

## CHAPTER SIXTEEN

# Speciation and extinction in the fossil record of North American mammals

JOHN ALROY

### Introduction

Paleontological data have been used for decades to address a series of very general and intrinsically interesting questions concerning speciation. Many of them are essentially microevolutionary, morphological or both. What is the relative prevalence of anagenesis and cladogenesis (Wagner & Erwin 1995)? Do constraints on morphology cause occupation of morphospace to slow down as diversity increases (Foote 1993)? Is morphological change gradual or punctuated across speciation events (Simpson 1944)?

A survey of the analytical paleobiology literature would reveal, however, that interest in all of these questions has waned over the last decade or two. The one topic relating to speciation that remains very popular is the quantification and modelling of turnover rates (Foote 1994b, 2000, 2003; Sepkoski 1998; Newman & Eble 1999; Kirchner and Weil 2000; Allen *et al.* 2006; Alroy 2008). Coincidentally and fortuitously, the explosion of molecular data sets and great improvements in phylogenetic methods have led to quantifying speciation rates by tracking the accumulation of lineages through time (Nee *et al.* 1992, 1994; Purvis *et al.* 1995; Magallon & Sanderson 2001; Roelants *et al.* 2007).

Nonetheless, paleontological research has focused far more strongly on taxonomic diversity than on speciation in recent years (Alroy 1996, 1998b, 2000; Miller & Foote 1996; Sepkoski 1997; Alroy *et al.* 2001, 2008; Connolly & Miller 2001; Peters & Foote 2001; Smith 2001; Jablonski *et al.* 2003; Bush *et al.* 2004; Krug & Patzkowsky 2004; Allen *et al.* 2006; Crampton *et al.* 2006). In addition to research on diversity at the global scale, there has been an explosion of studies on diversity at the community scale (Powell & Kowalewski 2002; Bush & Bambach 2004; Olszewski 2004; Peters 2004; Kosnik 2005; Kowalewski *et al.* 2006; Wagner *et al.* 2006; and many others). Because the latter kind of work concerns the richness and evenness of individual fossil collections from precise locations and stratigraphic horizons, it has no direct bearing on turnover.

When turnover rates do get attention, the paleontological literature has focused on two overriding issues: whether they can be explained using intrinsic dynamic mechanisms, such as either density dependence (Sepkoski 1978, 1979,

1984) or random sorting or competition between taxa that leads to long-term declines in average rates (Raup & Sepkoski 1982; Gilinsky 1994), or by extrinsic environmental controls, including catastrophic perturbations such as asteroid impacts and more gradual changes in sea level, climate and so forth (Vrba 1985; Brett & Baird 1995).

Much of this literature, and arguably a large majority of it, has depended on Jack Sepkoski's two monumental, global compendia of all marine animal age ranges, first of families (Sepkoski 1982) and later of genera (Sepkoski 2002). However, it has become apparent that these data are compromised by a variety of sampling biases (Raup 1976; Alroy *et al.* 2001, 2008; Peters & Foote 2001; Smith 2001). A steady progression of papers has emphasized this point (Foote 2003; Krug & Patzkowsky 2004; Crampton *et al.* 2006). Defenses of these compendia have focused on side-arguments, such as an attack (Bush *et al.* 2004) on sampling-standardization methods that fail to show a large exponential radiation after the Jurassic (Alroy *et al.* 2001); the logically insufficient claim that because an increase in local-scale richness matches the difference between the ends of Sepkoski's curves, those curves are likely to be correct (Bush & Bambach 2004); and the entirely predictable high quality of the Plio-Pleistocene bivalve record (Jablonski *et al.* 2003), a group whose exceptional preservation is unrepresentative of marine invertebrates in general (Foote & Sepkoski 1999).

The Paleobiology Database (<http://paleodb.org>) has more than tripled since an initial, preliminary analysis of two large Phanerozoic intervals (Alroy *et al.* 2001) suggested that corrections for sampling intensity bias might alter key patterns in Sepkoski's genus-level data. Without any corrections, the full height of Sepkoski's curve can now be replicated with reasonable corrections, a dramatically different and much flatter curve is indeed recovered (Alroy *et al.* 2008). This paper asks similar questions of a smaller model system: the North American record of Cenozoic mammals (Alroy 1996, 1998a,b, 1999, 2000, 2002; Alroy *et al.* 2000).

Three things make this system fruitful, and even necessary, to reanalyse. First, another decade of information has accumulated, including large amounts of data on reptiles that are used below to generate the time scale undergirding the analysis. Second, many of the statistical methods have improved greatly, as discussed below. Third, this system is much more tractable than Phanerozoic marine invertebrate data sets, because:

1. The high turnover rates of mammals compared to other groups (Stanley 1979) mean that the mammalian time scale (Alroy 1996, 2000) is about far more precise than traditional marine stages and epochs. The scale can be broken down into uniform and objectively defined intervals.
2. Identifications of fossil taxa within collections (i.e. occurrences) are taxonomically standardized, although unrecognized synonyms will always

- remain (Alroy 2002). Such a thing is almost unimaginable of the overall marine record, which involves at least 30 000 genera and far more species.
3. The data largely derive from the Western Interior of the United States and neighbouring parts of Mexico and Canada (Alroy 1998b), so very large-scale shifts in the geography of sampling do not exist.
  4. Mammals have reasonably constant taphonomic properties across time and taxonomic groups: a tooth is a tooth, but body parts such as trilobite cranidia, crinoid stem ossicles, and bivalve tests that are all thrown together in marine invertebrate studies have little in common. Taphonomy remains of great interest, especially because large and small mammals preserve and are collected in quite different ways. However, this issue is put aside for now, because large changes in the distribution of body masses (Alroy 1998a, 2000) make it difficult to address without going into much detail.
  5. Finally, the timing of global climate changes is well understood in the Cenozoic (Zachos *et al.* 2001), if not much farther back than the late Cretaceous, so climate proxy data can be compared directly to turnover patterns (Alroy *et al.* 2000).

The following sections present new and powerful methods for preparing the data, and then test a series of straightforward and general hypotheses. How many intervals witness markedly low or high turnover rates? Do rates fall through the Cenozoic? Are origination and extinction rates correlated? Are origination rates depressed or extinction rates inflated when diversity reaches high levels? Finally, does diversity vary within a narrow enough range to suggest an equilibrium?

### Time-scale analysis

There are two key steps in preparing the data. The first is to define a numerical time scale, as opposed to traditional categories such as land mammal ages, and the second is to produce a diversity curve.

The time-scale analysis uses information on all North American Cretaceous and Cenozoic turtles, crocodylians, dinosaurs, and mammals, not just mammals as in previous studies (Alroy 1996, 1998b, 2000). These taxa were selected because of their high preservation potential. The resulting data set includes 7642 fossil collections that document 39 743 occurrences of 1493 genera and 3694 species. The last published analysis (Alroy 2000) captured 4978 collections, 30 951 occurrences, 1241 genera, and 3243 species.

The quantitative biochronological method used here is appearance event ordination (Alroy 1992, 2000), which generates absolute estimates of the ages of fossil collections in millions of years ago (Mya). It shuffles first and last appearances until it maximizes a likelihood function involving hypothesized but as yet undemonstrated overlaps of age ranges (conjunctions). Conjunctions

are demonstrated by co-occurrence within collections or by mutual stratigraphic superposition. In contrast to earlier analyses that used graph theory to resolve geographic biases, the current one uses the simpler and apparently more accurate method of ignoring cases where disjunct pairs of taxa also do not overlap geographically for the purpose of computing likelihoods.

The resulting sequence of first and last appearance events is numbered from oldest to youngest by counting consecutive runs of first or last appearances. The counts are compared to geochronological age estimates to translate them into an absolute, numerical time scale. The position of each geochronological date in the event sequence is equated with that of the particular collection it dates. A collection's position in the event sequence is just the span between the oldest first appearance and youngest last appearance of any taxa in it, i.e. its assemblage zone. Line segments in the calibration plot are created by selecting 41 tie points from a larger set of 127 high precision  $^{40}\text{Ar}/^{39}\text{Ar}$ , uranium series, and paleomagnetic stratigraphy dates. After calibration, the event sequence is subdivided into uniform, as in previous studies (Alroy 2000). Collections only are used in the diversity analysis if their zones fall entirely within one bin.

A previous nonparametric method for selecting tie points called shrinkwrapping (Alroy 2000) is replaced here with a simpler set of winnowing criteria that is less sensitive to outliers: (1) remove 22 dates with assemblage zones entirely spanning those of others; (2) create averages from 31 dates that form 12 sets of identical assemblage zones; (3) use a simple greedy algorithm to remove 11 dates that unambiguously conflict with the position of other dates in the event sequence; (4) remove 10 dates immediately following younger geochronological estimates that the event sequences implies are older, but only if (a) the preceding date is a maximum, or (b) the removed date is a minimum; and (5) remove 23 points that create abrupt and temporary changes in the slope. The latter are identified by dividing the current slope by the preceding slope and the current slope by the following slope; multiplying the two ratios, which should each be 1.0 if there is no change; and seeing if this product is more than the current slope, or less than 1 divided by the current slope. An exception is made for the short segment spanning the rapid pulse of turnover at the Cretaceous–Tertiary boundary, which is well documented from independent evidence.

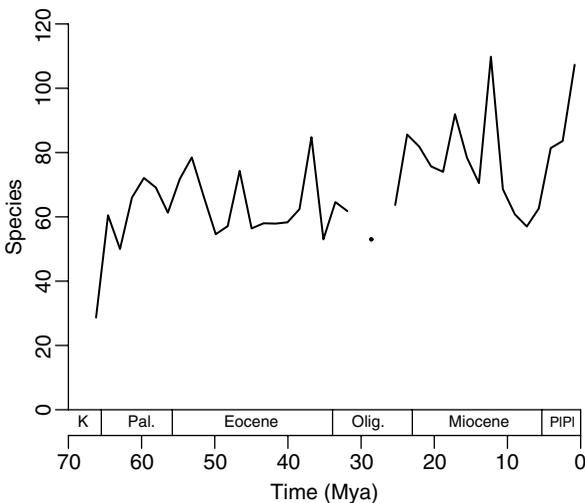
An alternative analysis employing all non-volant, terrestrial tetrapods resulted in a higher residual standard deviation (square root of the residual standard error) at the midpoint of each assemblage zone (1.34 Myr versus 1.06 Myr with the subset of four groups). Note that the error term computation forgives cases where the best interpolated age estimate is older than a minimum geochronological date or younger than a maximum, but not those where the date is a maximum or minimum but falls within an assemblage zone's range of age estimates, in which case the appropriate endpoint of the range is used to compute the residual.

The error figure is higher than before (Alroy 2000) because far fewer tie points are selected, so more of the dates fall off the calibration line. The difference might suggest that short sampling bins are no better than traditional land mammal ages. However, a simple nonparametric bootstrap simulation shows that short bins are robust. It draws from the observed ranges of age estimates for the calibration points and the observed offsets between these estimates and the geochronological dates. At a bin size of 0.1 Myr, just 37% of the few collections that have been assigned randomly to an exact bin have been assigned to the right one. This fraction rises swiftly to 64% at a bin size of 1.0 Myr and then slowly asymptotes, reaching 72, 77, and 84% at bin sizes of 1.5, 2.0, and 3.0 Myr. An entirely different, also nonparametric analysis with an earlier version of the data set showed much the same tradeoff of precision and accuracy (Alroy 1996).

Earlier studies (Alroy 1996, 1998b, 2000) used a bin size of 1.0 Myr, but doing so with this data set would create a noisy diversity curve with low autocorrelation, large changes between adjacent points, and low potential sampling quotas. These problems are essentially resolved by using a 1.635 Myr bin length that divides the Cenozoic into exactly 40 sampling intervals.

### Diversity estimation

The next task is to produce a diversity curve (Fig. 16.1) that is a proper statistical estimate based on comparably sized, randomly drawn subsets of the data, instead of a simple summation (Alroy 1996, 2000). Subsets only capture ecological dominant species that are abundant in samples, but no part of the fossil record is complete, so any count of actual species will reflect both richness and dominance. Methods of extrapolating instead of interpolating diversity, such as mark-recapture estimation (Connolly & Miller 2001), do seek to estimate overall species richness. However, they are not applied here because they



**Figure 16.1** Cenozoic trend in the standing diversity of mammalian species from western North America. Data are sampling standardized by randomly drawing enough entire fossil collections to reach a quota of 890 estimated individual specimens, with values averaged over 100 trials. The estimate for each collection is made by squaring the number of species it contains (Alroy 2000).

introduce noise while failing to remove the sampling overprint from this particular data set. Common species are ecologically important and biologically interesting, and it is better to have an accurate and precise measure of diversity in the general sense than an inaccurate and imprecise measure of species richness in the strict sense.

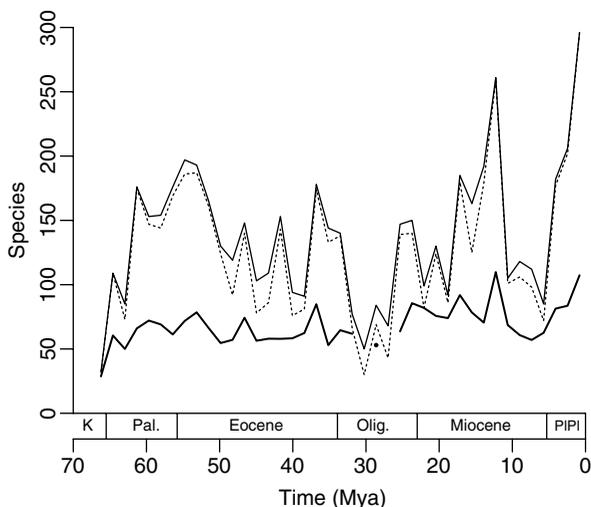
All of the steps involved in subsampling are improved upon here. They are drawing a uniform set of collections in each temporal interval (called by-list subsampling by Alroy 1996, 2000, and sample-based rarefaction by Gotelli & Colwell 2001); counting taxa from the subsamples; and estimating turnover rates, as discussed in the following section. Although these methods may seem difficult, the logic is sound, dramatically different results would be obtained by omitting any one step, and the estimates appear to be highly robust.

The argument has been made (Alroy 2000) that collections should be counted not one by one (Gotelli & Colwell 2001; Alroy *et al.* 2001, 2008; Allen *et al.* 2006), but by estimating the number of specimens in each, summing these estimates as collections are drawn, and stopping at a quota set in terms of specimens instead of collections. A good rule of thumb for fossil mammals is that the number of specimens is about equal to the square of the number of species occurrences within a collection (Alroy 2000). The simpler method of counting the occurrences with no transformation (Alroy 1996) is just a special case in which the relationship is assumed to be linear instead of exponential.

The problem with methods that use approximated specimen counts is underestimation of overall diversity when large collections predominate. In such cases, only a few collections representing a fraction of the landscape need be drawn to fill the quota. Thus, considerable beta diversity may go unsampled. The solution is to inversely weight the probability of drawing a collection by its estimated specimen count: if two collections have 5 and 10 taxa respectively, and therefore estimates of 25 and 100 specimens, it should be four times more probable to draw the smaller one. By doing this, sampling is dispersed spatially and environmentally, and on average each collection contributes exactly the same estimated number of specimens to each randomized draw.

Whether to estimate the specimen count for a collection by raising the number of taxa to a power 2 or, say, 1.4 has been debated (Bush *et al.* 2004). This problem turns out to be unimportant after dispersing the sampling by inverse weighting, because nearly identical curves are produced. The real issue is thus the spatial concentration of data, not the accuracy of the collection size estimates.

Until recently, most data sets such as Sepkoski's compendia (Sepkoski 1982, 2002) were simply lists of families or genera with geological first and last appearance dates. Such data only could be turned into diversity curves by assuming the presence of each taxon within every interval spanned by its age range. Thus, curves were based on range-through counts, which represent the minimum



**Figure 16.2** Alternative curves showing the diversity of North American mammals. Upper, thin line: range-through diversity, the sum of species actually sampled and species not sampled but inferred to exist because they are seen before and after a bin. Data are not standardized by subsampling. Dotted line: directly sampled diversity with no standardization. Thick line: sampled diversity after standardization (data used to construct Fig. 16.1).

numbers of taxa that must actually have existed in the bins. The effects of using these counts and not standardizing the data are great (Fig. 16.2; see also Alroy 1998b). For example, a very large apparent spike in the Pleistocene simply reflects the exceptional fossil record in that interval.

As sampling degrades, taxa found only in one interval (singletons) come to be more common in relative terms (Foote & Raup 1996). This problem was addressed previously by counting not the taxa within each interval, but those spanning the boundaries between intervals (Alroy 1996, 1998b, 2000). Recent simulation analyses to be presented elsewhere show that, in fact, these boundary-crosser counts are just as biased as range-through counts (Fig. 16.2): they fall at the edges of time series and both well before and after mass extinction events, and they increase as turnover rates decrease. They also rise sharply as the data approach a very thoroughly sampled interval, such as the Recent in many paleontological data sets (the Pull of the Recent: Raup 1979) combined with the Pleistocene in this one (Fig. 16.2).

Instead, simple counts of the number of taxa actually found within an interval's random subsample (sampled in bin or SIB diversity: Miller & Foote 1996; Alroy *et al.* 2001) turn out to be quite robust to all such biases, as long as sampling really is rendered uniform by drawing constant amounts of data, and the geographic, environmental, and taxonomic scope of sampling does not change much.

That said, SIB counts can be misleading because they scale to the sum of all taxa to have existed during each interval, so they may rise with rising turnover rates. Therefore, they may present a composite signal of diversity and turnover. For purposes of comparing diversity to the turnover rates, the interaction can be removed by obtaining an estimate of the average number of taxa existing at any

one time within a bin (standing diversity). Raup (1985, eqn. A29) showed that total diversity (the number tracked by SIB) can be approximated by multiplying a simple function of the origination and extinction rates ( $\lambda$  and  $\mu$ ) either by diversity at the base of a bin or diversity at the top of a bin. Therefore, standing diversity at either the bin bottom or bin top can be extracted by rearranging the expression.

Because separate standing diversity estimates are obtained from each turnover rate, their geometric mean is used to arrive at a final curve. These values are very congruent, with the median difference on a log scale being 0.020. Values from one calculation or the other are used in cases where only one estimate is available because of gaps in the subsampled data. The  $\lambda$  and  $\mu$  calculations respectively contribute two and one of these values. The raw SIB and standing diversity curves are nearly mirror images, with the median standing diversity estimate being 28% lower.

### Turnover rates

Turnover rates are notoriously difficult to estimate (Foote 1999, 2003), and numerous equations with different biases have been proposed (Foote 1994b). One robust method is to count the taxa at the base of a bin (i.e. the lower boundary crossers), see how many are present at the upper boundary, and compute an instantaneous extinction rate by taking the ratio of the logged values (Foote 1999; Alroy 2000). The origination rate is just the opposite: the log of the upper boundary crosser count divided by the lower boundary count of the same cohort. These exponential rates are directly equivalent to the ones computed in molecular lineage through time studies (Nee *et al.* 1992, 1994). The term 'origination' is used here because a relatively small number of appearances are of immigrants from Eurasia or South America.

For standing diversity counts, the rate can be computed analogously: the extinction rate is the absolute value of the logged standing diversity estimate for a bin  $i$  divided by the count of those same taxa sampled once again in the following bin  $i + 1$  (the two timers), and the origination rate is the mirror image (Alroy 2008). Taxa in the cohort must only be counted if they are actually sampled in bin  $i + 1$  because their later appearances in bins are subject to all the above counting method biases.

Such rates are biased in three ways that must be accounted for. First, the rates have a de facto upper bound imposed by the finite number of specimens drawn during subsampling. For example, if the quota is 100 specimens and two consecutive bins share 50 taxa, the upper bound on the origination rate is the absolute value of  $\log(100/50)=0.693$ , and the realistic limit is much lower, because many taxa will be sampled repeatedly. The bound, and therefore the bias, can be removed easily by dividing the counts of sampled and shared taxa by the sampling quota, logit-transforming these proportions,

and taking their difference. The difference of logits is just the log of the odds ratio, a common measure of effect size. With the current data the correction has little effect because the number of sampled species never closely approaches the number of sampled specimens.

Second, the rates are biased upwards by failure to resample the cohort. A reasonable correction involves a variant of gap analysis (Paul 1982), in which all taxa ranging across any given bin  $i$  are examined, and the proportion that is not sampled is computed. Such counts are inaccurate due to the above-mentioned problems with simple range data, such as edge effects. The solution is to only examine the group of taxa found in both bins  $i - 1$  and  $i + 1$ . The count of those not found (the part timers) is divided by that count plus the count of taxa found in all three intervals (the three timers). The probability of sampling a taxon  $s$  is therefore one minus this proportion (Alroy 2008). The corrected extinction rate can be shown with simple algebra to be approximately, although not exactly,  $\mu + \log p$ .

Small sample sizes may create high noise in separate estimates of  $s$  for each bin. Thus, the part and three timer counts are summed across all the data and put through this simple equation to produce an average  $s$  value (here 0.561) for use in correcting each bin's rates. Simulation analyses again not detailed here show that the estimates are unbiased regardless of edge effects and the like, as is desired.

Third, standing diversity figures in the turnover rate equations, but standing diversity itself is estimated from the rates, as mentioned in the previous section. Both things therefore must be generated recursively. First, standing diversity is estimated using the simplified expressions  $SIB/(1 + \lambda)$  or  $SIB/(1 + \mu)$ , as appropriate (Raup 1985, eqn. A28). These equations assume that  $\lambda$  and  $\mu$  are about equal. The above-mentioned ratio of the logit-transformed standing and two timer diversity values is then used to estimate  $\lambda$  or  $\mu$ , and the procedure is iterated repeatedly. In a second round, the full equation for SIB allowing  $\lambda$  and  $\mu$  to differ (Raup 1985, eqn. A29) is used instead. Here the estimate of  $\lambda$  at each iteration uses the median value of  $\mu$  determined in the first round, and vice versa.

## Data

The data employed in diversity analyses include all mammals except the volant order Chiroptera and four fully marine groups (Cetacea, Desmostylia, Pinnipedimorpha, and Sirenia). The geographic spread of sampling extends to the eastern USA during the late Neogene, most importantly with the addition of many fossil localities in southeastern Texas and Florida (Alroy 1998b). To remove the resulting biogeographic signal, the data are restricted to a rectangle spanning 20–60°N and 95–130°W, i.e. the entire western region of the Canada, the USA, and Mexico that yields substantial numbers of Paleogene in addition to Neogene vertebrates. This restriction leaves 24 690 out of 27 014 occurrences (91.4%), which include 3088 out of 3315 Cenozoic mammals

(93.2%). Several analyses also included a bin representing the terminal Cretaceous.

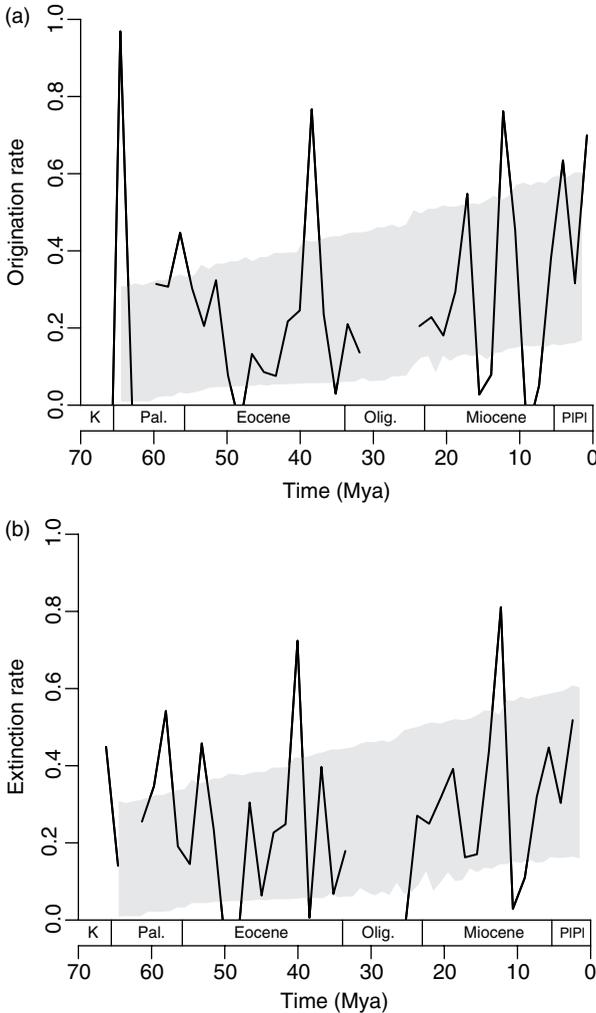
Two points within the Oligocene were removed because their estimated specimen counts would not meet any realistic sampling quota. The other bins all meet a quota of 890, which is generally high enough to sample 60–80 species, comparable to earlier studies (Alroy 1996, 1998b, 2000). Similar patterns are seen in a curve based on a doubled specimen quota. The geometric mean of raw subsampled diversity across bins in this larger data set is 90.6 instead of 68.3. However, the curves are visually almost identical, the rank-order correlation between them (Spearman's  $\rho$ ) is 0.966, and after logging both data sets the slope of a least-squares regression of the lower curve on the higher one is  $0.997 \pm 0.038$ . Thus, there is no significant change in shape as the quota is raised because the increase in diversity is essentially uniform.

### General patterns

The new sampling standardized diversity curve (Fig. 16.1) and turnover rates (Fig. 16.3) differ in some ways from the previous ones (Alroy 2000), most of which are not biologically important. These are: (1) the curve does not show the crash at the Cretaceous–Tertiary boundary because it represents diversity within bins instead of at bin boundaries, and the recovery is fast; (2) likewise, the end-Cretaceous extinction rate is high relative to the confidence intervals, but not very impressive, because almost all turnover actually was concentrated at the boundary but the equations assume continuous turnover through 1.635 Myr-long bins; (3) there is a much less pronounced diversity drop in the late Paleocene, an earlier pattern considered to be possibly artifactual (Alroy *et al.* 2000); (4) although this drop is followed by a small rebound, there is no large earliest Eocene origination spike and the Eocene is not consistently higher than the Paleocene; (5) there are consecutive extinction and origination peaks around 39 Ma that relate to poor sampling within the Duchesnean land mammal age; (6) there are brief diversity peaks in the late middle Miocene and Plio-Pleistocene that relate to intense sampling of small mammals, the former matched with concurrent origination and extinction rate excursions; and (7) there is a trough throughout the late Miocene, between about 11 and 5 Ma, instead of a short downwards excursion near the Mio-Pliocene boundary.

Despite these differences, the previous curve's most robust and biologically important key features (Alroy 2000) are still evident: (1) a weak overall trend; (2) a high earliest Paleocene origination rate; (3) a substantial offset between the Paleogene and Miocene; and (4) a full rebound from the late Miocene drop.

As noted before (Alroy *et al.* 2000), the most severe climate changes throughout the Cenozoic (Zachos *et al.* 2001) seem to have inconsistent effects, or perhaps none at all. A rapid global warming event at the Paleocene–Eocene boundary does trigger the invasion of North America by key groups such as crown-group



**Figure 16.3** Cenozoic origination (speciation plus immigration) and extinction rates for mammals from western North America, based on the same sampling standardized data shown in Figs. 16.1 and 16.2. Shaded areas show 95% confidence limits based a parametric bootstrap analysis employing 10 000 trials. (a) Instantaneous origination rates, based on the log ratio of estimated standing diversity and the number of these species also sampled in the previous bin (the two timers). (b) Instantaneous extinction rates computed using the log ratio of standing diversity and the following two timer count.

primates, artiodactyls, and perissodactyls (Clyde & Gingerich 1998; Smith *et al.* 2006), but the new turnover and diversity data show nothing dramatic at this scale of resolution. Likewise, there is no evidence for a response to the abrupt growth of ice sheets in the earliest Oligocene. Two missing data points in the data set (Fig. 16.1) make it hard to say whether the global late Oligocene warming event corresponds with a jump in the curve, although a match is certainly plausible. Finally, the late Miocene low comes during by a relatively minor global cooling trend (Zachos *et al.* 2001), and starts well before a major shift from  $C_3$  to  $C_4$  vegetation in the Western Interior that signals high seasonality (Fox & Koch 2004). Moreover, the possibility that this drop reflects poor sampling of small mammals cannot be dismissed, because it does not correspond clearly to changes in turnover rates (Fig. 16.3). As before (Alroy 2000), the certainly anthropogenic

end-Pleistocene megafaunal mass extinction (Alroy 1999) is not addressed, because an extinction rate for the last bin cannot be computed with the current methodology.

### Hypothesis tests

All reported correlations are nonparametric (Spearman's  $\rho$ ). The basic results are:

1. Cenozoic diversity has a trend (Fig. 1): there is a moderate correlation of subsampled diversity with time ( $\rho = -0.391$ ;  $p = 0.015$ ). The  $p$ -value shows nothing about possible mechanisms because the data are autocorrelated. It is useful simply to illustrate that there is indeed an upwards drift, justifying the use of an exponential trend as a null model in the following tests. A model assuming constant diversity would be easier to reject and therefore not as conservative.
2. The turnover rates (Fig. 16.3) show a weak upwards trend if any instead of declining, as found with less accurate methods (Alroy 2000). Regressions of origination and extinction rates against time are insignificant ( $\rho = -0.096, -0.183$ ;  $p = 0.584, 0.298$ ). The apparent decline was generated by high Paleocene rates, but here it makes no difference if the Paleocene points (the first six) are removed ( $\rho = -0.259, -0.329$ ;  $p = 0.166, 0.082$ ). Note that all tests reported here exclude large, paired outliers in the rates that are associated with an unusual burst of pseudoextinction between the second and third bins, which straddle the Puercan–Torrejonian boundary.
3. The rates are not predictable from one interval to the next. Each rate shows no serial correlation, i.e. a correlation with itself lagged by one interval ( $\rho = 0.136, -0.028$ ;  $p = 0.455, 0.880$ ). These tests are liberal because the data do trend very weakly.

From here on, a simple parametric bootstrap simulation iterated 10 000 times is used to compute significance levels. The null model is that the unseen, overall counts underlying the diversity and turnover rate estimates follow a long-term, exponential trend with no noise. The slopes and intercepts are estimated by linear regressions against time of SIB and two timer counts. The counts are divided by the average sampling probability  $s$  to obtain estimated figures for the entire species pool. In each interval, two timer and SIB subsamples are drawn at random with the same probability  $s$  to mimic the empirical data, and these values are used to generate standing diversity estimates and turnover rates. The results are:

1. The turnover rates show no great variability. Almost all the rates other than the few discussed above fall within or very near the confidence intervals generated by bootstrapping (Fig. 16.3). The standard deviations of the real

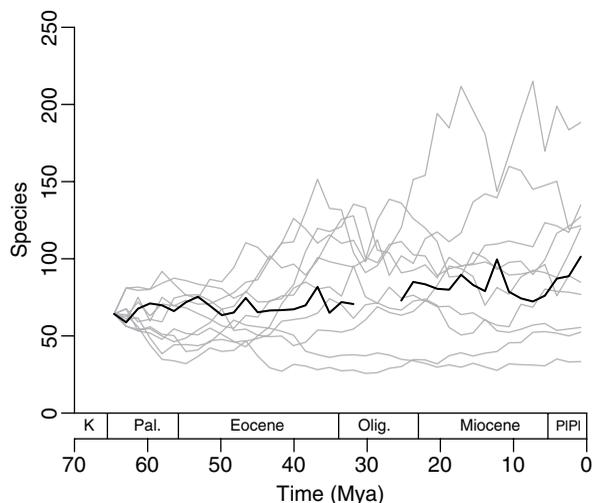
origination and extinction rates on a log scale are respectively 0.913 and 0.979, nearly equal to each other. The median origination and extinction rates are also subequal, being 0.228 and 0.249 species per species (spp/sp) per Myr. The extinction figure implies that the median species duration is 2.78 Myr. In combination with the preceding tests, these results indicate that the turnover rates are mostly white noise.

2. The rates seem not to correlate with each other. The bootstrapped 95% confidence limit of  $\rho$  is  $-0.241$  to  $0.493$ , and the observed value is  $0.199$ . Shifting the origination rates forward by one bin produces a seemingly great but insignificant cross-correlation of  $0.754$  (95% limits  $0.433$  to  $0.885$ ). The reason for the high correlations after lagging in both the real and simulated data is that in this case the rate estimate in each interval involves the same two timer count, which is small and therefore quite variable. High shared error is the price paid for removing such biases as edge effects.
3. Density dependence of either origination or extinction would produce logistic growth (Sepkoski 1978; Alroy 1996, 1998b; Foote 2000), and a diversity-origination relationship was found previously (Alroy 2000) with less robust data preparation methods. Here, however, there is no suggestion that origination rates fall with diversity or that extinction rates rise with diversity. Even after an outlier in the Paleocene, there are weak correlations ( $0.201$  and  $-0.008$ ) of standing diversity with origination and extinction during the next interval, and these correlations are insignificant based either on bootstrapping or on conventional tests. Here all of the time series are largely independent because either turnover rate in the next bin is computed from the diversity estimate in that bin, and in any case the simulation would build in any spurious correlation that might exist. Additionally, no detrending of the rates (Foote 2000) is necessary because the simulation includes the weak trend as an assumed parameter.

### Diversity curve simulation

In contrast to the turnover rates, the shape of the diversity curve itself provides strong evidence for logistic dynamics. The trend is much more flat than would be expected of a random walk with no density dependence, as shown by a simple, realistically parametrized simulation (Fig. 16.4) in the tradition of Raup (1977).

In this analysis, the starting count of species is the value predicted by the regression of log subsampled diversity against time. The count is multiplied at each step by a random number drawn from all the ratios of neighbouring points in the real diversity curve. The real curve's standard deviation is compared to the standard deviations of the simulated curves to test the hypothesis that density dependence flattens the trajectory. In a second test, the variance around a regression of log diversity against time is compared to the variance around similar



**Figure 16.4** Simulated diversity trends. Thick black line shows empirical data (Fig. 16.1); thin gray lines show ten representative random walks generated by drawing from the observed distribution of ratios between neighbouring diversity values.

regression lines fit to the random walks. All of the curves are logged before computing these figures.

Random counting error increases the real curve's variance, which makes the simulated curves more variable and therefore less likely to match the real one. The raw data curve is therefore dampened by (1) taking residuals to remove the temporal trend; (2) squaring the residuals to compute a variance measure; (3) separately estimating the variance of each data point that is created by sampling error as  $\sqrt{\text{SIB}}/\text{SIB}$  (Foote 1994a); (4) taking one minus the ratio of the sampling-related and observed variances; (5) rescaling the residuals from step 1 with this factor; (6) and adding back the trend to reconstitute the diversity curve. To make the test even more conservative, the ceiling on diversity imposed by the sampling quota of 890 estimated specimens is lifted by multiplying each diversity point by the ratio  $890 \text{ SIB}/(890 - \text{SIB})$ , which is analogous to the logit transformation of turnover rates discussed above. This correction is applied to the raw data immediately before the step that rescales the residuals.

Based on 10 000 trials, there is a 0.030 probability of generating a standard deviation less than the 0.126 log units seen in the rescaled data, with a median simulated standard deviation of 0.269. Thus, 53% of the expected variance is suppressed. Using the test that accounts for temporal trends, the probability of seeing such a smooth progression is 0.010.

## Discussion

### Turnover rates

As mentioned, the median extinction rate of 0.249 spp/sp/Myr implies a median expected duration of 2.78 Myr, which is substantially but not greatly more than the estimate of 2.14 produced by the last study of these data using very different

equations (Alroy 2000). It also is not very different from the rate of 1.7 Myr derived using yet another, even more divergent method, and an even older version of the same data set (Foote & Raup 1996).

However, the almost identical origination rate of 0.228 spp/sp/Myr is much higher than rates found in most molecular studies. Nee *et al.* (1992) estimated net speciation rates for all birds equivalent to 0.090 spp/sp/Myr. Magallon & Sanderson (2001) estimated a speciation rate of no more than 0.089 spp/sp/Myr for angiosperms over a wide range of possible extinction rates. Roelants *et al.* (2007) performed similar calculations for amphibians that produced a remarkably low maximal speciation rate of 0.0217 spp/sp/Myr. Of more direct relevance, Purvis *et al.* (1995) found negligible extinction rates and net diversification rates ranging from 0.070 – 0.342 spp/sp/Myr, but mostly on the low end of the range, in an analysis of all four major primate groups.

To some extent these differences are unsurprising, because mammals have long been recognized by paleontologists as evolving far more quickly than angiosperms and amphibians (Stanley 1979), although paleontological evidence is lacking for birds. However, the differences are large enough to suggest a real conflict.

One could argue that the molecular analyses are less biased because they work directly with dated branching events. However, the current analysis uses an almost unbiased method for computing origination rates, because edge effects are eliminated by working with counts of species that survive into and are sampled within an immediately following interval. The method also employs sampling-standardized data and takes failure to resample cohort members into account.

Nonetheless, there is still a legitimate reason to doubt that the rates derived in this study are comparable to true rates of cladogenesis and extinction. Specifically, perhaps much of the apparent turnover is due to pseudoextinction, as appears to be the case at the Puercan–Torrejonian boundary (see above). Because each pseudoextinction must lead to a pseudospeciation, this problem would tend not just to lower but to equalize calculated speciation and extinction rates. However, phylogenetic analyses of mammalian fossil clades suggest that pseudoextinction rates are normally around 10–25%. For example, Hulbert (1993) tallied 15 pseudoextinctions and 62 real extinctions in Neogene North American horses (19%), whereas Wang *et al.* (1999) inferred nine pseudoextinctions and 57 true extinctions (14%). Pseudoextinction rates that low would not create a large gap between paleontological and molecular results, or remove a large gap between speciation and extinction rates if any existed.

Although the current analyses are strictly at the species level, previously an ad hoc attempt was made to account for pseudoextinction. The correction involved identifying possible anagenetic lineages by finding cases in which two species of the same genus did not overlap in time and jointly spanned an interval that did

not otherwise include the genus (Alroy 1996, 2000). In other words, it estimated the absolute minimum possible diversity level. This correction is not used here, because it has a small effect; it does not employ real phylogenetic data; it works within the flawed range-through paradigm; and it assumes that genera are circumscribed in a similar way in all parts of the time scale.

A related objection is that the nominal turnover rates may largely reflect changes in preservability that cause lineages to enter and leave the pool of potentially sampled species. However, these evolutionary shifts are unlikely to be common, because all of the factors governing preservability are strongly heritable between species: local abundance, geographic range size, geographic range location relative to the area being sampled, taphonomic robustness of parts, ease of collection of parts, and taxonomic specificity of preserved parts.

### **Macroevolutionary theories**

The data are relevant to several important macroevolutionary hypotheses. First, Raup and Sepkoski (1982) showed that the average extinction rate across all marine animal families declined substantially through the Phanerozoic, and the same was later shown for genus-level origination rates (Sepkoski 1998; Foote 2000; Alroy 2008). There has been much interest in explaining these patterns with such mechanisms as simple natural selection (Raup & Sepkoski 1982) or sorting out of higher taxa (Sepkoski 1984; Gilinsky 1994), as well as interest in exploring the trends' implications for theories of diversity dynamics (Newman & Eble 1999). A similar pattern was seen earlier in the North American mammal data, and attributed to high volatility during the Paleocene (Alroy 1998b). The new data show either no real trend, or perhaps the opposite pattern.

Meanwhile, two established theories predict that turnover rates should be cross-correlated, one developed based on African Plio-Pleistocene mammals and the other on eastern North American early Paleozoic invertebrates. First, the 'turnover pulse' hypothesis (Vrba 1985, 1992) suggests that global climate cycles on the Milankovitch scale restrict and fragment the ranges of species that are poorly adapted to certain climate phases, driving bouts of speciation and extinction. Second, the 'coordinated stasis' hypothesis (Brett & Baird 1995) posits that entire regional biotas show little change not just in terms of taxonomic composition, but morphology and local-scale community composition. These patterns are maintained for millions of years, not within Milankovitch cycles, by strong species interactions and stabilizing selection. They are disrupted only by major environmental shifts. The observed lack of a cross-correlation in the current data, despite a long time series and presumably robust methods, casts real doubt on both theories.

It could be argued that the 1.635 Myr time scale used here is not relevant to the turnover pulse theory because multiple Milankovitch cycles will fall in each bin. However, there is no reason that the cycles should have exactly the same

cumulative effect in each bin, and the data set also spans several important changes in the relative intensity of the cycles (Zachos *et al.* 2001).

Indeed, there is nothing in the first place to motivate an argument that climate change has an important effect on a bin to bin basis, because most of the time the rates merely show random variation (Fig. 16.3; Alroy *et al.* 2000). The only clear exceptions are the extinction and origination pulses at either side of the Cretaceous–Tertiary boundary, because the few other excursions are all plausibly explained as sampling artifacts. Certain climate changes may still have caused changes in the biota's composition, if not in diversity per se. For example, there was an immigration pulse in the earliest Eocene that was coincident with a global warming event and does not register in the current data set. However, this dispersal episode was driven directly by biogeographic shifts and therefore only indirectly by climate (Clyde & Gingerich 1998; Smith *et al.* 2006). Together, the data suggest that routine climate cycles and changes in these cycles might be less important causes of extinction than biotic interactions such as competition, predation, and epidemic disease that are altered by adaptation and immigration.

Finally, the data do seem to confirm the Red Queen hypothesis (Van Valen 1973), which assumes that the constant evolution of predators, prey, and competitors prevents any species from establishing strong incumbency. Thus, extinction rates are predicted to show no trend through time, as in the current data set (Fig. 16.3), instead of declining, as in the marine data (Raup & Sepkoski 1982; Alroy 2008). Van Valen's theory does not directly address adaptive radiations and instead assumes a crowded world, so it is consistent with density dependence of rates as long as enough time has elapsed for diversity to rise and strong competition to come into play.

### **Diversity equilibrium**

The only clear overall pattern in the data, a flat diversity trajectory, could be explained by several dynamic mechanisms that are worth considering:

1. A secular decline in speciation or a secular increase in extinction. Gilinsky and Bambach (1987) found that origination rates typically decline within major marine animal groups, but extinction rates do not, even though they seem to decline when all groups are considered at once. This kind of a tradeoff between rates might explain the logistic growth patterns seen by Sepkoski (1979, 1984) in his Phanerozoic familial diversity data, even in the absence of density dependence. A symmetrical decline of both rates instead would allow exponential growth to continue.
2. Cross-correlations of rates. A positive correlation between origination and extinction rates could generate a diversity equilibrium regardless of density dependence, with extinction pulses being countered by radiations and vice versa (Webb 1969; Mark & Flessa 1977).

3. Density dependence of rates. If origination is a negative function of diversity, extinction is a positive function of diversity, or both rates show these properties, then an equilibrium will be reached at some point (Sepkoski 1978).

None of these hypotheses are directly supported by the analyses. However, models involving correlations are supported by marine data (Alroy 2008) and are more likely to be correct here because any strong trends through time would have been easy to spot. Furthermore, unless noise is a problem, the failure to recover any correlations might be a catch-22: the stronger an equilibrium, the less variation in turnover rates there is to correlate. Additionally, Foote (2000) argued that density dependence of both origination and extinction could simultaneously generate a cross-correlation between them and logistic growth. In other words, cross-correlations could be an effect and not a cause of equilibrium. Thus, direct density dependence seems like the most parsimonious interpretation.

Regardless of this dilemma, the evidence for a constrained diversity trajectory per se is fairly strong. The model used to test for a constraint (Fig. 16.4) accounts for biases that might create excessive variation in simulated diversity curves, and the true diversity curve (Fig. 16.1) is about as robust as could be hoped for. In contrast, evidence for density dependence is mixed in lineage through time studies: it is absent in primates (Purvis *et al.* 1995), but present in birds (Nee *et al.* 2002) and plausible in amphibians (Roelants *et al.* 2007). There are also serious technical difficulties with such analyses that might create the appearance of logistic growth, such as failure to discriminate newly evolved species (Purvis *et al.*, this volume).

If density dependence is real, we can make some inferences about the tempo of recovery from the current mass extinction (Alroy 2008). Given that the variance in diversity explained by saturation is 53% and that this saturation is caused by a relationship between the net diversification rate and standing diversity, it is easy to back-compute the slope of a regression between these variables. The slope and the respective means imply an intercept of 0.849 spp/sp/Myr, i.e. the rate when there is only one species. This value is more than three times higher than the background origination rate, much less normal net diversification rates. It means that a single species should give rise to 90% of the regional equilibrium number in about 18 Myr, and that the same level should be attained about 9 Myr after a 50% extinction pulse. A much faster recovery occurred after the Cretaceous–Tertiary boundary mass extinction (Fig. 16.1), but it seems safe to predict that any recovery from the current disruption will be fast on a geological time scale and extremely slow on a human time scale. Furthermore, a large radiation starting with a handful of species, some of them invasive and others closely related to each other, will do nothing to replace any phylogenetic, morphological, functional, trophic, or biogeographic diversity that will be lost.

The exact, population-level mechanism governing the statistically significant lack of variation through time in diversity cannot be pinpointed. Furthermore, one might wonder whether this pattern relates to real chances of speciation and extinction, or to chances that species will increase above or fall below the levels of abundance and geographical breadth required to be sampled in the fossil record. This difference is somewhat academic, and it stands to reason that the mechanism creating an equilibrium is primarily ecological, just as ecological factors such as abundance and reproductive rate govern variation among mammalian clades in speciation rates (Isaac *et al.* 2005). For example, competition for resources may cause average population sizes to fall as diversity increases, an inevitable outcome at some point. Small populations are more likely to go extinct, and must as population sizes limit on one, so an equilibrium eventually must result.

A more detailed mechanistic interpretation would be desirable, and recent work that predicts speciation rates based on explicit assumptions about population-level processes and their relationship to metabolism (Allen *et al.* 2006) shows great promise. However, this model would seem to imply a decline in origination rates because it posits that speciation rates scale to environmental temperature, and there was a strong overall decline through the Cenozoic in global temperatures (Zachos *et al.* 2001). There is no such decline in rates (Fig. 16.3).

A key assumption of this new theory is that although only species speciate, speciation rates need to be expressed in terms of per-individual, not per-species, rates. Moreover, the model predicts that not just per-individual but therefore per-species rates will be diversity-independent, if other factors such as total community abundance and temperature are constant. Thus, if any equilibrium exists, it must relate to extinction rates. However, because again species become extinct and individuals do not, extinction rates can vary only if species-level properties such as geographic range and population size do. This contradiction calls into question whether per-species rates are needless. The metabolic model (Allen *et al.* 2006) also does not address the mathematical certainty that constant total speciation rates will eventually produce more species than individuals, which could only lead to a diversity equilibrium.

Both fossil (Alroy 1998b) and molecular (Nee *et al.* 1992; Roelants *et al.* 2007) data show that the diversity of vertebrates does not change randomly, much less increase exponentially. This fact suggests that more process-oriented work needs to be done, although it need not be entirely reductionistic. Witnessing limits to diversity is just a matter of time: an infinite evolutionary process must produce them. Thus, the open issue is not whether limits exist, but rather whether they are approached quickly on a geological time scale (Sepkoski 1978, 1979, 1984; Benton 1995). In the case of North American mammals, the answer is yes.

### Acknowledgements

I thank Jon Bridle, Roger Butlin, and Dolph Schluter for the opportunity to contribute to this volume, and the British Ecological Society for allowing me to participate in the related Symposium. This work was conducted while a Center Associate at the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant #DEB-0553768), the University of California, Santa Barbara, and the State of California. I thank NSF's Division of Earth Sciences for recognizing the need for the Paleobiology Database to serve its discipline, and an anonymous donor for funding that effort. This is Paleobiology Database publication number 69.

### References

- Allen, A. P., Gillooly, J. F., Savage, V. M. and Brown, J. H. (2006) Kinetic effects of temperate on rates of genetic divergence and speciation. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 9130–9135.
- Alroy, J. (1992) Conjunction among taxonomic distributions and the Miocene mammalian biochronology of the Great Plains. *Paleobiology* **18**, 326–343.
- Alroy, J. (1996) Constant extinction, constrained diversification, and uncoordinated stasis in North American mammals. *Palaeogeography, Palaeoclimatology, Palaeoecology* **127**, 285–311.
- Alroy, J. (1998a) Cope's rule and the dynamics of body mass evolution in North American mammals. *Science* **280**, 731–734.
- Alroy, J. (1998b) Equilibrial diversity dynamics in North American mammals. In: *Biodiversity Dynamics: Turnover of Populations, Taxa, and Communities* (ed. M. L. McKinney and J. A. Drake), pp. 232–287. Columbia University Press, New York.
- Alroy, J. (1999) The fossil record of North American mammals: evidence for a Paleocene evolutionary radiation. *Systematic Biology* **48**, 107–118.
- Alroy, J. (2000) New methods for quantifying macroevolutionary patterns and processes. *Paleobiology* **26**, 707–733.
- Alroy, J. (2002) How many named species are valid? *Proceedings of the National Academy of Sciences of the United States of America* **99**, 3706–3711.
- Alroy, J. (2008) Dynamics of origination and extinction in the marine fossil record. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 11536–11542.
- Alroy, J., Aberhan, M., Bottjer, D. J., *et al.* (2008) Phanerozoic trends in the global diversity of marine invertebrates. *Science* **321** 97–100.
- Alroy, J., Koch, P. L. and Zacos, J. C. (2000) Global climate change and North American mammalian evolution. Deep time: paleobiology's perspective (ed. D. H. Erwin and S. L. Wing), pp. 259–288. *Paleobiology* **26** (supplement).
- Alroy, J., Marshall, C. R., Bambach, R. K., *et al.* (2001) Effects of sampling standardization on estimates of Phanerozoic marine diversification. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 6261–6266.
- Benton, M. J. (1995) Diversification and extinction in the history of life. *Science* **268**, 52–58.
- Brett, C. E. and Baird, G. C. (1995) Coordinated stasis and evolutionary ecology of Silurian to Middle Devonian faunas in the Appalachian Basin. In: *New Approaches to Speciation in the Fossil Record* (ed. D. H. Erwin and R. L. Anstey), pp. 285–315. Columbia University Press, New York.

- Bush, A. M. and Bambach, R. K. (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., Markey, M. J. and Marshall, C. R. (2004) Removing bias from diversity curves: the effects of spatially organized biodiversity on sampling standardization. *Paleobiology* **30**, 666–686.
- Clyde, W. C. and Gingerich, P. D. (1998) Mammalian community response to the latest Paleocene thermal maximum; an isotaphonomic study in the northern Bighorn Basin, Wyoming. *Geology* **26**, 1011–1014.
- Connolly, S. R. and Miller, A. I. (2001) Joint estimation of sampling and turnover rates from fossil databases: capture-mark-recapture methods revisited. *Paleobiology* **27**, 751–767.
- Crampton, J. S., Foote, M., Beu, A. G., *et al.* (2006) The ark was full! Constant to declining Cenozoic shallow marine biodiversity on an isolated midlatitude continent. *Paleobiology* **32**, 509–532.
- Foote, M. (1993) Discordance and concordance between morphological and taxonomic diversity. *Paleobiology* **19**, 185–204.
- Foote, M. (1994a) Morphological disparity in Ordovician-Devonian crinoids and the early saturation of morphological space. *Paleobiology* **20**, 320–344.
- Foote, M. (1994b) Temporal variation in extinction risk and temporal scaling of extinction metrics. *Paleobiology* **20**, 424–444.
- Foote, M. (1999) Evolutionary radiations in Paleozoic and post-Paleozoic crinoids. *Paleobiology Memoir* **25**, 1–115.
- Foote, M. (2000) Origination and extinction components of taxonomic diversity: Paleozoic and post-Paleozoic dynamics. *Paleobiology* **26**, 578–605.
- Foote, M. (2003) Origination and extinction through the Phanerozoic: a new approach. *Journal of Geology* **111**, 125–148.
- Foote, M. and Raup, D. M. (1996) Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology* **22**, 121–140.
- Foote, M. and Sepkoski, J. J., Jr. (1999) Absolute measures of the completeness of the fossil record. *Nature* **398**, 415–417.
- Fox, D. L. and Koch, P. L. (2004) Carbon and oxygen isotopic variability in Neogene paleosol carbonates: constraints on the evolution of the C-4-grasslands of the Great Plains, USA. *Paleogeography, Palaeoclimatology, Palaeoecology* **207**, 305–329.
- Gilinsky, N. L. (1994) Volatility and the Phanerozoic decline of background extinction intensity. *Paleobiology* **13**, 427–445.
- Gilinsky, N. L. and Bambach, R. K. (1987) Asymmetrical patterns of origination and extinction in higher taxa. *Paleobiology* **13**, 427–445.
- Gotelli, N. J. and Colwell, R. K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* **4**, 379–391.
- Hulbert, R. C., Jr. (1993) Taxonomic evolution in North American Neogene horses (subfamily Equinae): the rise and fall of an adaptive radiation. *Paleobiology* **8**, 159–167.
- Isaac, N. J. B., Jones, K. E., Gittleman, J. L. and Purvis, A. (2005) Correlates of species richness in mammals: body size, life history, and ecology. *American Naturalist* **165**, 600–607.
- Jablonski, D., Roy, K., Valentine, J. W., Price, R. M. and Anderson, P. S. (2003) The impact of the Pull of the Recent on the history of bivalve diversity. *Science* **300**, 1133–1135.
- Kirchner, J. W. and Weil, A. (2000) Correlations through time in fossil extinctions and originations. *Proceedings of the Royal Society of London, Series B*, **267**, 1301–1309.
- 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., Kiessling, W., Aberhan, M., *et al.* (2006) Ecological, taxonomic, and taphonomic components of the post-Paleozoic increase in sample-level species diversity of marine benthos. *Paleobiology* **32**, 533–561.
- Krug, A. Z. and Patzkowsky, M. E. (2004) Rapid recovery from the Late Ordovician mass extinction. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 17605–17610.
- Magallon, S. and Sanderson, M. J. (2001) Absolute diversification rates in angiosperm clades. *Evolution* **55**, 1762–1780.
- Mark, G. A. and Flessa, K. W. (1977) A test for evolutionary equilibria: Phanerozoic brachiopods and Cenozoic mammals. *Paleobiology* **3**, 17–22.
- Miller, A. I. and Foote, M. (1996) Calibrating the Ordovician radiation of marine life: implications for Phanerozoic diversity trends. *Paleobiology* **22**, 304–309.
- Nee, S., Holmes, E. C., May, R. M. and Harvey, P. H. (1994) Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions of the Royal Society Series B* **344**, 77–82.
- Nee, S., Mooers, A. O. and Harvey, P. H. (1992) Tempo and mode of evolution revealed from molecular phylogenies. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 8322–8326.
- Newman, M. E. J. and Eble, G. J. (1999) Decline in extinction rates and scale invariance in the fossil record. *Paleobiology* **25**, 434–439.
- Olszewski, T. D. (2004) A unified mathematical framework for the measurement of richness and evenness within and among multiple communities. *Oikos* **104**, 377–387.
- Paul, C. R. C. (1982) The adequacy of the fossil record. In: *Problems of Phylogenetic Reconstruction* (ed. K. A. Joysey and A. E. Friday), pp. 75–117. Academic Press, London.
- Peters, S. E. (2004) Evenness of Cambrian–Ordovician benthic marine communities in North America. *Paleobiology* **30**, 325–346.
- Peters, S. E. and Foote, M. (2001) Biodiversity in the Phanerozoic: a reinterpretation. *Paleobiology* **27**, 583–601.
- Powell, M. G. and Kowalewski, M. (2002) Increase in evenness and sampled alpha diversity through the Phanerozoic: comparison of early Paleozoic and Cenozoic marine fossil assemblages. *Geology* **30**, 331–334.
- Purvis, A., Nee, P. and Harvey, P. H. (1995) Macroevolutionary inferences from primate phylogeny. *Philosophical Transactions of the Royal Society Series B* **260**, 329–333.
- Raup, D. M. (1976) Species diversity in the Phanerozoic: an interpretation. *Paleobiology* **2**, 289–297.
- Raup, D. M. (1977) Stochastic models in evolutionary paleontology. In: *Patterns of Evolution as Illustrated by the Fossil Record* (ed. A. Hallam), pp. 59–78. Elsevier Scientific, Amsterdam.
- Raup, D. M. (1979) Biases in the fossil record of species and genera. *Bulletin of the Carnegie Museum of Natural History* **13**, 85–91.
- Raup, D. M. (1985) Mathematical models of cladogenesis. *Paleobiology* **11**, 42–52.
- Raup, D. M. and Sepkoski, J. J., Jr. (1982) Mass extinctions in the marine fossil record. *Science* **215**, 1501–1503.
- Roelants, K., Gower, D. J., Wilkinson, M., *et al.* (2007) Global patterns of diversification in the history of modern amphibians. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 887–892.
- Sepkoski, J. J., Jr. (1978) A kinetic model of Phanerozoic taxonomic diversity I. Analysis of marine orders. *Paleobiology* **4**, 223–251.
- Sepkoski, J. J., Jr. (1979) A kinetic model of Phanerozoic taxonomic diversity II. Early Paleozoic families and multiple equilibria. *Paleobiology* **5**, 222–252.

- Sepkoski, J. J., Jr. (1982) A compendium of fossil marine families. *Milwaukee Public Museum Contributions in Biology and Geology* **51**, 1–125.
- Sepkoski, J. J., Jr. (1984) A kinetic model of Phanerozoic taxonomic diversity III. Post-Paleozoic families and mass extinctions. *Paleobiology* **10**, 246–267.
- Sepkoski, J. J., Jr. (1997) Presidential address: Biodiversity: past, present, and future. *Journal of Paleontology* **71**, 533–539.
- Sepkoski, J. J., Jr. (1998) Rates of speciation in the fossil record. *Philosophical Transactions of the Royal Society Series B* **353**, 315–326.
- Sepkoski, J. J., Jr. (2002) A compendium of fossil marine animal genera. *Bulletins of American Paleontology* **363**, 1–560.
- Simpson, G. G. (1944) *Tempo and Mode in Evolution*. Columbia University Press, New York.
- Smith, A. B. (2001) Large-scale heterogeneity of the fossil record: implications for Phanerozoic biodiversity studies. *Philosophical Transactions of the Royal Society Series B* **356**, 351–367.
- Smith, T., Rose, K. D. and Gingerich, P. D. (2006) Rapid Asia-Europe-North America geographic dispersal of earliest Eocene primate *Teilhardina* during the Paleocene-Eocene Thermal Maximum. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 11223–11227.
- Stanley, S. M. (1979) *Macroevolution: Pattern and Process*. W.H. Freeman, San Francisco.
- Van Valen, L. (1973) A new evolutionary law. *Evolutionary Theory* **1**, 1–30.
- Vrba, E. S. (1985) Environment and evolution: alternative causes of the temporal distribution of evolutionary events. *South African Journal of Science* **81**, 229–236.
- Vrba, E. S. (1992) Mammals as a key to evolutionary theory. *Journal of Mammalogy* **73**, 1–28.
- Wagner, P. J. and Erwin, D. H. (1995) Phylogenetic tests of speciation hypotheses. In: *New Approaches to Speciation in the Fossil Record* (ed. D. H. Erwin and R. L. Anstey), pp. 87–122. Columbia University Press, New York.
- Wagner, P. J., Kosnik, M. A. and Lidgard, S. (2006) Abundance distributions imply elevated complexity of post-Paleozoic marine ecosystems. *Science* **314**, 1289–1292.
- Wang, X., Tedford, R. H. and Taylor, B. E. (1999) Phylogenetic systematics of the Borophaginae (Carnivora, Canidae). *Bulletin of the American Museum of Natural History* **243**, 1–659.
- Webb, S. D. (1969) Extinction-origination equilibria in late Cenozoic land mammals of North America. *Evolution* **23**, 688–702.
- Zachos, J., Pagani, M., Sloan, L., Thomas, E. and Billups, K. (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693.