

DIVERSITY ESTIMATES, BIASES, AND HISTORIOGRAPHIC EFFECTS: RESOLVING CETACEAN DIVERSITY IN THE TERTIARY

Mark D. Uhen and Nicholas D. Pyenson

ABSTRACT

The number of cetacean genera varied greatly over their evolutionary history. Cetaceans originated in the Eocene and then rapidly diversified, reaching around 30 genera during the Middle Eocene. The number of genera crashed in the late Eocene, with only eight described from the early Oligocene. This time also represents the putative extinction of archaeocetes and the origin of Neoceti, the clade including extant cetaceans. By the late Oligocene, the number of genera recovered and had begun an expansion that peaked in the middle Miocene, at over 75 genera. Since that time, the number of genera has decreased to around 40 genera extant today.

Ideally, generic counts reflect the actual number of genera in the past (generic richness), but our understanding of past diversity is confounded by many factors. Generic counts are potentially too high because of the prevalence of genera based on non-diagnostic material and taxonomic over-splitting. Taxonomic counts can also appear to be too high if the counts are made in time intervals that are excessively long. Conversely, generic counts are potentially too low because of fossil non-preservation; non-deposition, destruction, or obscured fossil-bearing, continental shelf rocks; and lack of collecting or publication effort. We evaluate each of these factors as a cause of bias in cetacean diversity. Individual causes may have significant effects on both true generic richness and the amount of bias. These factors must be addressed in any evaluations of past diversity, especially for clades represented by an abundance of fossils.

Mark D. Uhen. National Museum of Natural History, Department of Paleobiology, Washington DC, USA
uhenm@si.edu

Nicholas D. Pyenson. Museum of Paleontology and Department of Integrative Biology, University of California, Berkeley, California, USA. pyenson@berkeley.edu

KEY WORDS: diversity, generic; bias; collections; Cetacea; Sirenia; Tertiary

INTRODUCTION

Identifying accurate measures of diversity is a task that preoccupies paleontologists and neontol-

ogists alike. Across extinct and extant studies, biological diversity may equate only in terminology but not in biological actuality (i.e., species, alpha diversity, generic richness). The interplay between

PE Article Number: 10.2.10A

Copyright: Society of Vertebrate Paleontology August 2007

Submission: 21 January 2007. Acceptance: 12 July 2007

Uhen, Mark D., and Pyenson, Nicholas D., 2007. Diversity estimates, biases, and historiographic effects: resolving cetacean diversity in the Tertiary. *Palaeontologia Electronica* Vol. 10, Issue 2; 11A:22p, 754KB;
http://palaeo-electronica.org/paleo/2007_2/00123/index.html

these deep time and recent diversity studies, however, has generated a wealth of literature in the past 30 years (Jablonski 1999) and has spurred a profusion of ideas about diversity over time that carry both macroecological and conservational implications (Jackson and Johnson 2001; Jackson et al. 2001). At a fundamental level, the value of any measure of diversity, for paleontologists, directly relies on how well we think our measures of diversity separate biological signal from noise (Raup 1976b).

Traditionally, paleontologists use taxonomic counts, i.e., a tally of the number of fossil taxa (often families or genera) from a given time, as quantitative measurement of biological diversity across geologic time (Sepkoski 1997). Ideally, generic counts reflect the actual number of genera per time interval, but many biases can distort measured diversity from the actual diversity of any time interval (Raup 1976a; Sepkoski 1997). Moreover, for any given taxon, there are specific and, possibly, a limited number of factors that bias measures of diversity through geologic time. These factors arise directly from organismal behavior, life history, taphonomy, and morphology, and vary widely from group to group. Some studies in the paleontological literature use generic compilations within a higher taxon to reflect actual evolutionary trends in a broad fashion, despite apparent biases and distortional issues (Benton et al. 2000; Foote and Sepkoski 1999), whereas other studies seek to develop statistical techniques that account for various biases at broad temporal scales (i.e., 10^6 yr; Miller and Foote 2003; Peters and Foote 2001)

Our study of generic diversity through geologic time differs from previous work in both scale and taxonomic focus: we use cetaceans (a specific clade with a distinct and known temporal and spatial origin) as an exemplar taxon to assess specific issues associated with generic diversity estimates at a scale smaller than that of Phanerozoic marine diversity (Jablonski and Sepkoski 1996). By using genera as the primary units, we aim to circumvent problems associated with large-scale analytical and taxic paleobiological studies (eg., Adrain and Westrop 2000; Sepkoski 1978). Because our study examines generic richness within a single mammalian order through time, we can attain an acceptable and qualitative level of consistency with taxonomy and nomenclature. Although some Linnean ranks may not have biological equivalence in actuality (Mishler 1999), given the current practice of descriptive paleontological research, generic counts provide reliable data at the highest practical

taxonomic resolution for analyzing changing diversity through geologic time.

Cetacea as a Case Study

Cetacea originated in the late early Eocene as mostly terrestrial mammals with a few characteristics for aquatic predation and sensation (Gingerich and Uhen 1998). By the late Eocene, basilosaurid archaeocetes had become obligate marine cetaceans, and their fossils from both hemispheres suggest that they achieved a cosmopolitan distribution in the world's oceans (Uhen 1998). After the radiation of Eocene archaeocetes, cetacean evolutionary history can be simply summarized by the origin of Neoceti, in the Oligocene and their diversification during the middle Miocene (Fordyce and de Muizon 2001). Notwithstanding this widely accepted view of cetacean history, the potential effects on apparent cetacean diversity (including potentially biasing factors) have been identified in broad terms (Barnes et al. 1985; Fordyce 2003b and others) but these effects, particularly the potentially biasing factors, have remained largely unaddressed in specifics and in categorization. We address these potential sources of bias herein.

Potential Diversity Biases and Solutions

Increases bias.

One source of increased diversity bias may arise from naming new taxa based on limited, non-diagnostic fossils. Potentially, different parts of the same cetacean can be given different names. In paleontology, this process, called taxonomic inflation, results in multiple names being applied to what is actually a single taxon. Taxonomic inflation appears to be most acute in taxa with a sparse and fragmentary fossil record. Also, taxonomic inflation can arise from a literal reading of the taxonomic literature without some form of taxon vetting, which can conflate diversity estimates with taxonomic practice (Alroy 2002; White 2003). In this study, we attenuate the effect of potential over-splitting by strictly counting only genera instead of species.

Taxonomic inflation could still be a problem if genera being counted are based on species named with non-diagnostic type specimens. To prevent further burdening cetacean literature with an abundance of taxa named from fragmentary material (e.g., isolated periotics and teeth), Barnes (1977) proposed to limit new species descriptions only to specimens with diagnostic cranial, periotic and post-cranial elements. Despite this wise advice (even if it had been followed since the time it was offered), many non-diagnostic cetacean genera

have persisted in the literature. In this study, genera represented by type species with poor type specimens were designated *nomina dubia* and left out of the taxonomic counts. It is important to note that for some taxa, describing new taxa represented solely by complete specimens may bias diversity in the opposite manner (Donovan 2001). Incomplete, but diagnostic type specimens are common in vertebrate paleontology, and thus, they are very important to include in this kind of analysis.

Another potential source of apparent diversity is a lack of comparison of the type specimens leading to an overcount of the number of named taxa. Counting genera instead of species should help overcome this problem, but it may persist even on the generic level. The inability to directly compare type specimens of different species is particularly problematic for Cetacea because of the large size of many specimens. Logistically, it is difficult to compare cetacean type specimens side by side when the specimens are as large as a researcher, and especially so when key specimens are on different continents. Current research practices, such as distributing casts of critical type specimens, publishing informative figures, and reproducing descriptive monographs using modern electronic media are all helping to make cetacean specimens more comparable.

Another source of increased diversity bias may result from counting taxa in time intervals that are excessively long. As Foote (2000) noted, the varying duration of different time intervals can distort diversity counts significantly, particularly if some time intervals vary greatly in length. This problem can be minimized by using the finest time scale possible, as well as using time intervals that are evenly divided. Ideally, the length of each time interval would be less than the duration of the taxa being counted, and taxa would occur in multiple time intervals. This type of idealized pattern would allow researchers to use taxa that are boundary crossers for counts of diversity rather than those that are confined to the time interval alone, a procedure recommended by Foote (2000) to avoid problems of variations in interval length. Unfortunately, many cetacean genera are based on single occurrences, and most cetacean genera are confined to a single time interval, so the use of boundary crossers would be inappropriate for measuring cetacean diversity.

Decrease bias

Generic diversity counts may be lowered by lack of available fossil-bearing rock. We assessed this factor by measuring geologic map area to see if available rock correlated with generic diversity (see Crampton et al. 2003 for a recent application of this approach). Similarly, the presence of fossils and the lack of collection effort may decrease generic counts. We thus examined the uniformity of collecting effort (quantified by publication data) in time and space (Raup 1977; Sheehan 1977). In parallel with collecting effort, non-preservation and/or destruction of fossils may similarly decrease diversity counts, and we thus also assessed this biasing factor by determining the uniformity of preservation biases in time and space. Finally, informative paleontological data are hard won: the stepwise progress of information from field discovery to publication (summarized in White and Folkens 2000) can stop at any point in the sequence of field, museum, laboratory, or publication preparation. The presence of unpublished specimens can become “phantoms” of diversity in an analogous manner to ghost lineages in phylogenetic analyses. Our study also includes assessments of unstudied and undescribed cetacean genera to determine if study effort is uniform in time and space (see Jablonski et al. 2003).

DATA SOURCES AND DATA PROCESSING

Data on the geographic and temporal distribution of cetacean genera were tabulated from the primary original literature and from subsequent reviews of the original literature. The bibliographic data, as well as the taxonomic and occurrence data in this literature, were entered into the Paleobiology Database (PBDB). This database is publicly available and can be accessed at www.pbdb.org. Great effort has been taken to include every publication that includes taxonomic or distributional data on any fossil cetacean taxon through the June 2007. Certainly some small number of fossil cetacean publications still remain to be entered into the PBDB, but these should be only randomly associated with any particular taxa, time interval, or geographical region, which means that their absence should not bias our results in any predictable manner.

This data compilation resulted in 248 cetacean genera currently in use (218 of which have a known fossil record) and 188 additional generic names that have been previously applied to cetacean fossils. Of these 188, 74 are now junior syn-

onyms of other cetacean genera; seven were improperly formed genera that were subsequently replaced; three are objective synonyms of other genera; 77 of them are *nomina dubia* (names of doubtful application), *nomina nuda* (names without proper justification), *nomina oblita* (unused names), or *nomina vana* (names for which the proper application of which cannot be determined); and 26 are published misspellings of other genera. One genus was moved to Pinnipedia based on a new understanding of the type specimen (*Ontocetus emmonsii*), leaving a subsequently named cetacean species without a genus ("*Ontocetus oxymycterus*"). These data were compiled from 761 separate scholarly sources.

A similar, but perhaps slightly less comprehensive, data collection approach was also used for the mammalian order Sirenia, for use as a taphonomic control taxon. A slightly larger proportion of sirenian references probably still remain to be entered into the PBDB, but as stated above, these should be only randomly associated with any particular taxa, time interval, or geographical region, which means that their absence should not bias our results in any predictable manner.

This data compilation resulted in 40 sirenian genera currently in use and 33 generic names that have been previously applied to sirenian fossils. Of these 33, 30 are now junior synonyms of other sirenian genera; three were improperly formed genera (a homonym and two *nomina nuda*) that were subsequently replaced. These data were compiled from 301 separate scholarly sources.

Data from five museum specimen databases were compiled to help answer questions relating to the numbers of specimens, collections, and fossil collectors, and the habits of these collectors. These factors can be important biases that affect studies of diversity through time, especially for vertebrate paleontologists (e.g., Davis and Pyenson 2007). All records relating to the Order Cetacea were gleaned from the databases at the Florida Museum of Natural History (FLMNH), Natural History Museum of Los Angeles County (LACM), San Diego Museum of Natural History (SDMNH), University of California Museum of Paleontology (UCMP), and the United States National Museum (USNM). These museum databases were chosen because they all have very large collections of fossil cetaceans, and these museum collections together include fossil collections from all of the major fossil cetacean-producing areas of North America. The 48 contiguous United States were considered to be roughly equivalent to the conti-

nent of North America for several reasons. First, no fossil cetaceans are known from Hawaii, and only poorly identified material is known from two collections in Alaska. Second, of the fossil cetacean material in Canada, the most productive locality is from a very small area of southern Quebec, which has produced only two taxa (*Delphinapterus* and *Phocoena*) of Pleistocene age. The only other fossil cetacean localities from Canada are limited to the type locality of *Chonecetus* and two other collections producing indeterminate cetaceans from the Oligocene of British Columbia. Mexico includes six collections that have produced cetacean fossils identified to the generic level, resulting in eight genera. The effort to digitize the geologic maps of Canada and Mexico was considered excessive compared to the potential payoff. Also, keeping this part of the analysis to only a single country also helps to standardize the collection effort, since it is more likely to be closer to uniform within a single country.

Geologic maps of those 48 contiguous United States that contain continental shelf strata from the Eocene to Recent were digitized in order to determine the map area of rocks that could potentially produce fossil cetaceans. Any states with digital versions of geologic maps were unmodified from their original format. The latter include: Delaware (Delaware Geological Survey 1976), Florida (Scott et al. 2001), Maryland (Maryland Geological Survey 1968). For the remaining states, paper maps were digitally photographed in small sections. These include: Alabama (Osborne et al. 1989), California (Jennings 1977), Georgia (Pickering and Murray 1976), Louisiana (Snead and McCulloh 1984), Mississippi (Bicker 1969), North Carolina (North Carolina Geological Survey 1985), Oregon (Earnest et al. 1991), South Carolina (US Geological Survey 1936), Texas (Hartmann and Scranton 1992), Virginia (Johnson 1993) and Washington (Caruthers et al. 2002; Walsh et al. 1987). These sections, including the map scale, were photographed at exactly the same scale with a digital camera mounted on a tripod. These map sections were then reassembled in Adobe Photoshop to form a complete map. See Figure 1 for a graphical outline of the measurement procedure described here. Map areas of all Eocene to Recent formations were digitally measured using ImageJ software (Figure 1). These area data were recorded at the finest stratigraphic scale possible, based on the stratigraphic resolution of the map, and then assigned geologic ages based on the most recent publications available.

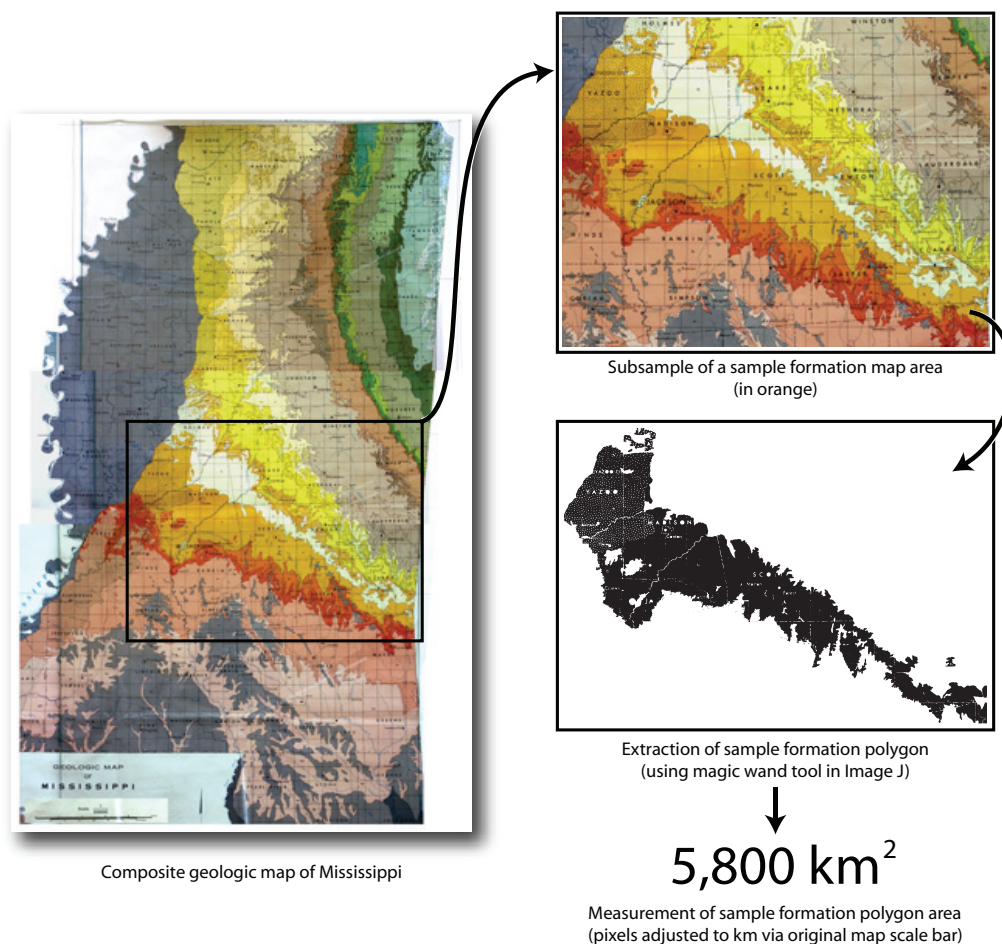


Figure 1. Graphic representation of the method used to measure map area of rocks from geologic maps. In this example, the undifferentiated Jackson Group (Eocene) of Mississippi is highlighted for measurement. The black and white bitmapped image of the Jackson Group is exported to ImageJ for measurement, based on the scale of the original map.

Population data were also tabulated to determine if the number of people in a potentially fossil cetacean-bearing area had any effect on the number of fossil cetaceans found in that area. Data on the population of each county that included potentially fossil cetacean-bearing strata in the states mentioned above were tabulated from the U.S. Census Bureau (available at <http://www.census.gov/population/www/censusdata/cen-counts.html>). Data from the censuses of 1900 to 1990 were included in this analysis.

We assessed the research effort by tabulating the number of papers published on fossil cetaceans (those including the keyword *Cetacea*) from the GeoRef database (www.georef.org, American Geological Institute). Thus, we use publication output as a proxy for research effort. GeoRef may not be the ideal database to use for this type of tabulation because it does not include many museum

publications, nor does it include many biologically oriented publications where many paleontologists regularly publish their work. The Bibliography of Fossil Vertebrates (BFV, www.bfvol.org/) may seem like a better choice, but it contains large gaps, covering only the years 1509 to 1968 and 1981 to 1993. Although GeoRef may be problematic because it may miss some obscure publications, it is less problematic for this type of analysis than the BFV and GeoRef has conceivably not changed the focus of its databasing effort over calendrical time.

To measure publication effort, we counted the number of papers on fossil cetaceans that referred to each time interval in our cetacean data set using the name of the time interval as a keyword. For instance, we tabulated a paper as “late Eocene *Cetacea*” if both keywords “late Eocene” and “*Cetacea*” are in the keyword field. Table 1 lists the

Table 1. Summary of the searches of papers in GeoRef that include the keyword Cetacea

Search Criteria	Number of references in GeoRef	Number of references in PBDB
Cetacea AND Holocene	67	N/A
Cetacea AND ((late Pleistocene) OR (upper Pleistocene))	19	6
Cetacea AND ((middle Pleistocene)	0	2
Cetacea AND ((early Pleistocene) OR (lower Pleistocene))	4	2
Cetacea AND ((late Pliocene) OR (upper Pliocene))	8	25
Cetacea AND ((early Pliocene) OR (lower Pliocene))	26	33
Cetacea AND ((late Miocene) OR (upper Miocene))	61	49
Cetacea AND middle Miocene	75	87
Cetacea AND ((early Miocene) OR (lower Miocene))	40	46
Cetacea AND ((late Oligocene) OR (upper Oligocene))	55	25
Cetacea AND ((early Oligocene) OR (lower Oligocene))	11	8
Cetacea AND ((late Eocene) OR (upper Eocene))	59	27
Cetacea AND middle Eocene	59	33
Cetacea AND ((early Eocene) OR (lower Eocene))	21	2
Cetacea AND Pleistocene NOT (((late Pleistocene) OR (upper Pleistocene)) OR (middle Pleistocene) OR ((early Pleistocene) OR (lower Pleistocene)))	25	N/A
Cetacea AND Pliocene NOT (((late Pliocene) OR (upper Pliocene)) OR (early Pliocene) OR (lower Pliocene))	70	N/A
Cetacea AND Miocene NOT (((late Miocene) OR (upper Miocene)) OR (middle Miocene) OR ((early Miocene) OR (lower Miocene)))	134	N/A
KW=Cetacea AND Oligocene NOT (((late Oligocene) OR (upper Oligocene)) OR (early Oligocene) OR (lower Oligocene))	76	N/A
Cetacea AND Eocene NOT (((late Eocene) OR (upper Eocene)) OR (middle Eocene) OR (early Eocene) OR (lower Eocene))	86	N/A

set of searches used to generate this data set. This procedure was complicated by the observation that GeoRef includes both time terms such as early, middle, and late, as well as time-stratigraphic terms such as lower, middle, and upper. We included papers that included either the time terms or time-stratigraphic terms. We avoided double counting of papers that included both types of terms by combining search criteria with the "OR" operator. The papers in GeoRef that include only a more general time term such as Eocene, without a modifier of any kind, presented another potentially confounding factor. These papers were tabulated by some additional searches for references that included the more general time terms exclusive of the more particular terms. The results of these tabulations made in December 2006 are shown in Table 1.

We also tabulated the number of references that contribute to the total data set in the Paleobiology Database, as well as the number of references that contribute to the generic counts and occurrence counts in each time interval. The results of these tabulations are also shown in Table 1. The number of genera described per paper is, of course, variable, but as long as there is no temporal trend (that is, a trend over the geologic times from which these fossils are described) that variability should not affect our results.

Several previous authors have also suggested excluding singletons (genera that are known from a single time interval) from this type of analysis to minimize Lagerstätten and monographic effects (Lu et al. 2006). This approach was not practical here, since many vertebrate genera, and fossil cetacean genera in particular, are known from single specimens or single collections. (See

Discussion for an expansion of this point.) Moreover, we suspect that excluding singletons would not matter much because: a) there are few cases of cetacean Lagerstätten (e.g., Sharktooth Hill bonebed); b) such cetacean Lagerstätten exhibit spectacular preservation and do not impact diversity (e.g., Pisco Formation); and c) few cetacean publications that have produced many generic names are still in use.

METHODS OF ANALYSIS AND METHODS TO ASSESS POTENTIAL BIASES

Several plots of cetacean generic richness from the Eocene to Recent were created from our generic data set. The first is a bar chart showing the number of cetacean genera described from each Cenozoic subepoch for the entire world (Figure 2.1). The second is a bar chart showing the number of cetacean genera described from each Cenozoic subepoch only for North America (Figure 2.2). Although we would have preferred to assess the potential biases in this data set for the entire world, compiling some of the necessary data (particularly the map area of rocks and population of various areas over time) would have been prohibitive. A total of 240 cetacean genera were tabulated with 101 of them represented in North America (about 42%). This figure is a higher proportion of the total than any other single continent, except Europe with 109; North America also includes all of the stages from which fossil cetaceans are known with the exception of the early Eocene (Ypresian).

Unless otherwise noted, plots of generic diversity are shown with taxa “ranged through.” That is, if a taxon is reported before and after but not in a particular time interval, it is shown as existing in the time interval. For instance, if a taxon were reported from the fossil record in the Miocene and Pleistocene, but not reported in the Pliocene, it would be tabulated for the Pliocene. This method forces us to assume generic continuity and monophyly, which are also assumptions made by the taxonomists who assigned these particular specimens to these genera in the first place. In some instances (i.e., when looking at taphonomic bias for instance), the plots are made of genera as sampled in the time interval, rather than ranged through, so they are only tallied for a particular time interval if they actually occur in that time interval.

Taxonomic Bias

Ideally, we would tabulate the number of fossil cetacean species over time rather than the number of genera. Unfortunately, many species of fossil

cetaceans have been named on very limited fossil material. Often these fossils are not diagnostic at even the family level, much less the genus or species level. Much of this problem can be avoided by simply tabulating genera rather than species. This step presumes that the number of genera is similarly correlated with the number of species at all times in the geologic past.

Using genera instead of species does not completely eliminate the potential of poorly delimited taxa artificially inflating taxonomic counts. To further combat this potential bias, we eliminated from our tabulation genera that were based on type species with non-diagnostic type material. This determination was made either by following the opinions of previous authors [(Kellogg 1968) for Cope’s mysticetes; (Fordyce and de Muizon 2001) for most odontocetes], or by reading the original description of the type material and making the determination ourselves. These taxonomic opinions were added to a set of unpublished opinions to the PBDB for all cetacean genera (entered as Mark’s opinions, Uhen 2008: the future date ensures that these opinions will be the most recent in the database and override all other opinions). In most instances, these opinions followed the most recent expert opinion for the genus in question. In a few cases however, taxa were designated *nomina dubia* because of the poor quality of the type material for the genus, a new designation for these taxa. This procedure resulted in 34 genera being designated as *nomina dubia* that had not been so designated by the most recent expert opinion, although many of these taxa had been judged to be non-diagnostic by some author or authors at some point in the past. These taxa are listed in the Appendix.

Time Scale Bias

As noted above, and by Uhen (1996), the time scale used in analysis of diversity can have a significant effect on the shape of the diversity curve and on the metrics calculated from the diversity data (Foote 2000). To determine the effect of time scale coarseness on our analysis of cetacean diversity, we plotted our data at every time scale available in the PBDB. Time scales that did not reflect the patterns found in the finest time scale were rejected from future study (see Figure 3).

Additionally, we emphasize how it is also important to note how well fossil collections can be placed temporally. If a fossil collection can only be placed to the nearest epoch, it is not very helpful when analyzing diversity at the stage level. To

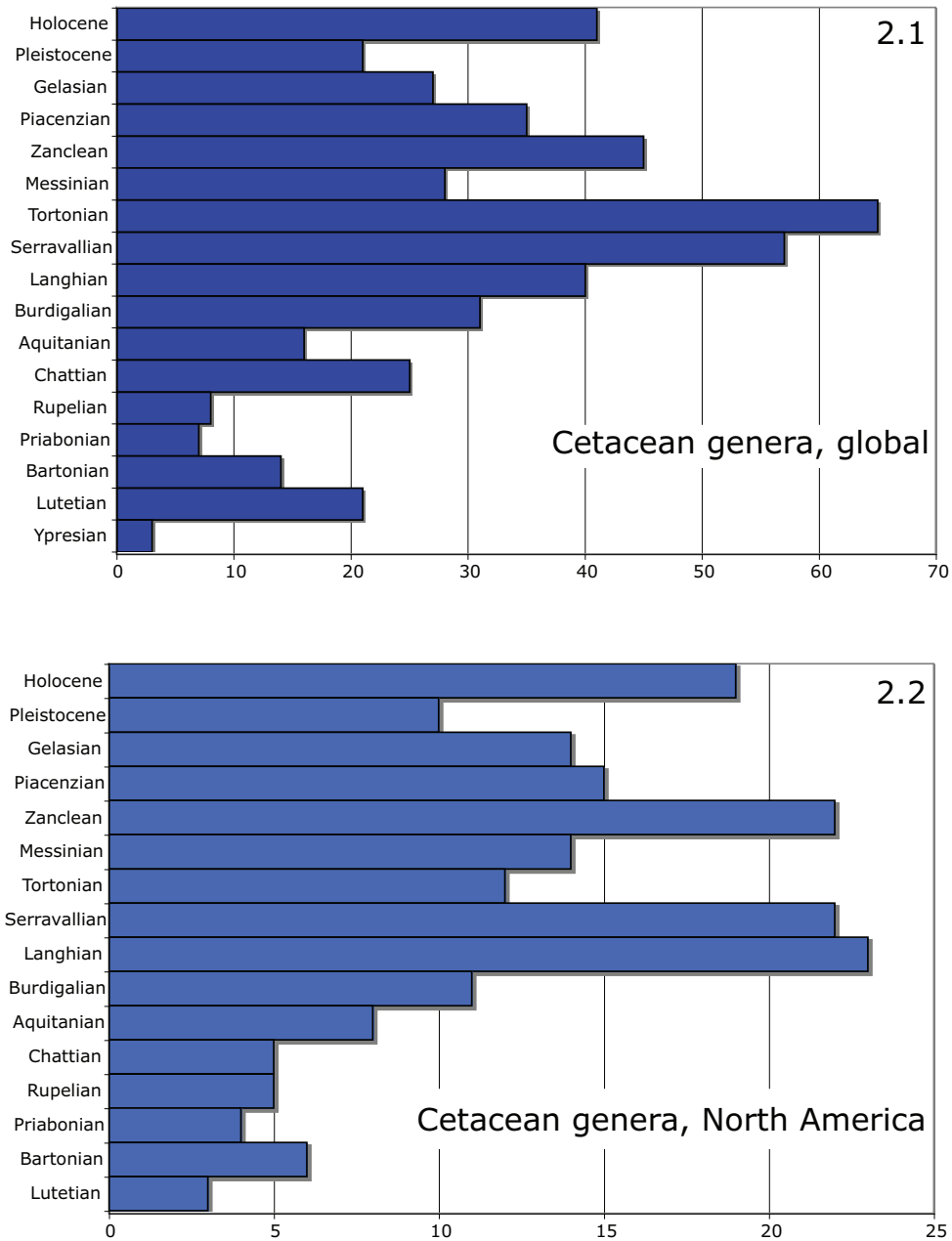


Figure 2. Cetacean generic richness. **2.1.** Bar chart showing the number of cetacean genera described from each time interval for the entire world (dark blue). **2.2.** Bar chart showing the number of cetacean genera described from each time interval for North America (sky blue). Note the similarity in the patterns of diversity, despite the different scales. Although the general shape of these distributions is similar, the first differences (the differences between one time interval and the next) are not correlated.

accommodate this potential problem, the PBDB discards fossil collections that cannot be allocated to a single time interval when generating a diversity curve. Clarifying the geologic age of fossil collections that were temporally poorly placed when originally described was a serious objective in this

study, which we accomplished by searching the geological literature for additional information on the age of deposits via paleomagnetic dating, radiometric dating, and biostratigraphic (particularly microfossil) dating. In the final global data set, 378 of 953 fossil cetacean collections with fossils iden-

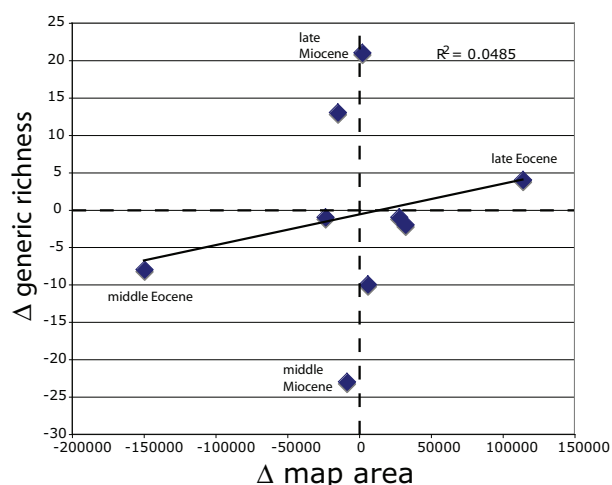


Figure 3. Subepoch-to-subepoch changes in the map area of continental shelf rocks for each time interval in North America and sampled North American cetacean diversity (not ranged through). The correlation between the two factors is poor. Spread along the rock area axis is mainly driven by the huge map area of middle Eocene rocks. Spread along the cetacean diversity axis is mainly driven by the huge diversity spike in the middle Miocene.

tified to genus (about 40%) were correlated with global ages. In North America, we correlated global ages for 180 of 349 (about 51%) fossil cetacean collections with fossils identified to genus.

Collection Bias

The process of collecting fossils can significantly affect taxon counts resulting from those collections. Prolific collectors often work in particular geographic areas and also within narrow stratigraphic (and thus chronologic) bands. To determine the extent of this potential bias, we tabulated the numbers of specimens that could be attributed to particular collectors in each of the museum databases that we examined. If a collector was listed individually, or as a member of a party, the specimens collected by the party were attributed to the individual collector only if that collector was listed first or was noted as the primary collector.

Available Rock Bias

Though comprehensive, our method of measuring map area of continental shelf rocks is very time consuming. It is also dependent on good maps being available at similar scales for all areas of analysis. For these reasons, we limited our study of map area to North America. Our analyses of the map areas compared North American continental

shelf rock only to the diagnostic cetacean genera of North America. Because the amount of rock available for collecting of fossils has been shown to affect fossil diversity counts (Crampton et al. 2003; Raup 1976a; Smith 2001), our measure of the geologic map area of potentially cetacean-producing rocks represents an attempt to determine if the same is true for cetaceans. Map area is less than ideal for this purpose for several reasons. First, geologic map area does not reflect the area that is visible on the ground. Many such deposits are covered with soil, vegetation, and/or human development, making them relatively inaccessible for fossil collecting. Second, geologic map area can underestimate the rock available for fossil collecting when the deposits are relatively flat lying and exposed in nearly vertical surfaces. Even if geologic map area inaccurately measures the potential area in which to find fossils, it was the only tractable measure of this factor available. Formation name data from the GEOLEX database were also explored to measure bias inherent in the rock record (Peters and Foote 2001), but these data were only recorded to the epoch level, whereas much of the variation in cetacean diversity is apparent only at finer time scales.

Population Bias

We assessed human population size in fossil cetacean-bearing areas to determine if the number of people living in these areas had any effect on the number of fossils found. If we assume that the chance of discovering fossils is essentially random for any person who is not necessarily looking for them, and that many fossil cetaceans are found by these individuals, population size could have a large impact on the number of fossils found. If, however, most fossil cetaceans are found by people who are intent on finding them, this situation is less likely to be the case. To resolve this issue, data from California, Florida, North Carolina, and Oregon were explored in detail. The number of fossil cetacean specimens in each state was tabulated by decade in which they were found to compare to the number of people living in the state in the same decades. If fossil cetaceans are found by random individuals, then more people in the state should show an increase in the number of fossil cetaceans found. If, on the other hand, fossils are found by collectors intent on finding them, more fossils should be found in the decades during which these collectors were most active.

Research Effort Bias

Research effort was assessed by plotting the number of papers in the GeoRef database on fossil cetaceans by geologic age to determine if this correlated with the number of cetacean genera known in those ages. Sheehan (1977) performed a similar test by tabulating the time intervals of interest of paleontologists as listed in the *Directory of Palaeontologists of the World*. One of his concerns about using these data were that some paleontologists listed multiple times of interest, and he was forced to split their "interest units" equally among their listed time intervals, even though they may not have split their efforts evenly. The use of GeoRef keyword data avoids this problem because it actually measures research output, not stated areas of interest. Whereas this measure is closer to the research effort than the area of interest noted by Sheehan, it is still not the same. Individual researchers may invest a great deal of effort into projects that either produce no published results or a relatively small number of published works.

We also assessed research effort by plotting the number of references in the PBDB that contributed to the cetacean record. These records are automatically associated with time and taxonomic units as these data are entered into the PBDB.

Taphonomic Bias

Taphonomic biases can greatly affect the distribution of fossils and could potentially alter our perceptions of generic richness (Behrensmeyer et al. 2000). Cetaceans, as obligate aquatic mammals at least since the late Eocene, have very different taphonomic processes affecting them as compared with most other mammals. To control for potential taphonomic biases, we compared the generic richness of cetaceans with that of sirenians over the same time interval. There is an inherent potential taphonomic problem early in cetacean history because the earliest cetaceans were somewhat terrestrial to semi-aquatic (Thewissen et al. 1996). Thus, different taphonomic processes were affecting semi-aquatic early cetaceans when compared to fully aquatic late Eocene basilosaurids and all later cetaceans. Fortunately, sirenians underwent the same terrestrial to aquatic transition at the same time and in the same geographic regions as cetaceans (Domning 2001), so our taphonomic control taxon should be affected by the same taphonomic processes as our taxon of interest (Behrensmeyer et al. 2000).

RESULTS

Taxonomic Bias

Our tabulation of described cetacean genera totaled 248 genera that had not been explicitly synonymized with other diagnosable genera. After combing the literature to determine which of these genera are based on diagnostic material and which are not, we found that of the 436 generic names that have been applied to fossil cetaceans, only 56% of them are still in use. Of the 43% that are no longer in use, 6% are simple misspellings, 18% are considered *nomina dubia*, *nomina vana*, or *nomina oblita* (mainly based on poor type material), and 19% are synonyms of other taxa.

Figure 4 shows a comparison of the global cetacean fossil record with junior synonyms (both subjective and objective) included (unvetted) and excluded (vetted). The two distributions are virtually identical indicating that the specimens identified as junior synonyms of other genera were not more likely to live in any time interval more than any other. This result also shows that, in hindsight, this taxonomic housekeeping step was not necessary if all we were interested in was the relative change in diversity from time interval to time interval. This step has been completed, however, and we can now be confident in negating the bias from bad taxonomic practice at the generic level. This result is similar to results found by Wagner et al. for Paleozoic gastropods, Jurassic pelecypods, and Cenozoic pelecypods (Wagner et al. 2007). Furthermore, we can also report absolute values for generic richness rather than just relative changes in generic richness.

Time Scale Bias

As noted above, and by Uhen (1996), the time scale used in analysis of diversity can have a significant effect on the shape of the diversity curve. Figure 5 shows the effect time scale coarseness on perceptions of taxon duration. Note that the mean, maximum, and range of the perceived taxon durations all increase as the scale becomes coarser. The epoch scale is somewhat different from all of the others. The result shows most taxa with a long range because those taxa are found in the Miocene, which is the longest of all Cenozoic epochs, and because several late Miocene genera range into the early Pliocene, making their perceived ranges even longer.

Foote (2000) recommended using only boundary crossing taxa in diversity analyses to avoid the problems associated with interval length.

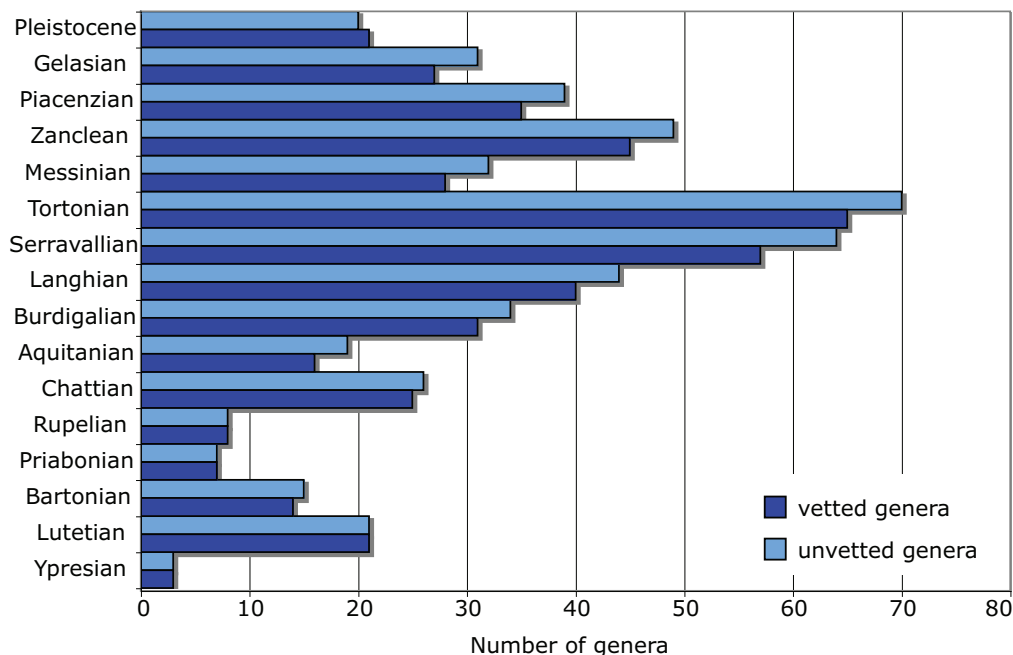


Figure 4. Bar chart showing the number of cetacean genera that have diagnostic type material described from each time interval for the entire world plus junior synonyms (unvetted genera, light blue), compared to the number of cetacean genera that have diagnostic type material described from each time interval for the entire world (vetted genera, dark blue). Note that both distributions are virtually identical. Correlation of the first differences displays a significant correlation ($r^2 = 0.988$).

In this study, such an approach proved problematic for cetaceans because many of the genera are known only from single time intervals. Of the 218 cetacean genera with a known fossil record, 103 of them are known from a single subepoch. See the Discussion below for a further development of this point.

Collection Bias

Collector habits.

Analysis of the museum collection data from the FLMNH, SDMNH, UCMP, and USNM shows that in each museum, a small number of collectors contribute a majority of the specimens to each collection (see Table 2). Many of these collectors are

also avocational paleontologists who have a particular geographic interest, and thus they have collected within a narrow stratigraphic range. For example, 99% of the specimens attributed to the most prolific collector at the USNM (P.J. Harmatuk) are from a single locality, the Lee Creek phosphate mine near Aurora, North Carolina.

This type of analysis emphasizes a particular point about the nature of fossil collecting localities and how they influence our ideas about diversity. Authors in the past have noted fossil Lagerstätten can cause a spike in diversity against the background of normal fossil deposits (Behrensmeier et al. 2000). In this case, the Lee Creek Mine has produced a great number of specimens (6585 speci-

Table 2. Summary of data on fossil cetacean collectors from the FLMNH, SDMNH, UCMP, and USNM.

Museum	Number of Collectors with > 1% of the collection	Percent of collection in top collectors	Percent of top collectors who are avocationalists
FLMNH	23	78%	52%
SDMNH	12	83%	67%
UCMP	13	63%	38%
USNM	11	80%	55%

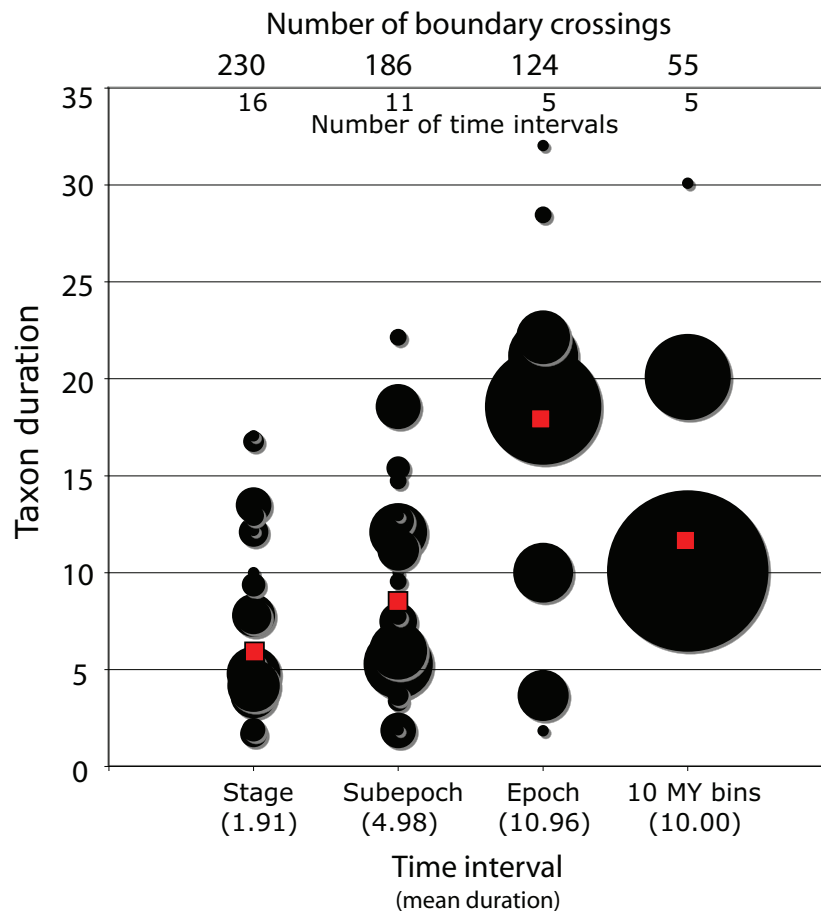


Figure 5. Bubble chart showing the effects of interval size on the perception of taxon longevity. The bubbles on the chart are centered on the duration of cetacean genera, and the size of the bubbles is proportional to the number of taxa having that duration. The red squares show the mean duration for each time scale. The numbers across the top indicate the number of boundary crossings from one time interval to the next. The mean duration of the time intervals is also shown across the bottom. Note that at the finest scales, the mean durations are not dramatically different. The epoch scale shows an interesting elevation in taxon duration, which is caused by a large number of genera that cross from the late Miocene into the early Pliocene. This pattern is not shown in the 10 million year bin scale because this group of taxa is broken up in that scale. Also note that the number of boundary crossings decreases as the number of time intervals decreases, as one might expect.

mens or 75% of the USNM specimen database), but few have been identified to genus. For the Lee Creek Mine, this situation may be resolved in a forthcoming volume on the marine mammal fossil of this site, with seven new fossil whale species to be named, including two new genera (Whitmore and Barnes, in press; Whitmore and Kaltenbach, in press).

Although the Lee Creek Mine greatly increased the number of specimens of fossil cetaceans, the site only marginally increased our knowledge of the diversity of cetaceans during the time of deposition of the Lee Creek assemblage. We attribute this paucity of diversity to the taphonomic setting of Lee Creek, which produces mostly

fragmentary and isolated bones and teeth, and thus only a very small proportion of the specimens are generically or specifically diagnostic.

Collecting localities.

There are 953 collecting localities (called “collections” in the PBDB) that have produced cetaceans listed in the PBDB. The first differences (the difference between one time interval and the subsequent time interval) of sampled cetacean diversity and the number of collections are shown in Figure 6. The two factors are positively correlated, albeit very weakly, which indicates that the number of collections has an effect on perceived diversity,

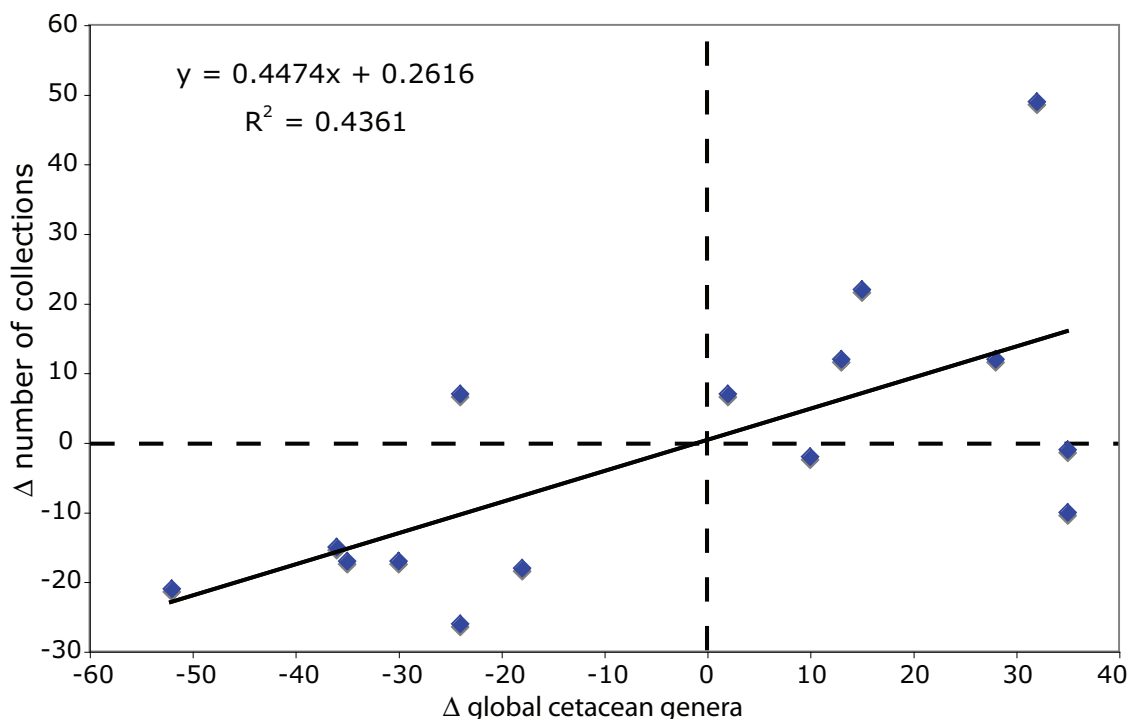


Figure 6. Subepoch-to-subepoch changes in the number of collections with diagnostic cetacean genera for each time interval in North America and sampled diversity (not ranged through). There is only a weak correlation between the two factors, indicating that the pattern of cetacean diversity is driven by more than just the distribution of cetacean collections in geologic time.

but it is by no means the only factor controlling its value.

Available Rock Bias

Unlike previous studies of Phanerozoic diversity (e.g., Crampton et al. 2003; Raup 1976a), the number of cetacean genera over time does not appear to be significantly linked to map area of rocks in North America (see Figure 3). This result is characterized by an abundance of middle Eocene continental shelf rocks, particularly in Texas, and these rocks have produced very little cetacean material. We find these results not completely surprising because cetaceans only evolved in the late Early Eocene in Asia (Bajpai and Gingerich, 1998) and only arrived in North America after crossing the Atlantic Ocean during the middle Eocene (Uhen, 1999). Even ignoring the middle Eocene however, there appears little correspondence between map area and the number of genera. Despite the large number of genera in the middle Miocene, middle Miocene deposits are not overrepresented when compared to the other time intervals. Also, despite the relatively large number of genera in the early Pliocene, there is very little in

the way of early Pliocene continental shelf deposits. The one instance of correspondence between the two data sets that makes intuitive sense is the complete lack of cetaceans from the early Oligocene and a relatively small map area of early Oligocene rocks. Fordyce (2003a) has suggested that the lack of Rupelian cetacean fossils is probably attributable to widespread erosion and reworking of shallow marine deposits caused by an eustatic sea-level fall, which is marked by a widespread unconformity at 29-30 Ma.

Population Bias

No discernable relationship exists between the number of people in a cetacean-bearing state and the number of fossils found. Table 3 shows the ratio of the number of fossils found to the number of people in California, Florida, North Carolina, and Oregon over the last century. The ratio fluctuates wildly with noticeable peaks that are often attributable to particular collectors. The aforementioned dominance of particular collectors in museum collections explains these data rather succinctly and intuitively.

Table 3. Summary of population data in relation to the number of fossil cetaceans specimens collected per decade. Data for this table was compiled from the FLMNH, UCMP, LACM, SDMNH, and USNM specimen databases. Entries in each cell represent the number of fossil specimens found divided by the population (multiplied by 100,000 just for presentation purposes). Note that there is no discernable relationship between population and the number of fossils found, which would be indicated by a steady value or steadily changing value in each column. In fact, spikes (shown in bold) in the values are caused by spikes in the number of fossils collected and can be attributed to individual collectors. For California, it is attributable to L.G. Barnes; for North Carolina, it is attributable to P.J. Harmatuk; and for Oregon, it is attributable to D. Emlong.

Year	California	Florida	North Carolina	Oregon
1990	0.25	3.43	5.38	0.87
1980	0.95	0.88	46.60	17.70
1970	3.46	1.72	25.80	40.19
1960	0.41	0.28	0.08	
1950	0.32	0.15	0.09	0.58
1940	0.81	0.28	0.21	0.87
1930	0.49	0.42	0.11	1.04
1920	0.26	3.04	0.13	1.87
1910	0.05	0.71	0.15	1.50
1900	0.09	0.20	0.18	2.16

Research Effort Bias

Research effort is well correlated with the number of diagnostic cetacean genera. First differences of the number of papers in GeoRef is tightly correlated with the number of genera of fossil cetaceans (Figure 7, $r^2 = 0.546$, $p < 0.01$). This result matches the findings of Sheehan (1977) for the readily fossilizable marine shelf invertebrates over the entire Phanerozoic, despite the fact that Sheehan estimated effort in a completely different way. Also, the first differences of PBDB references is even better correlated with first differences in cetacean genera in the PBDB (Figure 7.2, $r^2 = 0.751$, $p < 0.001$). The significance of research effort bias suggests that the number of diagnostic cetacean genera is directly related to the amount of research output relating to fossil cetaceans.

Taphonomic Bias

Figure 8 shows the absolute pattern of sirenian diversity over time (8.1) and correlation of the first differences of sirenian genera and cetacean genera (8.2). These two distributions are weakly, but significantly correlated ($r^2 = 0.460$, $p < 0.01$), indicating that a common factor is affecting both cetacean and sirenian diversity. The total number of sirenian genera (31) remains much lower than the total number of cetacean genera (240). Interestingly, sirenian generic richness peaks early on,

declining toward more recent time. Because neither group is completely absent from the fossil record at any time during their history, this comparison strongly indicates that environments conducive to marine mammals and to their fossil preservation were continuously present from the early Eocene to the Recent.

DISCUSSION

The results presented here for Cetacea match some of the results that have been presented for similar studies of the entire Phanerozoic biota. Our measures of research effort (papers in the PBDB and in GeoRef) had an even closer relationship with the number of fossil cetacean genera than did the measure of research effort used by Sheehan (1977) for marine invertebrates. In that case, however, both research effort and the measure of diversity matched the estimated map area and volume of rock in which the fossils were found (Raup 1977; Sheehan 1977), whereas our study found only occasional correspondence of diversity with map area. Raup (1977) was neither surprised nor concerned by the correlation of research effort with his measure of diversity, because he thought that paleontologists studied what was available for study, and that was in turn related to the amount of rock available in which to find fossils.

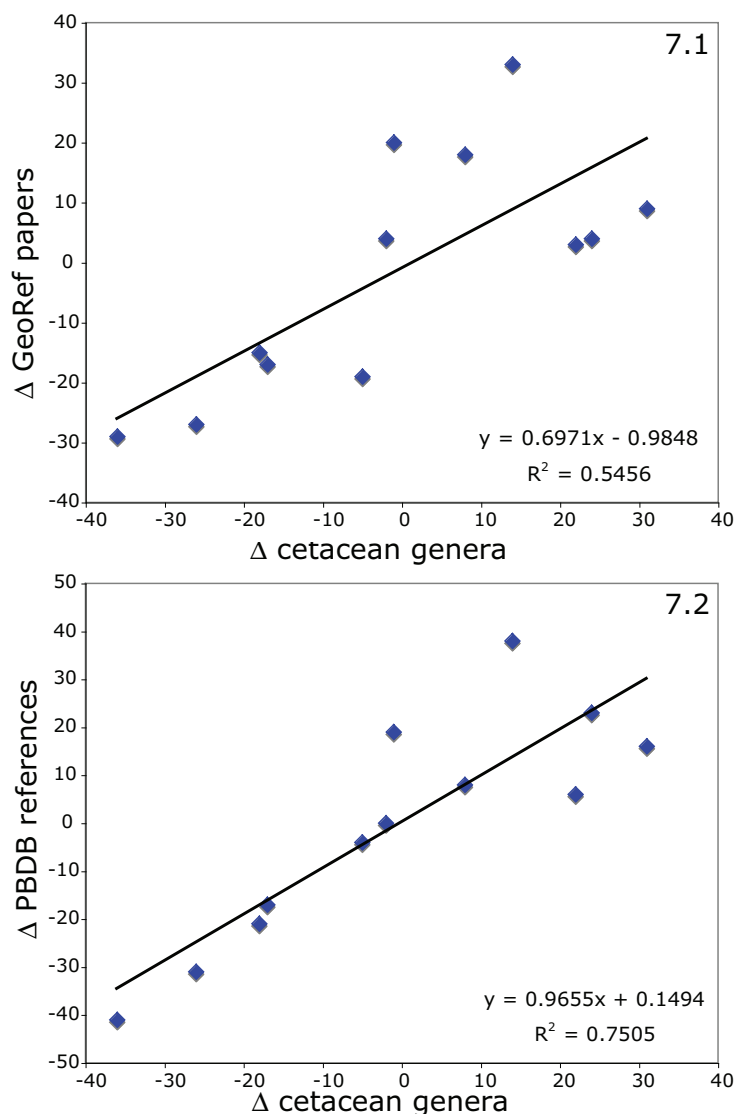


Figure 7. Correlations of first differences of references with global cetacean richness (as sampled, not ranged through). **7.1.** Correlation of GeoRef papers with cetacean genera. **7.2.** Correlation of PBDB references with cetacean genera. Note that both display a positive correlation, but the relationship is much tighter with PBDB references, as one might expect. Note also that the GeoRef and PBDB reference samples are not independent since most of the PBDB references are in GeoRef and vice versa.

It is difficult to discern whether there is a problem with research effort. It is not surprising that there is more research where and when there are more fossils (Sheehan 1977). Without prior knowledge of the fossils yet to be studied, we cannot determine whether a particular taxon, area, or time interval is understudied relative to other time intervals. We can however get a glimpse at an earlier stage of the journey from fossil preservation to publication by looking at the fossils in museum databases. These fossils have at least been recovered, but many remain unstudied. If the distribution

of studied fossils does not match that of unstudied ones, then we might suspect that there is a problem with research effort. To estimate how serious a problem this might be, we tallied how many of the specimens in the USNM specimen database were assigned to particular genera for each Cenozoic epoch. These results show that 74% of Eocene specimens, 75% of the Miocene specimens, 86% of the Pliocene specimens, but only 14% of the Oligocene specimens are assigned to named genera. The proportion of Oligocene taxa in the USNM collections assigned to the “*incertae sedis*” category

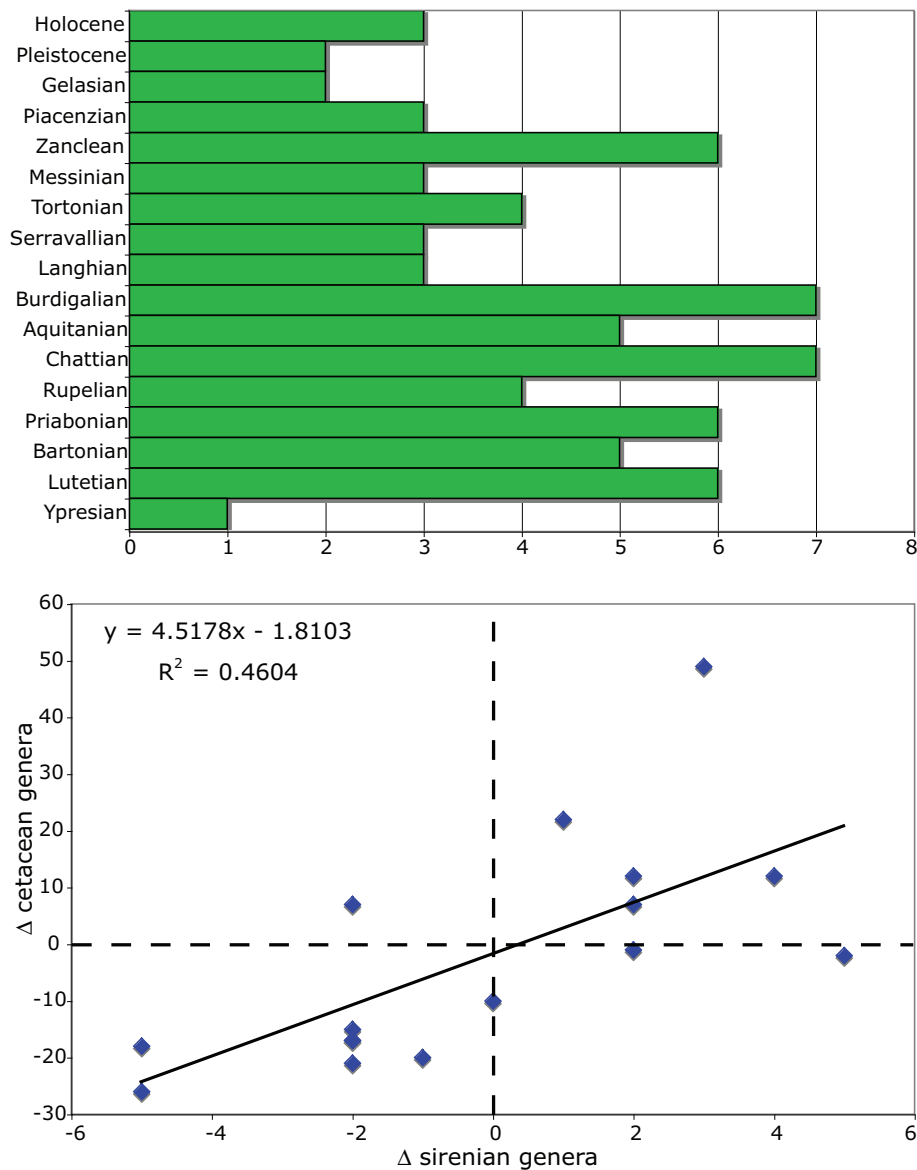


Figure 8. Sirenian diversity. **8.1.** Bar chart showing the number of sirenian genera described from each time interval for the entire world. Note the much lower absolute diversity of sirenians in each time interval when compared to cetaceans. **8.2.** Correlation of first differences of cetacean genera and sirenian genera (as sampled, not ranged through). Note that there is a general positive correlation, but not a significant correlation. This relationship suggests that different factors are affecting sirenian and cetacean diversity.

supports our suspicion that there are likely to be many more Oligocene genera than are currently in our data set. This preponderance of “phantom” Oligocene genera is consistent with our own intuition based on many years of work in the USNM and other museums. Such “phantoms” of diversity may not be a problem for diversity over time studies for invertebrate paleontology, but undescribed taxa are real problems in vertebrate paleontology where many taxa are represented by singular specimens. Although it remains to be demonstrated empirically,

it is possible that when more of the Oligocene specimens already in museums are studied, the Oligocene (at least the late Oligocene) will look much more like the Miocene, and the origins of modern families of cetaceans will be much better understood.

Recently, Smith (2001) demonstrated a correlation between rock area and measured diversity using outcrop area of rocks in France and the United Kingdom using Sepkoski’s (1997) generic level diversity database; similarly, Crampton et al.

(2003) demonstrated a correlation between rock area and mollusk species in New Zealand. The high correlation between rock units and diversity changes over time compelled Raup (1977), Smith (2001), and Crampton et al. (2003) all to call into question whether the number of taxa over time actually represent what diversity was like in the past. Our results depart sharply from these studies in that we did not find a correlation between the amount of rock available for study and the number of genera named from those rocks. It is then tempting to interpret our strong correlation of research effort with the number of diagnostic genera as having a different meaning. Perhaps both the number of papers on fossil Cetacea and the number of diagnostic cetacean genera are driven by the actual pattern of generic diversity, which is what we first attempted to discern.

Unlike the entire Phanerozoic marine invertebrate record, a working knowledge of the entirety of the order Cetacea is within the grasp of individual researchers. We suspect that the pattern of diagnostic generic diversity that we report is a fair representation of the pattern of generic diversity for the order, with the caveat for the Oligocene noted above. This observation supports the hypothesis of an early Oligocene rapid radiation of Neoceti (Fordyce 1992; Fordyce 2002a; Fordyce and de Muizon, 2001; Whitmore and Sanders 1977), and also calls into question whether generic richness was truly different in the Oligocene vs. the Miocene.

Some recent studies such as those of Smith (2001), Peters and Foote (2002), and Crampton (Crampton et al. 2003) strongly link patterns of observed diversity or changes in taxonomic richness to geologic processes or patterns that may not relate directly to the true pattern of taxonomic richness over time. All of these studies link changes in taxonomic richness to the amount of rock available for study, even though they use very different approaches. Peters and Foote (2002) use the number of named formations as a proxy for available rock whereas Smith (2001) noted the presence of rocks of particular ages on a series of small scale geologic maps of France and the United Kingdom. Crampton et al. (2003) found that mollusk species richness correlated with both rock area and formation names, but the fit was better with rock area. Although our data do not show a strong relationship between available rock and generic richness, two time intervals that have a low map area of rocks are worth discussing in more detail.

The first time interval with a low area map is the early Oligocene. Generally, it is thought that the Eocene-Oligocene boundary demarcates the origin of modern cetacean suborders from derived archaeocetes, with the nearly simultaneous (pseudo) extinction of the archaeocetes (see Barnes et al. 2001; Fordyce 2002b for possible alternatives). Our study provides qualitative and quantitative justification for paleocetologists to continue focusing their efforts on the Oligocene (Fordyce 2003a; Whitmore and Sanders 1977). As noted above, many fossils that have already been collected will help to fill in the diversity during this time interval.

The second time interval with a low map area is the early Pliocene. Despite the paucity of rocks in which to find cetacean fossils (Figure 4), observed diversity is relatively high. We attribute this fact to the willingness of taxonomists to assign a generic identification of a living genus to a fossil that they might not have assigned to a genus at all in the Pliocene vs. earlier time intervals. Support for this suggestion comes from the observation that many of the named genera in the Pliocene are extant genera (25%), whereas only 8% of Miocene genera are extant and none in the Oligocene or Eocene are extant.

Correlation of sirenian and cetacean diversity may reflect a common affinity for warm, tropical to subtropical environments. Today, all sirenians (with the exception of the recently extirpated *Hydrodamalis*) live in these types of environments. The correlation of the diversity of these two groups, though weak, may track the waxing and waning of these environments through geologic time. Another study currently under way will attempt to explore this idea further by actually tracing co-occurrence of Sirenia and Cetacea along with lithologic indicators of environmental factors.

Finally, it is worth noting that a great deal more work remains to be done on fossil cetaceans in general, not just during the Oligocene. Figure 9 shows the number of fossil cetacean genera vs. the number of collections in which they occur. Many genera occur in only a single collection and often are represented by single individuals. As a direct result, very few fossil cetacean genera range for more than a single stage. It is clear that many taxa remain to be discovered even in localities that have been collected for decades. Collection of new specimens from these localities with good age control will help clarify the age range of fossil cetacean taxa.

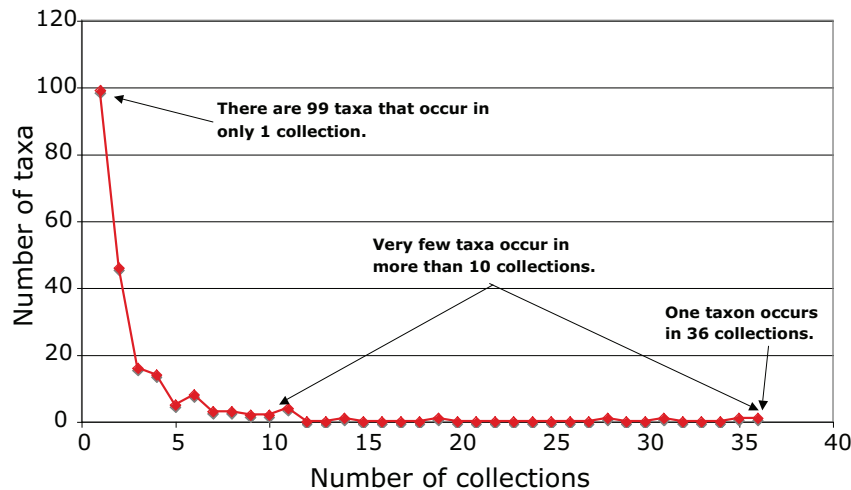


Figure 9. Frequency distribution of the number of taxa occurring in the number of collections. This plot shows that most taxa are known from a very small number of collections, usually less than 10. The taxon that occurs most frequently is *Squalodon*, which is probably over-identified in the fossil record. Most occurrences of *Squalodon* are identified based on teeth, and *Squalodon*-like teeth occur in many taxa, not just the genus *Squalodon*. See Dooley (2005). Whereas many of the taxa that occur in more collections are recent genera (such as *Balaenoptera* and *Delphinapterus*), many are not, including the second most frequently occurring taxon, *Basilosaurus*, which is Eocene in age, which also suggests that this phenomenon is not completely due to the pull of the Recent.

Figure 10 shows a plot of origination rate and extinction rate for global cetacean genera from the Eocene to Holocene. Flat areas at the beginning and end are caused by the lack of genera crossing stage boundaries in those intervals. It appears that increases in diversity from the late Oligocene to Middle Miocene are driven by an increase in origination rate, coupled with a relatively low extinction rate. Extinction rate peaks in the late Miocene, causing a drop in diversity during the Messinian when coupled with a low origination rate. This result may be caused by a data quality problem in the Messinian caused by the drying of the Mediterranean, which in earlier time intervals was a source of many fossil cetacean collections. Origination picks up again in the Pliocene, adding several genera that persist to the present.

So, what controls our understanding of fossil cetacean diversity? Primarily, it appears to be the underlying diversity of the fossil cetaceans themselves. The diversity pattern does not seem to be affected by the population of potential collectors in a fossil bearing area. This pattern is perhaps distorted by the paucity of early Oligocene rocks. The pattern may be distorted by zealous individual collectors or groups of collectors in particular times and areas, although this pattern may equally reflect the abundance of fossils in those particular places. Cetacean diversity is certainly distorted by the lack of study of existing late Oligocene fossils, a prob-

lem which is decidedly solvable. Fordyce (1992) graphically demonstrated this problem in his Figure 18.2. He shows ~ 40 species of late Oligocene cetaceans, of which twelve are informally designated.

In conclusion, we reiterate the need for using appropriate scales when discussing diversity through geologic time. Because patterns of diversity over the entire Phanerozoic may be too overwhelmed by serious biases to draw detailed conclusions about the history of life, the evolution of smaller groups of organisms may present more tractable cases of diversity through time if potential biases are addressed and corrected, as we have done with cetaceans. For Cetacea, it appears that the fossil record is generally good, with a problem only in the Oligocene. With more research effort geared toward finding more fossils and preparing and describing those fossils that have already been discovered, we should be able to come to a better understanding of the basic pattern of cetacean diversity through time, in coordination with the manifold biotic and abiotic factors related to their evolutionary history.

ACKNOWLEDGMENTS

We would like to thank P.A. Holroyd at the University of California Museum of Paleontology; T.A. Deméré, San Diego Museum of Natural History; S.A. McLeod, Natural History Museum of Los

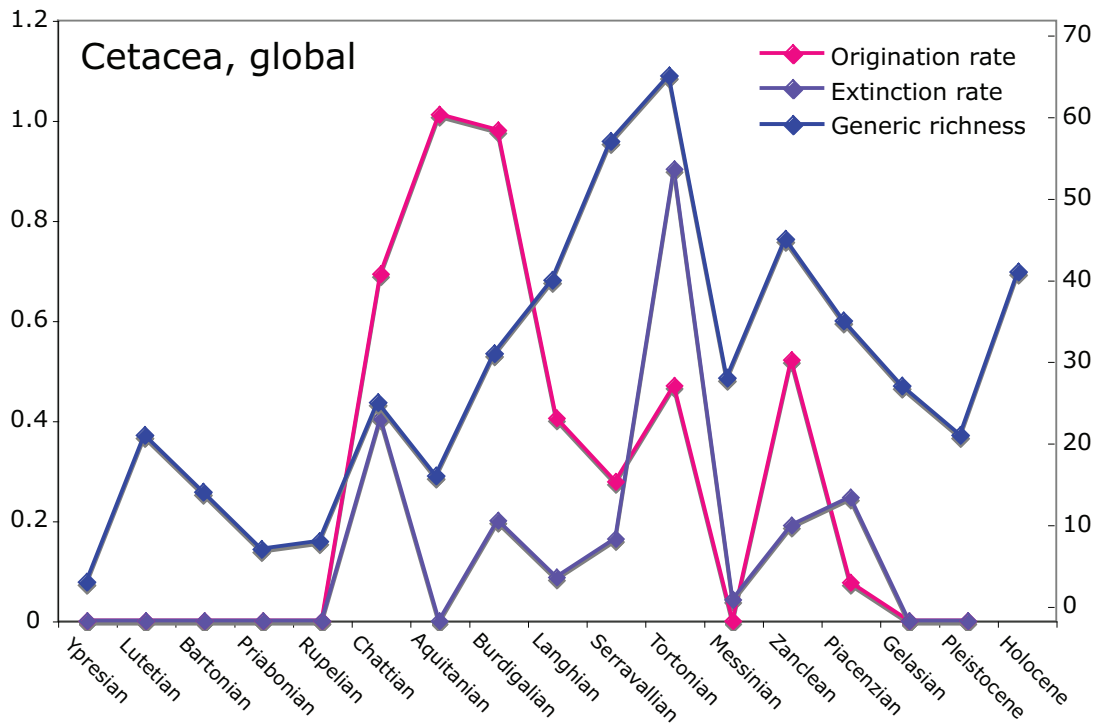


Figure 10. Origination and extinction rates for Cetacea along with generic richness. Origination and extinction rates are not able to be calculated in the earlier portion of the graph due to a lack of data. Note that richness values are primarily driven by higher extinction rates or higher origination rates at different times in cetacean history.

Angeles County; R. Hulbert, Florida Museum of Natural History; and M. Brett-Surman, United States National Museum who all provided data on the cetaceans in their collections management databases. N.D.P. would like to thank A.B. Barnosky, E.B. Davis, P.A. Holroyd, D.R. Lindberg, and G.P. Wilson for comments on early drafts. Insightful and constructive comments from two anonymous reviewers greatly improved the quality of this manuscript. This research was supported by a grant from the UCMP Remington Kellogg Fund and the NSF Graduate Research Fellowship (both to N.D.P.). This paper is UCMP Contribution No. 1862 and PBDB contribution number 65.

REFERENCES

- Adrain, J.M. and S.R. Westrop. 2000. An empirical assessment of toxic paleobiology. *Science*, 289:110-112.
- Alroy, J. 2002. How many named species are valid? *Proceedings of the National Academy of Sciences*, 99(6):3706-3711.
- Bajpai, S. and P.D. Gingerich. 1998. A new Eocene archaeocete (Mammalia, Cetacea) from India and the time of origin of whales. *Proceedings of the National Academy of Science*, 95:15464-15468.
- Barnes, L.G. 1977. Outline of eastern North Pacific fossil cetacean assemblages. *Systematic Zoology*, 25(4):321-343.
- Barnes, L.G., Domning, D.P., and Ray, C.E. 1985. Status of studies on fossil marine mammals. *Marine Mammal Science*, 1(1):15-53.
- Barnes, L.G., Goedert, J.L., and Furusawa, H. 2001. The earliest known echolocating toothed whales (Mammalia; Odontoceti): Preliminary observations of fossils from Washington State. *Mesa Southwest Museum Bulletin*, 8(Western Association of Vertebrate Paleontologists with Mesa Southwest Museum and Southwest Paleontological Society: First Meeting of the New Millennium):91-100.
- Behrensmeier, A.K., Kidwell, S.M., and Gastaldo, R.A. 2000. Taphonomy and paleobiology. *Paleobiology*, 26(Deep Time: Paleobiology's Perspective, Supplement to no. 4):101-147.
- Benton, M.J., Wills, M.A., and Hitchin, D.R. 2000. Quality of the fossil record through time. *Nature*, 403:534-537.
- Bicker, A.R., Jr. 1969. *Geologic Map of Mississippi*. Mississippi Geological Survey, Jackson, Mississippi.
- Caruthers, C., Heinritz, A., Schuster, E., Ikerd, K., and Harris, C. 2002. *Geologic Map of Washington-North-*

- west Quadrant, Washington State Department of Natural Resources.
- Crampton, J.S., Beu, A.G., Cooper, R.A., Jones, C.M., Marchall, B., and Maxwell, D.P.A. 2003. Estimating the Rock Volume Bias in Paleobiodiversity Studies. *Science*, 301:358-360.
- Davis, E.B. and Pyenson, N.D. 2007. Diversity biases in terrestrial mammalian assemblages and quantifying the differences between museum collections and published accounts: a case study from the Miocene of Nevada. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 250:139-149.
- Delaware Geological Survey. 1976. *Generalized Geologic Map of Delaware*. Delaware State Planning Office, Dover, Delaware.
- Domning, D.P. 2001. Evolution of the Sirenia and Desmostylia, p. 151-168. In J.-M. Mazin and V. d. Buffr n il (eds.), *Secondary Adaptation of Tetrapods to Life in Water*. Verlag Dr. Friedrich Pfeil, M nchen, Germany.
- Donovan, S.K. 2001. Evolution of Caribbean echinoderms during the Cenozoic: moving towards a complete picture using all of the fossils. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 166:177-192.
- Dooley, A.C., JR. 2005. A new species of *Squalodon* (Mammalia, Cetacea) from the Middle Miocene of Virginia. *Virginia Museum of Natural History Memoirs*, 8:1-14.
- Earnest, P., Newman, M., Garcia, D., Goodman, J., and Myers, E. 1991. *Geologic Map of Oregon*. U. S. Department of the Interior.
- Foote, M. 2000. Origination and extinction components of taxonomic diversity: Paleozoic and post-Paleozoic dynamics. *Paleobiology*, 26(4):578-605.
- Foote, M. and Sepkoski, III, J.J. 1999. Absolute measures of the completeness of the fossil record. *Nature*, 398:415-417.
- Fordyce, R.E. 1992. Cetacean evolution and Eocene/Oligocene environments, p. 368-381. In D. R. Prothero and Berggren, W.A. (eds.), *Eocene-Oligocene Climatic and Biotic Evolution*. Princeton University Press, Princeton, New Jersey.
- Fordyce, R.E. 2002a. Fossil Record, p. 453-471. In Perin, W.F., W rsig, B. and Thewissen, J.G.M. (eds.), *Encyclopedia of Marine Mammals*. Academic Press, San Diego.
- Fordyce, R.E. 2002b. Oligocene archaeocetes and toothed mysticetes: Cetacea from times of transition. *Geological Society of New Zealand Miscellaneous Publication*, 114A (Secondary Adaptation to Life in Water):16-17.
- Fordyce, R.E. 2003a. Cetacean Evolution and Eocene-Oligocene Oceans Revisited, p. 154-170. In Prothero, D.R., Ivany, L.C., and Nesbitt, E.A. (eds.), *From Greenhouse to Icehouse: The Marine Eocene-Oligocene Transition*. Columbia University Press, New York.
- Fordyce, R.E. 2003b. Early crown-group Cetacea in the southern ocean: the toothed archaic mysticete *Llanocetus*. *Journal of Vertebrate Paleontology*, 23(3):50A.
- Fordyce, R.E. and de MUIZON, C. 2001. Evolutionary history of cetaceans: a review, p. 169-223. In Mazin, J.-M. and Buffr n il, V. (eds.), *Secondary Adaptation of Tetrapods to Life in Water*. Verlag Dr. Friedrich Pfeil, M nchen, Germany.
- Gingerich, P.D. and Uhen, M.D. 1998. Likelihood estimation of the time of origin of Cetacea and the time of divergence of Cetacea and Artiodactyla. *Paleontologica Electronica*, 1(2).
- Hartmann, B.M. and Scranton, D. 1992. *Geologic Map of Texas*. Bureau of Economic Geology, the University of Texas at Austin, Austin, Texas.
- Jablonski, D. 1999. The future of the fossil record. *Science*, 284:2114-2116.
- Jablonski, D., Roy, K., Valentine, J.W., Price, R., and Anderson, P.S. 2003. The impact of the pull of the recent on the history of marine diversity. *Science*, 300(May 16):1133-1136.
- Jablonski, D. and Sepkoski, Jr., J.J. 1996. Paleobiology, community ecology, and scales of ecological pattern. *Ecology*, 77(5):1367-1378.
- Jackson, J.B.C. and Johnson, K.G. 2001. Measuring past biodiversity. *Science*, 283:2401-2404.
- Jackson, J.B.C., Kirby, M.X., Berger, W.H., Bjorndal, K.A., Botsford, L.W., Bourque, B.J., Bradbury, R.H., Cooke, R., Erlandson, J., Estes, J.A., Hughes, T.P., Kidwell, S., Lange, C.B., Lenihan, H.S., Pandolfi, J.M., Peterson, C.H., Steneck, R.S., Tegner, M.J., and Warner, R.R. 2001. Historical overfishing and the recent collapse of coastal ecosystems. *Science*, 293:629-638.
- Jennings, C.W. 1977. *Geologic Map of California*. Department of Conservation, Division of Mines and Geology, State of California, Sacramento, California.
- Johnson, S.S. 1993. *Geological Map of Virginia*. Commonwealth of Virginia Department of Mines, Minerals, and Energy Division of Mineral Resources, Charlottesville, Virginia.
- Kellogg, R. 1968. Miocene Calvert mysticetes described by Cope. *Bulletin of the United States National Museum*, 247(5):103-132.
- Lu, P.J., Yogo, M., and Marshall, C.R. 2006. Phanerozoic marine biodiversity dynamics in light of the incompleteness of the fossil record. *Proceedings of the National Academy of Science*, 103(8):2736-2739.
- Maryland Geological Survey. 1968. *Geologic Map of Maryland*. Maryland Geological Survey.
- Miller, A.I. and Foote, M. 2003. Increased longevities of Post-Paleozoic Marine Genera after mass extinctions. *Science*, 302:1030-1032.
- Mishler, B.D. 1999. Getting rid of species?, p. 307-315. In Wilson, R. (ed.), *Species: New Interdisciplinary Essays*. MIT Press, Cambridge, Massachusetts.
- North Carolina Geological Survey. 1985. *Geologic Map of North Carolina*. North Carolina Department of Natural Resources and Community Development.

- Osborne, W.E., Snabo, M.W., Copeland, Jr., C.W., and Neathery, D.T.L. 1989. *Geologic Map of Alabama*. Geological Survey of Alabama, Tuscaloosa, Alabama.
- Peters, S.E. and Foote, M. 2001. Biodiversity in the Phanerozoic: a reinterpretation. *Paleobiology*, 27(4):583-601.
- Peters, S.E. and Foote, M. 2002. Determinants of extinction in the fossil record. *Nature*, 416:420-424.
- Pickering, S.M. and Murray, J.B. 1976. *Geologic Map of Georgia*. Georgia Geological Survey, Atlanta.
- Raup, D.M. 1976a. Species diversity in the Phanerozoic: an interpretation. *Paleobiology*, 2:289-297.
- Raup, D.M. 1976b. Species diversity in the Phanerozoic: a tabulation. *Paleobiology*, 2:279-288.
- Raup, D.M. 1977. Systematists follow the fossils. *Paleobiology*, 3:328-329.
- Scott, T.M., Campbell, K.M., Rupert, F.R., Arthur, J.D., Missimer, T.M., Lloyd, J.M., Yon, J.W., and Duncan, J.G. 2001. *Geologic Map of the State of Florida*. Florida Geological Survey.
- Sepkoski, J.J., Jr. 1978. A kinetic model of Phanerozoic taxonomic diversity I. Analysis of marine orders. *Paleobiology*, 4(3):223-251.
- Sepkoski, J.J., JR. 1997. Biodiversity: Past, present and future. *Journal of Paleontology*, 71(4):533-539.
- Sheehan, P.M. 1977. A reflection of labor by systematists? *Paleobiology*, 3:325-329.
- Smith, A.B. 2001. Large-scale heterogeneity of the fossil record: implications for Phanerozoic biodiversity studies. *Philosophical Transactions of the Royal Society of London, Series B, Biological Science*, 356:351-367.
- Snead, J.I. and McCulloh, R.P. 1984. *Geologic Map of Louisiana*. Louisiana Geological Survey, Baton Rouge, Louisiana.
- Thewissen, J.G.M., Madar, S.I., and Hussain, S.T. 1996. *Ambulocetus natans*, an Eocene cetacean (Mammalia) from Pakistan. *Courier Forschungsinstitut Senckenberg*, 191:1-86.
- Uhen, M.D. 1996. An evaluation of clade-shape statistics using simulations and extinct families of mammals. *Paleobiology*, 22(1):8-22.
- Uhen, M.D. 1998. Middle to Late Eocene Basilosaurines and Dorudontines, p. 29-61. In Thewissen, J.G.M. (ed.), *The Emergence of Whales*. Plenum Press, New York.
- UHEN, M.D. 1999. New species of protocetid archaeocete whale, *Eocetus wardii* (Mammalia, Cetacea), from the middle Eocene of North Carolina. *Journal of Paleontology*, 73(3):512-528.
- US Geological Survey. 1936. *Geologic Map of South Carolina*. United States Geological Survey.
- Wagner, P.J., Aberhan, M., Hendy, A., and Kiessling, W. 2007. The effects of taxonomic standardization on sampling-standardized estimates of historical diversity. *Proceedings of the Royal Society B*, 274:439-444.
- Walsh, T.J., Koorosec, M.A., Phillips, W.M., Logan, R.L., and Schasse, H.W. 1987. *Geologic map of Washington-Southwest Quadrant*. Washington Division of Geology and Earth Resources.
- White, T. 2003. Early Hominoids - Diversity or Distortion? *Science*, 299:1994-1998.
- White, T.D. and Folkens, P.A. 2000. *Human Osteology*. Academic Press, San Diego, 563 p.
- Whitmore, JR., F.C., and Barnes, L.G. In press. *The Herpetocetinae, a new subfamily of extinct baleen whales (Mammalia, Cetacea, Cetotheriidae)*. Virginia Museum of Natural History in association with the Smithsonian Institution.
- Whitmore, JR., F.C. and Kaltenbach, J.A. In press. *Neogene Cetacea of the Lee Creek Phosphate Mine, North Carolina*. Virginia Museum of Natural History in association with the Smithsonian Institution.
- Whitmore, Jr., F.C., and Sanders, A.E. 1977. Review of the Oligocene Cetacea. *Systematic Zoology*, 25(4):304-320.

APPENDIX

Cetacean genera here designated or confirmed as *nomina dubia* due to the poor quality of the type material. Genera in bold were designated as “non-diagnostic” by Fordyce and de Muizon (2001). The genera marked with an asterisk (*) were designated as “non-diagnostic” by Fordyce and de Muizon (2001) but have had more recent opinions reasserting their diagnosability.

Agriocetus*
Amphicetus
Araeodelphis
Balaenodon*
Belemnoziphius*
Dinoziphius
Eboroziphius
Graphiodon*
Helvicetus*
Heterocetus
Kogiopsis*
Macrochirifer
Megalodelphis
Microcetus*
Microzeuglodon*
Mioceta
Miokogia*
Miotursiops
Neosqualodon*
Pachycetus
Palaeocetus
Paleophoca*
Palaeoziphius*
Pelycorhamphus*
Physetodon*
Platyosphys
Priscophyseter*
Probalaena
Prophyseter*
Proterocetus
Rhytisodon*
Saurocetus*
Scaptodon*
Thalassocetus*