversities. Despite differences in protocol, all curves show essentially the same long-term pattern of roughly stable to declining diversity. Elsewhere in this paper our interpretations are based on the $O^{1.4}W$ sampling-standardization protocol (see text for further explanation). Note that, for clarity, y-axes for each plot have been shifted so as to separate the curves. All three curves are based on 500 resampling trials; error bars indicate ±1 SD.

gene minimum diversity was in the late Pliocene–Pleistocene.

3. There were diversity peaks in the early Miocene and Pliocene; the latter peak is probably smeared back in time in Figure 6 by the effects of spurious range extensions (cf. Fig. 4).

The possible significance of these patterns is discussed in more detail below. Elsewhere in this paper, for simplicity, we represent inferred genus-level biodiversity history using the central plot of Figure 6, based on $O^{1.4}W$. This curve is selected to represent the three because it displays intermediate levels of structure and variability

The Effects of Taxonomic Rank.—Comparative, sampling-standardized diversity plots at the family, genus, and species levels are shown in Figure 8. All three of these are based on the $O^{1.4}W$ sampling standardization protocol. As noted earlier, this protocol is apparently appropriate at the genus level, as determined empirically, but may be sub-optimal at other taxonomic ranks. For this reason, we cannot draw firm conclusions regarding relationships between diversity curves at different taxonomic levels. Figure 8 does suggest, though, that patterns of species-level diversity change are likely to be captured at the genus and family levels, albeit with progressively dampened signals as expected given the nesting of branches on the evolutionary tree (e.g., Sepkoski 1997). This tentative conclusion supports many previous studies that have assumed that trends of family- or genus-richness in the fossil record can provide valid proxies for species-level diversity patterns through time (e.g., Sepkoski et al. 1981; Valentine 1985a).

Interpretation of Taxic Rate Estimates.—Modeled taxic rates are shown in Figure 9. The origination and extinction curves are similar between taxonomic ranks and display consistent patterns of variation:

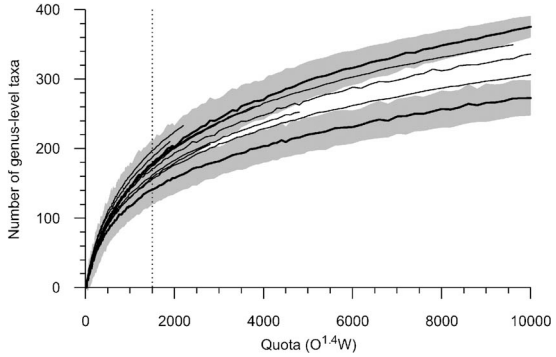

FIGURE 7. Full genus-level resampling curves for the FRED data set, the time bins shown in the middle panel of Figure 2, and the $O^{1.4}W$ sampling-standardization protocol. Each curve shows the expected diversity for a single time bin given increasing resampling quotas. The curves for different time intervals do not cross, indicating that rank-order diversity among the intervals is not sensitive to the particular quota used for sampling standardization. Representative uncertainty intervals (±1 SD, shaded regions) are shown for the two time bins indicated with bold lines (0–3.6 Ma, 12.7–16 Ma). The dotted line marks the quota used to generate diversity curves shown in other figures. The curves are based on 500 resampling trials.

1. Origination rates for genera and species were, episodically, relatively high prior to the middle Miocene, with major peaks in the late middle Eocene, late Oligocene, and late early Miocene.
2. Origination rates were comparatively low from the middle Miocene to Recent.
3. Extinction rates remained episodically high throughout the Cenozoic, although at the species level there may be a decreasing trend for the late Miocene to Recent. Peaks in extinction occurred in the early late Eocene, latest Oligocene to earliest Miocene, late early Miocene, late middle Miocene, early late Pliocene, and late late Pliocene.

Origination and extinction rates (Fig. 9) can be used to predict the total number of taxa that actually lived during each time interval (Fig. 10). If N_b is the number of taxa extant at the start of interval i and P is the origination rate, then $N_b(1 + P_i)$ is the total interval diversity (Foote 2003). N_b is itself proportional to $\Pi_{j=1}^{i-1}(1 + P_j)(1 - Q_j)$, where Q is the extinction rate. The actual value depends on an arbitrary scaling constant, namely the true diversity at the start of the time series, which can be set to unity without affecting the shape of the modeled diversity curve.

The fluctuations in modeled diversity do not consistently match those in sampling-standardized diversity. The two diversity estimates nonetheless both show a clear downward trend, in contrast to the increasing trend in raw diversity measures (Fig. 10 and see below). Given that the sampling-standardized diversity curve and the taxic rates are derived from different data sets and by different approaches, the observed consistency is encouraging.

Comparisons Between Sampling-Standardized and Raw Diversities.—Differences between both sampling-standardized and modeled diversity patterns and those inferred from the raw FRED data set are shown in Figure 10. To compare overall trends in the various data series, we computed robust linear regressions using Tukey's (1977) method, as implemented in the R programming language (R Development Core Team 2004). The equations for these regressions are given in Table 1. Because the data represent time series and points are, therefore, non-independent, we cannot evaluate the statistical significance of the slopes in a straightforward way. Instead, we implemented a bootstrap procedure and, for each series, resampled the data points as if they were independent. From 1000 bootstrap trials, we calculated the standard error of the slope and used this to test for a significant difference from zero using a two-tailed test and the normal approximation. In addition, we also calculated explicitly the 95% confidence interval on the slope using the bootstrap distribution (i.e., the values bracketing the mean ± 47.5% of the distribution). These probabilities and confidence intervals are reported in Table 1.

From these results, it is clear that the species raw diversity curve, and almost certainly the genus raw curve also, displays significant trends of increasing diversity toward the Recent (Table 1). From our analyses, we infer that the trend of increasing raw diversity through time is an artifact of a sampling bias that favors recovery of fossils from younger strata (Crampton et al. 2003, 2006). In contrast, the standardized and total progeny curves show

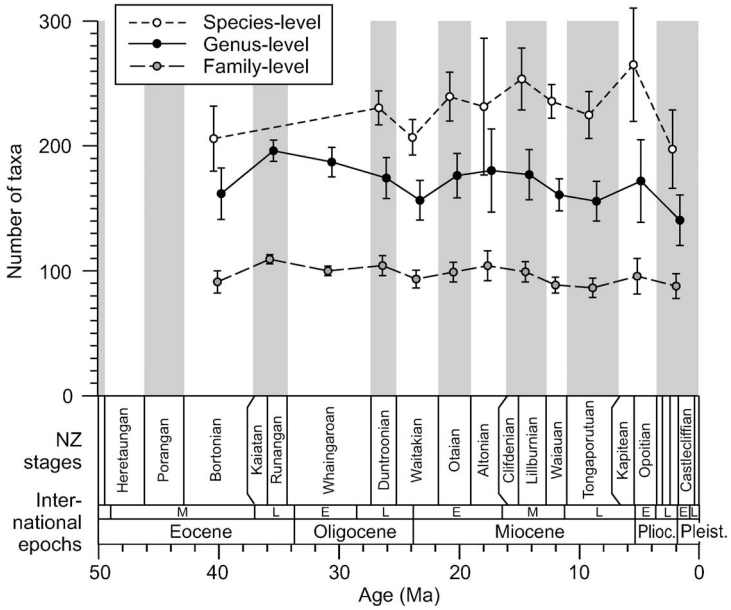

FIGURE 8. Family-, genus-, and species-level diversity curves for NZ shelf molluscs for the past 40 Myr, based on the $O^{1,4}W$ sampling-standardization protocol, a quota of 1500, and 500 resampling trials. Lack of long-term diversity increase is evident regardless of taxonomic level. Error bars indicate ±1 SD. Note that missing data points for the species-level curve indicate time bins that failed to satisfy the resampling quota.

apparently decreasing trends through time (Fig. 10); at the genus level, the slopes of these trends are significantly different from zero, whereas for species the slopes are only marginally different from zero (Table 1).

The other pronounced difference between the raw and sampling-standardized curves is the large amount of variation in the former. This variation results from uneven sampling between stages resulting from variations in outcrop area and preservation potential, factors that are themselves related (in part at least) to sequence stratigraphic controls (Crampton et al. 2003, 2006).

Discussion

To a first order, sampling standardized molluscan diversity curves derived here are robust to a range of biases and to variations in approach that reflect unresolved methodological uncertainties. For this reason, we suggest that the major patterns we describe are likely to reflect true underlying features of NZ marine diversity history. If anything, our inferred constant to declining diversity may be *conservative* and biased *upward* as we approach the Recent, owing to the effects of lithification bias, described above.

Predictably (perhaps), many aspects of our diversity and taxic rate curves presented above accord well with earlier, qualitative observations. In particular, Beu and Maxwell (1990) discussed in great detail patterns of Cenozoic molluscan faunal change in NZ. For example, they recorded the largest known increase in molluscan diversity beginning in the late Oligocene—the time of high genus- and species-level originations according to our analyses—and maximum diversity in the early Miocene. They went on to describe progressive extinction of taxa through the latest middle and late Miocene to result in an impoverished terminal Miocene fauna, a pattern consistent with our results. Lastly, the change from latest Pliocene to early Pleistocene molluscs ''constitutes one of the most dramatic of faunal turnovers in the NZ Cenozoic fossil record'' (Beu and Maxwell 1990: p. 329) and is clearly revealed in our analyses as a major drop in diversity corresponding to a peak in extinction rate.

Given our conclusion that certain aspects of

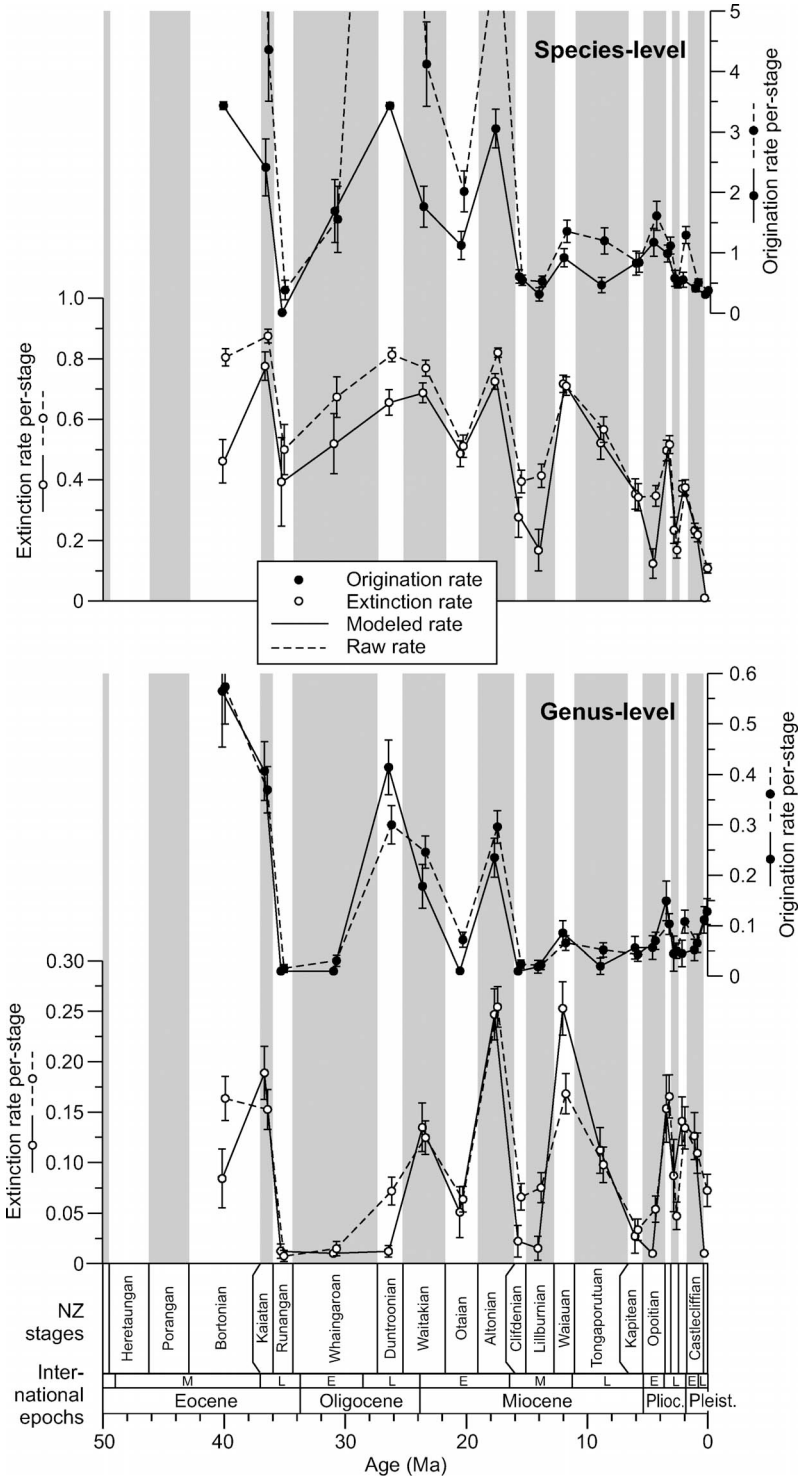

FIGURE 9. Raw and modeled genus- and species-level origination and extinction rates, based on the synoptic data set. The modeled rates were calculated using inverse survivorship modeling, with empirically constrained sampling probabilities, assuming pulsed turnover, and using ungrouped stages (see text for further explanation). Under the pulsed model, extinction rate is the total extinctions in an interval divided by the total interval diversity, and origination rate is total originations divided by diversity at the start of the interval. Error bars indicate ±1 SE. Raw origination values that extend off the top plot are 32.3 in the Bortonian, 16.9 in the Duntroonian, and 6.2 in the Altonian.

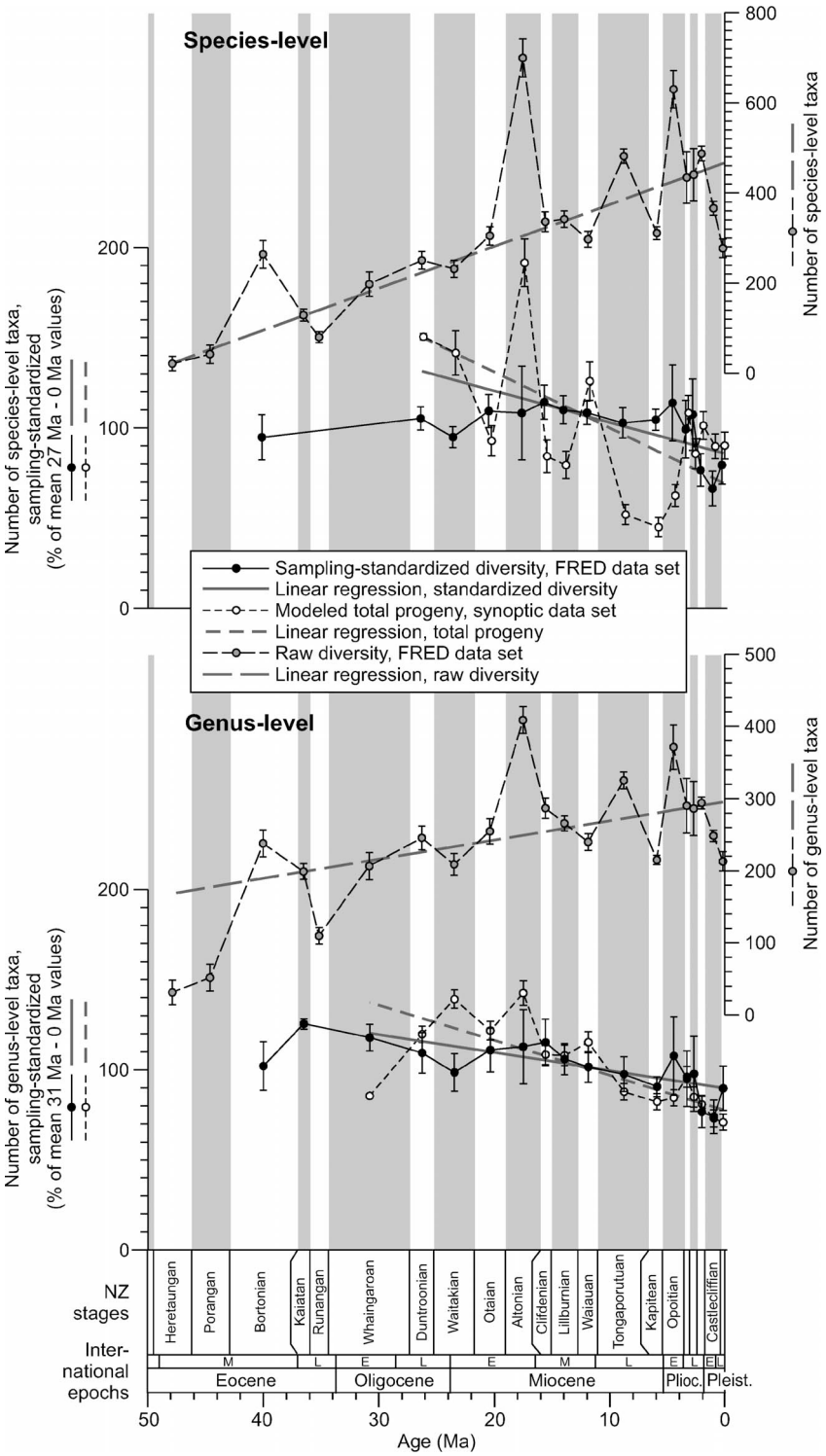
Species-level
Genus-level
Number of species-level taxa, sampling-standardized (% of mean 27 Ma - 0 Ma values)
Number of genus-level taxa, sampling-standardized (% of mean 31 Ma - 0 Ma values)
Number of species-level taxa
Number of genus-level taxa
Sampling-standardized diversity, FRED data set
Linear regression, standardized diversity
Modeled total progeny, synoptic data set
Linear regression, total progeny
Raw diversity, FRED data set
Linear regression, raw diversity
NZ stages
Heretaungan
Porangan
Bortonian
Kaiatan
Runangan
Whaingaroan
Duntroonian
Waitakian
Otaian
Altonian
Clifdenian
Lillburnian
Waiauan
Tongaporutuan
Kapitean
Opoitian
Castlecliffian
International epochs
Eocene
Oligocene
Miocene
Plioc.
Pleist.
Age (Ma)

TABLE 1. Equations of the robust linear regressions shown Figure 10, followed by standard errors of the slopes and corresponding probabilities that the slopes differ from zero, and independent estimates of the 95% confidence intervals for the slopes calculated from bootstrap distributions (see text). Bold values mark confidence limit estimates that support non-zero slopes at >95% level of confidence.

Regression		Equation of regression	Standard error of slope	Prob. that slope differs from 0*	95% confidence interval on slope†
Species	Standardized diversity	$y = 1.74t + 86$	0.97	0.073	−0.4, 3.2
	Total progeny	$y = 3.10t + 69$	2.32	0.180	−1.6, 7.0
	Raw diversity	$y = -9.30t + 468$	2.68	**<0.001**	**−13.0, −3.0**
Genus	Standardized diversity	$y = 1.00t + 90$	0.49	**0.041**	**0.2, 2.2**
	Total progeny	$y = 1.95t + 77$	0.75	**0.009**	**0.3, 3.5**
	Raw diversity	$y = -2.66t + 296$	1.83	0.150	**−6.9, −0.1**

* Two-tailed test using the normal approximation.
† Derived from the bootstrap distribution.

raw diversity are beset with sampling problems, it may seem odd that raw and modeled taxic rates appear at first glance to agree reasonably well in a variety of features. There are two important points to consider here, however. First, the agreement is not so complete that raw rates can be taken as a good proxy for true underlying rates. There are several discrepancies between raw and modeled rates that suggest that the raw rates are affected by backward and forward smearing of extinction and origination rates. Moreover, the average levels of the raw curves, while similar to the modeled curves at the genus level, are substantially higher than the modeled curves at the species level. The exaggeration of turnover rates under the pulsed model is predicted as an artifact of incomplete sampling (Foote 2000: Fig. 5); it therefore seems reasonable that species would be more severely affected.

The different response of species and genera to short-term variation in sampling is seen more clearly if we decompose total stage diversity into singleton taxa and taxa crossing one or both stage boundaries (Fig. 11). At the genus level, total diversity is dominated by long-ranging taxa extant and sampled both before and after the stage. Table 2 gives the detrended correlations between raw diversity and sampling probability estimated from gap analysis (Crampton et al. 2006: Appendix 1). Although sampling variation is positively correlated with total genus diversity, the correlation is not very strong. At the species level, by contrast, total diversity is dominated by singletons, which are considered to be comparatively sensitive to sampling variation (Foote 2000). Consistent with this view, the correlation between sampling probability and total diversity is substantially higher than at the genus level.

Second, the raw rates in Figure 9 are based on the synoptic Data, in which the effects of variable sampling may be partly mitigated by ranging taxa through between their first and last appearance. The FRED data, with which we count only taxa sampled actually within a time bin, are evidently more severely affected

←

FIGURE 10. Comparisons of genus- and species-level raw diversity curves with sampling-standardized curves and total progeny for NZ shelf molluscs for the past 40 Myr. General trends in the curves are indicated by linear regressions (equations given in Table 1). For both species and genera, the raw curves display increasing trends through time, whereas the sampling standardized and total progeny curves show marginally to significantly decreasing trends through time (Table 1). The linear regressions are based on Tukey's (1977) method for robust line fitting, as implemented in the R programming language (R Development Core Team 2004). The sampling-standardized curves are based on the FRED data set, the $O^{1.4}W$ sampling-standardization protocol, a quota of 1500, and 500 resampling trials; error bars indicate ±1 SD. The total progeny curves were calculated using the modeled taxic rates, as explained in the text; error bars indicate ±1 SE and were estimated from 100 model fits. Because the absolute values of standardized diversity and total progeny are dependent on arbitrary scaling factors, and to make the curves comparable, these two data series are shown as percentages of average within-curve, Oligocene to Pleistocene diversities. Error bars on the raw curves indicate ±1 SE and were estimated by bootstrapping (100 trials).

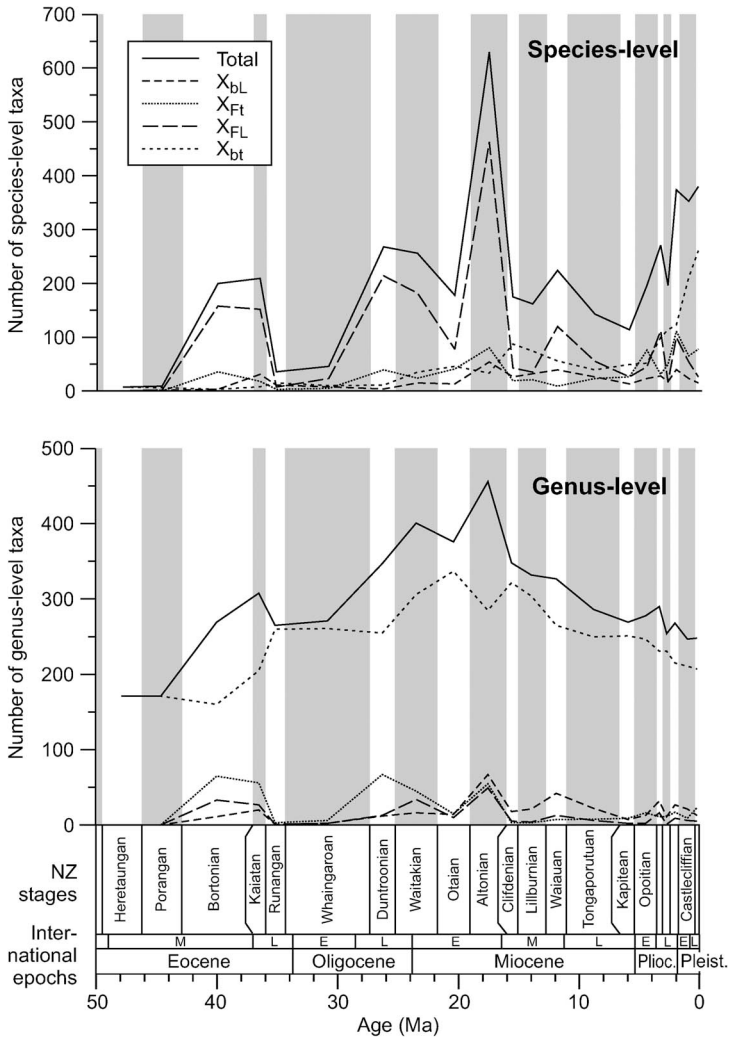

FIGURE 11. Raw diversity in the synoptic data set, decomposed into subgroups of taxa confined to a stage (X_{FL}) and those crossing one (X_{bL}, X_{Ft}) or both (X_{bt}) stage boundaries (Foote 2000). Genus diversity is dominated by long-ranging taxa, whereas species diversity is dominated by single-stage taxa.

by sampling variation (Table 2). Thus, despite some agreements between raw and modeled rates in Figure 9, there is clearly a substantial need for sampling-standardization and other procedures that offset temporal variability in sampling.

The inferred history of approximately constant diversity of shallow marine molluscs in the NZ region for much of the past 50 million years is contrary to the widely accepted pattern of increasing richness that has been recorded for the global marine fauna. This result raises a number of important questions: (1) is our result typical of other regions and other clades; (2) can changes in diversity be explained in part by the species-area effect; and (3) can the diversity curve be explained in part by changes in marine temperature?

TABLE 2. Rank-order correlation coefficients between per-stage sampling probability and raw measures of total stage diversity (40 Ma—late Pleistocene).

	Synoptic data set	FRED data set
Genus	0.375	0.904
Species	0.635	0.857

First, is the history of marine mollusc diversity in NZ typical of other clades and other regions? With respect to patterns in other clades, we cannot yet answer this question, al-

though suitable information exists in the FRED data set to test this in the NZ context (in particular, using foraminifera). Likewise, we cannot answer this question for other regions, and we emphasize the point made elsewhere that regional studies are now key to understanding global diversity history. In particular, if global marine diversity has indeed increased dramatically through the Cenozoic, and if NZ is representative of midlatitude regions, then we infer that the global increase must have been driven largely from the tropics or by increasing faunal provinciality. The only regional diversity history of tropical, shallow marine molluscs that is known to us—a study of Caribbean faunas by Jackson et al. (1999)—does not reach a firm conclusion as to whether diversity has increased or remained more or less constant over the past 10 Myr (but see Jackson and Johnson 2001). Alternatively, the apparent global increase in diversity may be nothing more than an artifact of sampling bias or analytical methods, a possibility raised by some recent studies (Alroy et al. 2001; Peters and Foote 2001).

Second, NZ is a continent that has been largely but variably inundated for much of the Cenozoic. The area of shelf has changed by a factor of almost five over the past 50 Myr as a result of tectonic processes. Given the well-established positive relationship between diversity and habitat area (e.g., MacArthur and Wilson 1967; Rosenzweig 1995), do we see a corresponding species-area effect in NZ Cenozoic molluscs? This relationship is typically expressed as $S = kA^z$, where S is species richness, A is the habitable area, and k and z are constants that are particular to a given taxonomic group and region; z is typically in the range 0.2 to 0.3 (Vermeij 2005). To examine this question, we compiled measurements of approximate shelf area based on digitized versions of the paleogeographic maps of King (2000) and King et al. (1999), with the addition of an unpublished map for 16 Ma. These maps show the inferred position of the coastline and the shelf break for nine times in the Cenozoic, based on interpretation and integration of seismic lines, seismic facies mapping, and stratigraphic analyses of a large amount of onshore and offshore sedimentological information (King et al. 1999, and references therein). To derive shelf area estimates for the time bins used in the diversity analysis, we interpolated between area measurements using a spline smoothing function (Fig. 12A). Plotting the sampling-standardized number of species against shelf area for the past 50 Myr, it is clear that the exponent z is statistically indistinguishable from 0 (Fig. 12B). In other words, we are unable to demonstrate a species-area effect for the NZ mollusc data. This conclusion is perhaps surprising, but it mirrors the results of other studies of marine molluscs (e.g., Schopf et al. 1978; Valentine and Jablonski 1991; McRoberts and Aberhan 1997; Roy et al. 1998) and may be interpreted in a number of ways.

1. Our species richness and shelf area values span only narrow ranges of magnitude and, therefore, may be simply inadequate for demonstrating an association between the two parameters.
2. Our measurements of shelf area are based on many underlying assumptions and approximations and are, at best, crude estimates. Furthermore, they fail to factor the effects of eustatic sea-level fluctuations that will have altered shelf area at a range of time scales finer than the first-order changes shown in Figure 12A. Depending on the local slope of the shelf, changes in shelf area related to eustasy may have been significant and it is possible that these effects have masked an underlying species-area relationship.
3. Marine invertebrates on the shelf typically have geographic ranges that are broadly parallel to the margin and narrow in the margin-perpendicular direction. For this reason, they may be insulated from the effects of changes in shelf area per se, which affect mainly the small, margin-perpendicular component of habitat area (Valentine and Jablonski 1991).
4. It is possible that habitat area per se is not a primary determinant of species richness in the marine realm, but instead, richness is the result of a complex interplay of many factors such as tectonism, nutrient supply, oceanographic factors, environmental sta-

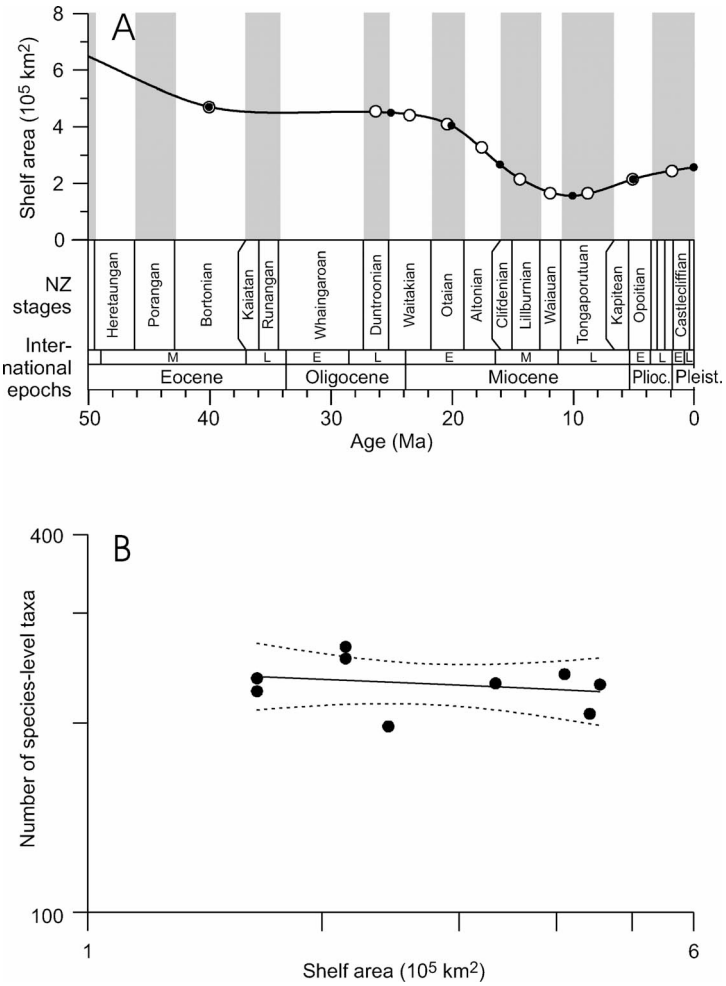

FIGURE 12. A, Area bounded by the inferred shoreline and shelf break on the NZ continent for the past 50 Myr as measured from the paleogeographic maps of King (2000) and King et al. (1999), with addition of an unpublished map for 16 Ma. Measured areas are shown as filled circles; the solid line is a spline smoothing function that has been fitted to these points (this curve is constrained by two additional points at 56 and 65 Ma). Open circles indicate interpolated values that are used in the species-area plot. Note that the calculated shelf area includes non-contiguous shelf ''islands'' and, at 40 Ma, was truncated along the Norfolk Ridge, to the north of NZ, at a modern-day latitude of about 30°S. B, Species-area plot shown against log-log axes. The sampling-standardized species data are based on the FRED data set, the $O^{1.4}W$ sampling-standardization protocol, a quota of 1500, and 500 resampling trials (the curve shown in Fig. 2 and elsewhere). The equation of the best-fit curve is $S = 460A^{-0.055}$; dotted lines are the 95% confidence interval about this line. Shelf area is not a predictor of species richness.

bility and stress, together with habitat area (McRoberts and Aberhan 1997; Martin 2003; Vermeij 2005). In particular, declining shelf area on the NZ continent during the Neogene resulted from propagation of the modern Australia-Pacific plate boundary through the region and was accompanied by increasing structural complexity, basin differentiation, and topographic relief—factors that may have elevated habitat and beta diversities even as shelf area declined.

Finally, there is a well-documented inverse relationship between latitude and molluscan diversity (e.g., Jablonski et al. 2000; Crame 2001; Crame and Rosen 2002; Hillebrand 2004) and increasing evidence to suggest that this relationship is a function of environmental temperature and available energy (Turner et al. 1996; Roy et al. 1998; Jablonski et al. 2000; Allen et al. 2002). Given that NZ has moved northward by about 10° latitude over the past 40 Myr (Sutherland et al. 2001), we might ex-

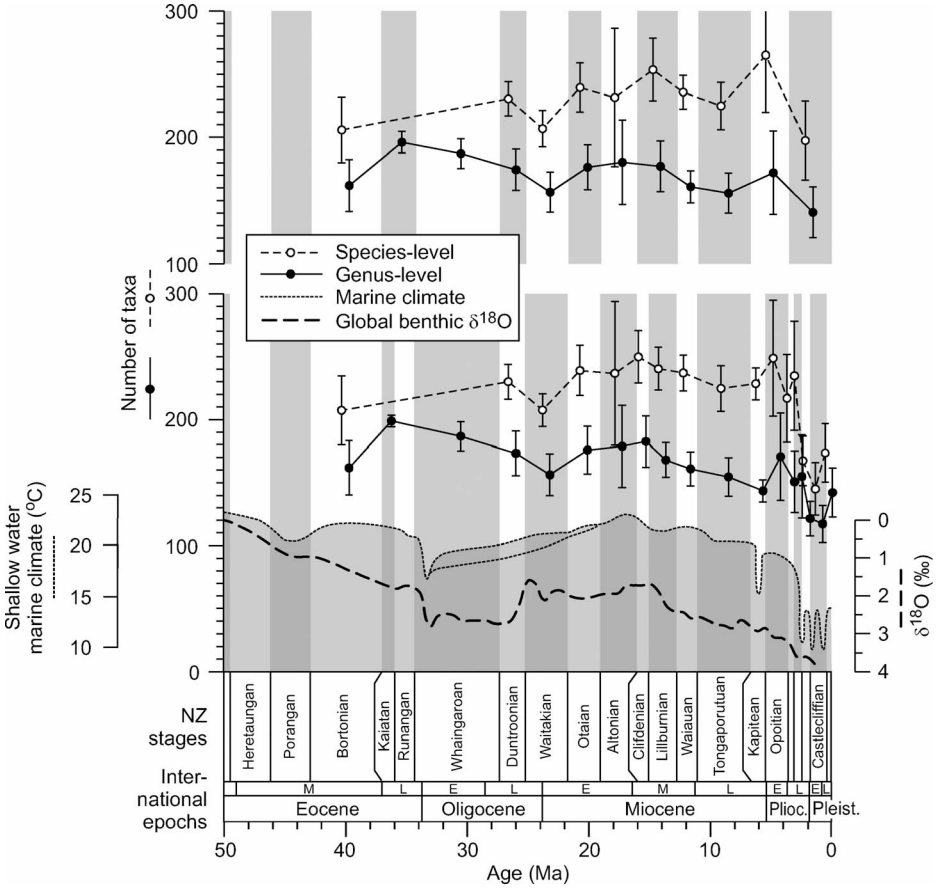

FIGURE 13. Genus- and species-level sampling-standardized diversity curves for NZ shelf molluscs for the past 40 Myr compared with regional marine climate. For reference, we also show the global benthic $\delta^{18}O$ curve, which depicts the combined effects of marine temperature and ice-volume increase through the Cenozoic. The diversity curves are based on the $O^{1.4}W$ sampling-standardization protocol, a quota of 1500, and 500 resampling trials. Diversity curves are shown for two different arrangements of stages into time bins, as indicated by the gray bars. The shallow water marine climate curve for NZ is normalized to latitude 42°S and is modified from Hornibrook (1992; see also Carter et al. 2004). The two climate curve segments in the Oligocene and early Miocene represent the south and west of the South Island (upper curve) and east coast of the South Island (lower curve). The global ocean benthic foraminifer oxygen isotope record is replotted from Zachos et al. (2001); the curve shown is a LOESS smooth based on the five-point moving average of the original data.

pect a corresponding increase in diversity toward the Recent. This effect, however, may have been offset by global cooling over the same period. To examine this question, we have used the regional shallow marine climate curve of Hornibrook (1992), which is based on the southernmost records of warm-water invertebrates and terrestrial biota and normalized to latitude 42°S (Fig. 13). This marine climate record represents the net effects of latitudinal drift combined with climate change. Correlations, based on first differences, between our diversity curves and both this regional temperature curve and the global $\delta^{18}O$ curve of Zachos et al. (2001) are low and nonsignificant (results not reported here; temperatures were estimated from the marine climate curve at midpoints of time bins). Given the nature of the temperature data available to us, this is hardly surprising: the Hornibrook (1992) curve is really only semiquantitative, and the $\delta^{18}O$ curve reflects convolved changes in both global average benthic temperature and global ice volume. Using existing data, therefore, we cannot demonstrate a link between marine biodiversity and paleotemper-

ature, although this question requires further testing using refined shallow marine climate data.

Conclusions

We see no evidence for a mid- to late Cenozoic increase in shelfal marine diversity in the NZ region. On the contrary, we see that sampling-standardized diversity has been approximately constant for much of the past 40 Myr and, at the species and genus levels, has declined over the past ~5 Myr (Figs. 8, 10). This result is counter both to the widely accepted global signal and to the raw NZ pattern (Fig. 10). Assuming that the NZ result is representative of other taxonomic groups and other temperate faunal provinces, then this suggests that the apparent global signal: (1) is an artifact of sampling bias or analytical methods; (2) resulted from increasing provinciality; or (3) was driven by large increases in diversity in tropical regions.

The inferred decline in diversity toward the Recent is counter to the likely—but unquantified—effects of a lithification bias that may have elevated diversity estimates in younger strata (Fig. 5); for this reason, the estimated diversity decline toward the Recent may be conservative. Major features of our sampling-standardized diversity curves are robust to a range of other potential taphonomic and systematic biases and methodological uncertainties—notably selective loss of aragonitic faunas (Fig. 5), taxonomic noise (Fig. 3), biostratigraphic range errors (Fig. 4), choice of time bins (Fig. 2), choice of resampling protocol (Fig. 6), and taxonomic rank (Fig. 8).

We are unable to demonstrate either a species-area effect on diversity (Fig. 12) or a relationship between marine temperature and diversity (Fig. 13), although the latter question should be reexamined once refined marine temperature estimates become available.

Acknowledgments

This work was funded as part of Marsden Fund contracts GNS001 and GNS0404, administered by the Royal Society of New Zealand, and also supported in part by the U.S. National Science Foundation (grant EAR-0105609). We acknowledge use of information contained in the New Zealand Fossil Record File. For discussions, we would like to thank D. Jablonski. J.S.C. acknowledges the support and hospitality of the Department of the Geophysical Sciences, University of Chicago, during a study visit in 2004. For constructive and helpful reviews, we acknowledge M. Kowalewski and A. Smith.

Literature Cited

Adrain, J. M., and S. R. Westrop. 2000. An empirical assessment of taxic paleobiology. Science 289:110–112.

Allen, A. P., J. H. Brown, and J. F. Gillooly. 2002. Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. Science 297:1545–1548.

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

———. 1998. Diachrony of mammalian appearance events: implications for biochronology. Geology 26:23–26.

———. 2000. New methods for quantifying macroevolutionary patterns and processes. Paleobiology 26:707–733.

Alroy, J., C. R. Marshall, R. K. Bambach, K. Bezusko, M. Foote, F. T. Fürsich, T. A. Hansen, S. M. Holland, L. C. Ivany, D. Jablonski, D. K. Jacobs, D. C. Jones, M. A. Kosnik, S. Lidgard, S. Low, A. I. Miller, P. M. Novack-Gottshall, T. D. Olszewski, M. E. Patzkowsky, D. M. Raup, K. Roy, J. J. J. Sepkoski, M. G. Sommers, P. J. Wagner, and A. Webber. 2001. Effects of sampling standardization on estimates of Phanerozoic marine diversification. Proceedings of the National Academy of Sciences USA 98:6261–6266.

Ausich, W. I., and S. E. Peters. 2005. A revised macroevolutionary history for Ordovician–Early Silurian crinoids. Paleobiology 31:538–551.

Bambach, R. K. 1977. Species richness in marine benthic habitats through the Phanerozoic. Paleobiology 3:152–167.

———. 1999. Energetics in the global marine fauna: a connection between terrestrial diversification and change in the marine biosphere. Geobios 32:131–144.

Behrensmeyer, A. K., F. T. Fürsich, R. A. Gastaldo, S. M. Kidwell, M. A. Kosnik, M. Kowalewski, R. E. Plotnick, R. R. Rogers, and J. Alroy. 2005. Are the most durable shelly taxa also the most common in the marine fossil record? Paleobiology 31: 607–623.

Beu, A. G. 1983. New Zealand Neogene molluscan biostratigraphy. Fifteenth Congress of the Pacific Science Association. Programme, abstracts and congress information, Vol. 1. University of Otago, Dunedin. [Unpaginated.]

———. 1990. Molluscan generic diversity of New Zealand Neogene stages: extinction and biostratigraphic events. Palaeogeography, Palaeoclimatology, Palaeoecology 77:279–288.

Beu, A. G., and P. A. Maxwell. 1990. Cenozoic Mollusca of New Zealand. New Zealand Geological Survey Palaeontological Bulletin 58:1–518.

Bush, A. M., and R. K. Bambach. 2004. Did alpha diversity increase during the Phanerozoic? Lifting the veils of taphonomic, latitudinal, and environmental biases. Journal of Geology 112:625–642.

Bush, A. M., M. J. Markey, and C. R. Marshall. 2004. Removing bias from diversity curves: the effects of spatially organized biodiversity on sampling-standardization. Paleobiology 30: 666–686.

Carter, R. M., I. N. McCave, and L. Carter. 2004. Leg 181 syn-

thesis: fronts, flows, drifts, volcanoes, and the evolution of the southwestern gateway to the Pacific Ocean, eastern New Zealand. Pp. 1–112 *in* C. Richter, ed. Proceedings of the Ocean Drilling Program, Vol. 181. Scientific Results; Southwest Pacific gateways. Ocean Drilling Program, College Station, Tex.

Cherns, L., and V. P. Wright. 2000. Missing molluscs as evidence of large-scale, early skeletal aragonite dissolution in a Silurian sea. Geology 28:791–794.

Cooper, R. A., ed. 2004. The New Zealand geological timescale. Institute of Geological and Nuclear Sciences Monograph 22: 1–284.

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

Crame, J. A. 2001. Taxonomic diversity gradients through geological time. Diversity and Distributions 7:175–189.

Crame, J. A., and B. R. Rosen. 2002. Cenozoic palaeogeography and the rise of modern biodiversity patterns. *In* J. A. Crame and A. W. Owen, eds. Palaeobiogeography and biodiversity change: the Ordovician and Mesozoic-Cenozoic radiations. Geological Society of London Special Publication 194:153–168.

Crampton, J. S., A. G. Beu, R. A. Cooper, C. M. Jones, B. Marshall, and P. A. Maxwell. 2003. Estimating the rock volume bias in paleobiodiversity studies. Science 301:358–360.

Crampton, J. S., M. Foote, A. G. Beu, R. A. Cooper, I. Matcham, C. M. Jones, P. A. Maxwell, and B. Marshall. 2006. 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.

Finnegan, S., and M. L. Droser. 2005. Relative and absolute abundance of trilobites and rhynchonelliform brachiopods across the Lower/Middle Ordovician boundary, eastern Basin and Range. Paleobiology 31:480–502.

Foote, M. 2000. Origination and extinction components of taxonomic diversity: general problems. *In* D. H. Erwin and S. L. Wing, eds. Deep time: *Paleobiology*'s perspective. Paleobiology 26(Suppl. to No. 4):74–102.

———. 2003. Origination and extinction through the Phanerozoic: a new approach. Journal of Geology 111:125–148.

———. 2005. Pulsed origination and extinction in the marine realm. Paleobiology 31:6–20.

Forey, P. L., R. A. Fortey, P. Kenrick, and A. B. Smith. 2004. Taxonomy and fossils: a critical appraisal. Philosophical Transactions of the Royal Society of London B 359:639–653.

Hendy, A. J. W. 2005. Lithification and the measurement of biodiversity—is missing alpha stuck between a rock and a hard place? Geological Society of America Abstracts with Programs 37(7):117.

Hillebrand, H. 2004. On the generality of the latitudinal diversity gradient. American Naturalist 163:192–211.

Hornibrook, N. de B. 1992. New Zealand Cenozoic marine paleoclimates: a review based on the distribution of some shallow water and terrestrial biota. Pp. 83–106 *in* R. Tsuchi and J. C. J. Ingle, eds. Pacific Neogene: environment, evolution, and events. University of Tokyo Press, Tokyo.

Jablonski, D., K. Roy, and J. W. Valentine. 2000. Analysing the latitudinal diversity gradient in marine bivalves. *In* E. M. Harper, J. D. Taylor, and J. A. Crame, eds. Evolutionary biology of the Bivalvia. of the Geological Society of London Special Publication 177:361–365.

Jablonski, D., K. Roy, J. W. Valentine, R. M. Price, and P. S. Anderson. 2003. The impact of the Pull of the Recent on the history of marine diversity. Science 300:1133–1135.

Jackson, J. B. C., and K. G. Johnson. 2001. Measuring past biodiversity. Science 293:2401–2404.

Jackson, J. B. C., J. A. Todd, H. Fortunato, and P. Jung. 1999. Diversity and assemblages of Neogene Caribbean Mollusca of lower Central America. Bulletins of American Paleontology 357:193–213.

Kidwell, S. M. 2005. Shell composition has no net impact on large-scale evolutionary patterns in mollusks. Science 307: 914–917.

King, P. R. 2000. New Zealand's changing configuration in the last 100 million years: plate tectonics, basin development, and depositional setting. Pp. 131–145 *in* 2000 New Zealand Petroleum Conference Proceedings, 19–22 March 2000, Christchurch Convention Centre, New Zealand. Ministry of Economic Development, Wellington.

King, P. R., T. R. Naish, G. H. Browne, B. D. Field, and S. W. Edbrooke. 1999. Cretaceous to Recent sedimentary patterns in New Zealand. Institute of Geological and Nuclear Sciences Folio Series 1:1–35.

Kosnik, M. A. 2005. Changes in Late Cretaceous–early Tertiary benthic marine assemblages: analyses from the North American coastal plain shallow shelf. Paleobiology 31:459–479.

Kowalewski, M., W. Kiessling, M. Aberhan, F. T. Fürsich, D. Scarponi, S. L. Barbour Wood, and A. P. Hoffmeister. 2006. Ecological, taxonomic, and taphonomic components of the post-Paleozoic increase in sample-level species diversity of marine benthos. Paleobiology 32 [this issue].

Maas, M. C., M. R. L. Anthony, P. D. Gingerich, G. F. Gunnell, and D. W. Krause. 1995. Mammalian generic diversity and turnover in the Late Paleocene and Early Eocene of the Bighorn and Crazy Mountains Basins, Wyoming and Montana (USA). Palaeogeography, Palaeoclimatology, Palaeoecology 115:181–207.

MacArthur, R. H., and E. O. Wilson. 1967. The theory of island biogeography. Princeton University Press, Princeton, NJ.

Martin, R. E. 2003. The fossil record of biodiversity: nutrients, productivity, habitat area and differential preservation. Lethaia 36:179–194.

McKinney, M. L. 1990. Classifying and analysing evolutionary trends. Pp. 28–58 *in* K. J. McNamara, ed. Evolutionary trends. Belhaven, London.

McRoberts, C. A., and M. Aberhan. 1997. Marine diversity and sea-level changes: numerical tests for association using Early Jurassic bivalves. Geologische Rundschau 86:160–167.

Miller, A. I., and M. Foote. 1996. Calibrating the Ordovician Radiation of marine life: implications for Phanerozoic diversity trends. Paleobiology 22:304–309.

Patterson, C., and A. B. Smith. 1989. Periodicity in extinction: the role of systematics. Ecology 70:802–811.

Peters, S. E. 2004. Evenness of Cambrian-Ordovician benthic marine communities in North America. Paleobiology 30:325–346.

———. 2005. Geologic constraints on the macroevolutionary history of marine animals. Proceedings of the National Academy of Sciences USA 102:12326–12331.

Peters, S. E., and M. Foote. 2001. Biodiversity in the Phanerozoic: a reinterpretation. Paleobiology 27:583–601.

Powell, M. G., and M. Kowalewski. 2002. Increase in evenness and sampled alpha diversity through the Phanerozoic: comparison of early Paleozoic and Cenozoic marine fossil assemblages. Geology 30:331–334.

R Development Core Team. 2004. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Raup, D. M. 1975. Taxonomic diversity estimation using rarefaction. Paleobiology 1:333–342.

———. 1976. Species diversity in the Phanerozoic: an interpretation. Paleobiology 2:289–297.

———. 1979. Biases in the fossil record of species and genera. Bulletin of the Carnegie Museum of Natural History 13:85–91.

Robeck, H. E., C. C. Maley, and M. J. Donoghue. 2000. Taxonomy and temporal diversity patterns. Paleobiology 26:171–187.

Rosenzweig, M. L. 1995. Species diversity in space and time. Cambridge University Press, Cambridge.

———. 1998. Preston's ergodic conjecture: the accumulation of species in space and time. Pp. 311–348 *in* M. L. McKinney and J. A. Drake, eds. Biodiversity dynamics: turnover of populations, taxa, and communities. Columbia University Press, New York.

Roy, K., D. Jablonski, and J. W. Valentine. 1996. Higher taxa in biodiversity studies: patterns from eastern Pacific marine molluscs. Philosophical Transactions of the Royal Society of London B 351:1605–1613.

Roy, K., D. Jablonski, J. W. Valentine, and G. Rosenberg. 1998. Marine latitudinal diversity gradients: tests of causal hypotheses. Proceedings of the National Academy of Sciences USA 95:3699–3702.

Schopf, T. J. M., J. B. Fisher, and C. A. F. I. Smith. 1978. Is the marine latitudinal diversity gradient merely another example of the species-area curve? Pp. 365–386 *in* B. Battaglia and J. A. Beardmore, eds. Marine organisms: genetics, ecology, and evolution. Plenum, New York.

Scotland, R. W., and M. J. Sanderson. 2004. The significance of few versus many in the tree of life. Science 303:643.

Sepkoski, J. J., Jr. 1993. Ten years in the library: new data confirm paleontological patterns. Paleobiology 19:43–51.

———. 1994. Limits to randomness in paleobiologic models: the case of Phanerozoic species diversity. Acta Palaeontologica Polonica 38:175–198.

———. 1997. Biodiversity: past, present, and future. Journal of Paleontology 71:533–539.

———. 2002. A compendium of fossil marine animal genera. Bulletins of American Paleontology 363:1–560.

Sepkoski, J. J., Jr., and M. L. Hulver. 1985. An atlas of Phanerozoic clade diversity diagrams. Pp. 11–39 *in* Valentine 1985b.

Sepkoski, J. J., Jr., and D. C. Kendrick. 1993. Numerical experiments with model monophyletic and paraphyletic taxa. Paleobiology 19:168–184.

Sepkoski, J. J., Jr., R. K. Bambach, D. M. Raup, and J. W. Valentine. 1981. Phanerozoic marine diversity and the fossil record. Nature 293:435–437.

Signor, P. W. I. 1985. Real and apparent trends in species richness through time. Pp. 129–150 *in* Valentine 1985b.

———. 1990. The geologic history of diversity. Annual Review of Ecology and Systematics 21:509–539.

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., and C. Patterson. 1988. The influence of taxonomic method on the perception of patterns of evolution. Evolutionary Biology 23:127–216.

Smith, E. P., P. M. Stewart, and J. J. Cairns. 1985. Similarities between rarefaction methods. Hydrobiologia 120:167–170.

Spencer, H. G., B. A. Marshall, P. A. Maxwell, J. A. Grant-Mackie, J. D. Stilwell, R. C. Willan, H. J. Campbell, J. S. Crampton, R. A. Henderson, M. A. Bradshaw, J. B. Waterhouse, and J. Pojeta Jr. In press. Phylum Mollusca: chitons, clams, tusk shells, snails, squids, and kin. Chapter 10 *in* D. Gordon, ed. The New Zealand Inventory of Biodiversity; Vol. 1. Kingdom Animalia: Radiata, Lophotrochozoa, and Deuterostomia. Canterbury University Press, Christchurch.

Sutherland, R., P. King, and R. Wood. 2001. Tectonic evolution of Cretaceous rift basins in south-eastern Australia and New Zealand: implications for exploration risk assessment. Pp. 3–13 *in* K. C. Hill and T. Bernecker, eds. Proceedings of the PESA Eastern Australasian Basins Symposium, 25–28 November 2001, Melbourne, Victoria. Petroleum Exploration Society of Australia Special Publication. Australasian Institute of Mining and Metallurgy, Carlton, Victoria.

Tipper, J. C. 1979. Rarefaction and rarefiction—the use and abuse of a method in paleoecology. Paleobiology 5:423–434.

Tukey, J. W. 1977. Exploratory data analysis. Addison-Wesley, Reading, Mass.

Turner, J. R. G., J. J. Lennon, and J. J. D. Greenwood. 1996. Does climate cause the global diversity gradient? Pp. 199–220 *in* M. E. Hochberg, J. Clobert, and R. Barbault, eds. Aspects of the genesis and maintenance of biological diversity. Oxford University Press, Oxford.

Valentine, J. W. 1985a. Diversity as data. Pp. 3–8 *in* Valentine 1985b.

Valentine, J. W., ed. 1985b. Phanerozoic diversity patterns; profiles in macroevolution. Princeton University Press, Princeton, NJ.

Valentine, J. W., and D. Jablonski. 1991. Biotic effects of sea level change: the Pleistocene test. Journal of Geophysical Research 96:6873–6878.

Vermeij, G. J. 2005. From phenomenology to first principles: toward a theory of diversity. Proceedings of the California Academy of Sciences 56:12–23.

Vermeij, G. J., and L. R. Leighton. 2003. Does global diversity mean anything? Paleobiology 29:3–7.

Wright, P., L. Cherns, and P. Hodges. 2003. Missing molluscs: field testing taphonomic loss in the Mesozoic through early large-scale aragonite dissolution. Geology 31:211–214.

Zachos, J. C., M. Pagani, L. Sloan, E. Thomas, and K. Billups. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. Science 292:686–693.