

TAXONOMIC INFLATION AND BODY MASS DISTRIBUTIONS IN NORTH AMERICAN FOSSIL MAMMALS

JOHN ALROY*

*National Center for Ecological Analysis and Synthesis, University of California,
Santa Barbara, CA 93101, USA*

Body mass distributions of mammalian species are a major focus of macroecological and macroevolutionary studies. However, these distributions may be obscured by taxonomic error, just like any other aspect of biodiversity. The key problem with taxonomy is that many currently used names are synonyms of each other or are biologically indeterminate. This article reassesses body mass patterns in the fossil record of North American mammals using the recently developed flux ratio method for estimating the underlying proportion of invalid names. Current name quality varies very strongly with body mass: small species names are highly unreliable, but names of large species have been evaluated thoroughly. The main reason is that there has been a dramatic fall through historical time in the average size of described species. Hence, there simply has not been enough time yet to reevaluate the names of most small species. This bias only accentuates the previously described bimodal diversity distribution for North American mammals, which suggests the existence of dual body mass optima—so not all evolutionary lineages converge on 100 g. The historical shift in the underlying quality and body mass of newly described species also differentially affects our picture of biodiversity in major taxonomic groups. On the one hand, ungulate and carnivoran names are much more likely to be invalid in the 1st place than are rodent and insectivoran names. On the other hand, most of the invalid names for large mammals already have been identified, but this is not true for the small-mammal groups. Therefore, the most fruitful strategy for future taxonomic research would be to focus on small- and medium-sized mammals.

Key words: body mass, Cenozoic, fossil record, Mammalia, North America, synonymy, taxonomic diversity

The fossil record is a valuable resource in evolutionary and ecological studies, and the North American mammalian record is particularly well studied. Indeed, this record has been used for decades (e.g., Lillegraven 1972; Simpson 1953; Stucky 1990) to examine many major events in Earth history and mammalian evolution, such as the Cretaceous–Tertiary boundary mass extinction and ensuing Paleocene radiation (Alroy 1999; Archibald 1993; Foote et al. 1999), the major faunal reorganization at the Paleocene–Eocene boundary (Alroy et al.

2000; Clyde and Gingerich 1998; Gingerich 1989; Maas et al. 1995), and the end-Pleistocene megafaunal mass extinction (Alroy 2001; Martin 1967). The same data show that the major control on diversification is biological interactions leading to diversity equilibrium (Alroy 1996, 1998b; Stucky 1990; Webb 1969) and that major climate shifts have unpredictable and frequently inconsequential effects (Alroy 1996, 1998b; Alroy et al. 2000; Barnosky 2001; Prothero 1999; Prothero and Heaton 1996). Finally, big-picture research has revealed dramatic trends in the evolution of morphology and

* Correspondent: alroy@nceas.ucsb.edu

ecological strategies, such as the rampantly convergent evolution of dental adaptations for carnivory (Van Valkenburgh 1988) and herbivory (Hunter and Jernvall 1995), the surprisingly decoupled acquisition of cursorial locomotor adaptations in carnivores and ungulates (Janis and Wilhelm 1993; Van Valkenburgh 1985), and Cope's rule of increasing body mass (Alroy 1998a; MacFadden 1986; McShea 1994; Stanley 1973).

Recent advances in our knowledge of mammalian evolution have depended not just on the steady acquisition of more data (e.g., Alroy 2000b) but also on methodological refinements. For example, over the last decade the subjective, traditional "land-mammal age" system, whose coarse and uneven intervals distort turnover rate data (Archibald 1993; Barnosky 2001; Maas et al. 1995; Stucky 1990), has been replaced with objective and fully quantitative time-scales that can be broken into uniform bins 1.0 million years long (Alroy 1992, 1994, 1996, 1998b, 2000b, 2000c); confounding variation through time in the amount of data (Maas et al. 1995; Stucky 1990) has been controlled with standardized subsampling methods (Alroy 1996, 1998b, 2000c; Barnosky 2001); both diversity metrics (Alroy 1996; Maas et al. 1995) and turnover rate metrics (Alroy 2000c; Foote 1999, 2000) have been improved; and ecomorphological shifts and evolutionary trends have been quantified more rigorously (Alroy 2000a, 2000c; Clyde and Gingerich 1998; Gingerich 1989; Gunnell et al. 1995; Janis and Wilhelm 1993; Janis et al. 2000; Jernvall et al. 1996).

Despite all this progress, the fossil record still has some important shortcomings. One persistent worry about big-picture diversity analyses is the quality of the underlying genus- and species-level taxonomy (Adrain and Westrop 2000; Niklas and Tiffney 1994; Raup 1972; Vermeij 1999; Wagner 1995). A large fraction of species and genera named in the past already are known to be invalid (Alroy 2002), so it stands to rea-

son that many more invalid names remain to be weeded out. Bad names could have many effects. For example, if the proportion of bad names is large enough, concentrations of these names in particular intervals of evolutionary time could severely distort a diversity curve. Likewise, if certain groups have relatively oversplit taxonomy, the overrepresentation of those groups might distort apparent changes in taxonomic composition, ecomorphology, and body mass distributions.

Surprisingly, the best way to tackle the taxonomic error problem turns out to involve historical analyses of taxonomic activity (Alroy 2002). Workers constantly disagree about whether names are valid, invalid because they are synonyms of previously proposed names, or invalid because they are biologically indeterminate (i.e., *nomen dubium*). Quantifying the constant back-and-forth between taxonomic lumpers and splitters by computing a "flux ratio" makes it possible to estimate the proportion of species names that will be accepted once these debates finally settle down. Assuming this is a good estimate of the number of true biological species, one can then compare it with the invariably larger tally of currently accepted species names. The ratio between the 2 counts is an estimate of the current taxonomy's reliability, and its inverse is the amount of diversity inflation due to taxonomic error.

In this article, I use the flux ratio method to reevaluate body mass distributions in the mammalian fossil record. I show that the average size of newly described species has fallen dramatically through historical time, spurred by the development of screen-washing collection techniques before World War II (Gingerich 1980; Hibbard 1975). There is a weak historical correlation between the body mass and underlying name quality of newly described species. Furthermore, large species are much older on average, and they have a much higher underlying probability of being invalid. This last pattern is particularly striking because

the current reliability of the species names varies in the opposite direction: younger species names are consistently less reliable. The reason is that small species are relatively new on average, so there has been much less time to evaluate them and remove synonyms. The broad-brush correlation between body size and taxonomic quality also translates into large differences among major taxonomic groups: small-mammal orders such as the Insectivora and Rodentia tend to be more poorly evaluated, so they clearly deserve more study.

MATERIALS AND METHODS

The taxonomic data used in this article have been detailed elsewhere (Alroy 2002) and are a part of the Paleobiology Database (<http://paleodb.org>). The data set comprises 4,861 originally proposed names of extinct fossil mammal species; some information on extant species having fossil records is available but was not used for reasons of methodological consistency. Historical opinions on the synonymy and validity of published names are tracked in detail, making it possible to reconstruct the per-year, per-taxonomic-worker rate at which proposed names are invalidated or revalidated. In addition to purely taxonomic data, most analyses also depend on previously published body mass data (Alroy 1998a, 2000c). These mass estimates depend on standard allometric equations (Alroy 1998a) that relate the log of lower 1st molar area to the log of body mass in grams, with each species having 1 average value computed from all available published molar measurements. A total of 22,335 measurements across 3,851 populations of 1,977 currently accepted species were used. Mass estimates for valid species also were applied to their junior synonyms, but *nomina dubia* had to be excluded because by definition they do not clearly pertain to narrowly defined biological species.

The basic idea of flux ratio analysis is to model the gradual historical process of assessing proposed species names (Alroy 2002). Proposed names are invalidated at some stochastic, per-name rate q that depends on the number of actively working taxonomists (Alroy 2002:3707, figure 3). The pool of invalid names is not fixed, however, because taxonomic splitters will argue for revalidating some of them. Revalidation

takes place at some per-name, per-worker rate r . Hence, taxonomic debate creates a steady back-and-forth flux between invalid and valid pools of species names. It can be shown that after an infinite amount of debate, the expected proportion of all names that will be recognized is just $r/(q + r)$, i.e., the flux ratio. The rates can be computed directly by tallying the number of times names have been invalidated or revalidated and the number of years they have been treated as valid and invalid. It is important to weight the year tallies for each species by counts of taxonomic workers who published in each of the relevant historical years. This worker-year scaling helps the flux ratio model to account for changes in the rate of taxonomic activity—a typical example being the marked slowdown in effort during World War II. In addition to estimating the underlying proportion of valid names, flux rates also can be used to project a quasi-exponential survivorship curve that shows the proportion of names expected to be treated as valid any particular number of years after being proposed.

In this article, I use statistics for skewness (g_1), significance (P -values based on t -tests), and rank-order correlation (r_s), which are standard throughout the scientific literature. Correlations are treated as potentially significant whenever $P < 0.10$.

TAXONOMIC QUALITY AND BODY MASS

Previous flux ratio analyses have established that there have been broad swings through historical time in taxonomic quality, which suggests that there are large differences in taxonomic reliability among major workers (Alroy 2002). These facts are of interest, but by themselves, they have no necessary significance for big-picture studies of evolution and ecology. After all, descriptions of species by different workers and from different historical periods could be distributed randomly with respect to any biological variable. However, it already is clear that the distribution of taxonomic bias across the Cenozoic is profoundly nonrandom (Alroy 2002). In this and the following sections, I test for 2 additional, interrelated, biologically interesting ways in which taxonomic bias might vary: across the shrew-

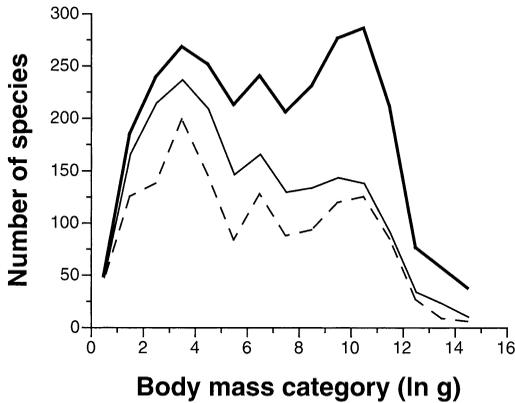


FIG. 1.—Counts of all named mammalian species (thick solid line), currently accepted mammalian species (thin solid line), and mammalian species estimated to be valid (dashed line) in each 1.0-ln g unit body mass category.

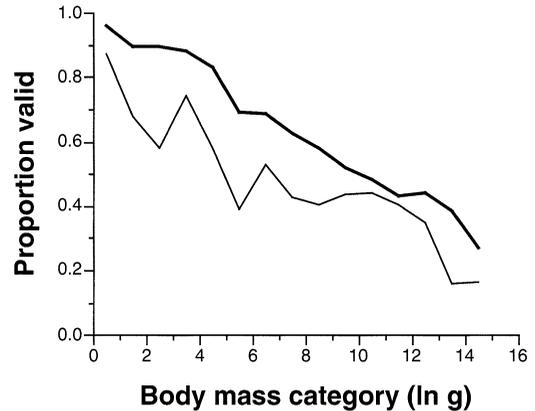


FIG. 2.—Proportions of named mammalian species in each 1.0-ln g unit body mass category that are currently accepted (thick line) estimated by flux ratio analysis to be valid (thin line).

to-mammoth gradient of body mass and across major phylogenetic groupings.

Basic pattern.—Body mass estimates are available for 1,883 (56.2%) of the 3,350 valid, named species. Better still, this subset includes the senior synonyms of an additional 944 invalid species, so the same estimates can be used for them. Several clear patterns emerge after binning species into 1.0-natural log (ln) unit slices of the body mass gradient (Fig. 1). The overall number of named species is remarkably flat between 2 and 13 ln g units (=7.4–442 kg), a span that includes 2,607 (92.2%) of the 2,827 names with body mass estimates. Indeed, the skewness of this curve actually is positive ($g_1 = 0.548$), meaning that its median is higher than its mean. However, the curve for currently valid species (Fig. 1) shows a marked peak centered between 3 and 4 ln g (20–55 g), with a steep falloff into a plateau that starts around 6 ln g (403 g) and weak skewness in the opposite direction ($g_1 = -0.261$). Because of this difference, the proportion of all species that are now valid drops with almost perfect linearity ($n = 15$, $r_s = -0.993$, $t = 30.004$, $P < 0.01$) from around 90% in the size range of most rodents down to less than 50% in the size range of ungulates (Fig. 2).

The flux ratio analysis tells a very different story. The underlying proportions of valid species (Fig. 2) still fall consistently with increasing body mass ($r_s = -0.871$, $t = 6.405$, $P < 0.01$). Because both curves plunge so dramatically, the ratio between the 2 (i.e., current quality) is uncorrelated with body mass ($r_s = 0.107$, $t = 0.389$, NS). However, the pattern is distorted by outlying values for the smallest bin and the largest 3 bins—i.e., the 4 bins with <100 species (Fig. 1). Focusing on the other 11 bins reveals a substantial, negative relationship ($n = 11$, $r_s = -0.527$, $t = 1.862$, $P < 0.1$). Where the flux ratios peak out at around 3/4 for rodent-sized mammals, they are mostly well below the currently accepted proportions (Fig. 2). However, for large mammals, ratios are only slightly lower than the currently recognized valid species proportions, and similar parity is seen among shrew-sized mammals (below 2 ln g units = 7 g). In other words, there is significant inflation of diversity for most rodent- and medium-sized mammals, peaking at 76% for 6- to 7-ln g unit (403–1,097 g) species and being at least 18% for all species between 2 and 11 ln g units (7.4–60 kg).

A simple correction for the bias is to multiply the total species count per bin by

the flux ratios to estimate the true number of valid species (Fig. 1). The resulting trend does confirm a peak between 4 and 5 ln g units (55–148 g), which is noteworthy because the same 100-g peak is observed in Recent mammals (Gardezi and da Silva 1999), not just in North America (Brown and Nicoletto 1991) but even in Australia (Chown and Gaston 1997) and South America (Marquet and Cofré 1999). Because of such patterns, a considerable body of theoretical literature has sprung up around a debate over whether a 100-g size is optimal for mammals in general (e.g., Alroy 1998a; Brown et al. 1993; Chown and Gaston 1997; Gardezi and da Silva 1999; Johst and Brandl 1997; Marquet and Taper 1998).

However, the corrected pattern (Fig. 1) just is not what one would expect if a simple, single optimum size governed species diversity. It clearly indicates a 2nd peak around 11 ln g units (60 kg). Indeed, the left-hand side is so depressed that the curve's skewness is now suddenly positive, not negative ($g_1 = 0.213$). The curve's positive skewness and slight bimodality is dramatically different from the gracefully negative skewing seen in curves for Recent mammals (Brown and Nicoletto 1991; Gardezi and da Silva 1999; Marquet and Cofré 1999) and assumed as the basis of many theoretical discussions (e.g., Brown et al. 1993; Chown and Gaston 1997; Gardezi and da Silva 1999; Johst and Brandl 1997; Marquet and Taper 1998). Even the long, flattened plateau in the uncorrected data (Fig. 1) runs against the standard picture.

The 2nd peak in the corrected data does make sense, however, in light of the overall temporal distribution of size estimates (Alroy 1998a:731, figure 1) and the underlying double-equilibrium, within-lineage evolutionary dynamic that has been demonstrated with a comparative statistical approach (Alroy 1998a). Specifically, the results confirm the argument that whatever might make 100-g mammals "optimal," the same rules do not apply to larger mammals, and spe-

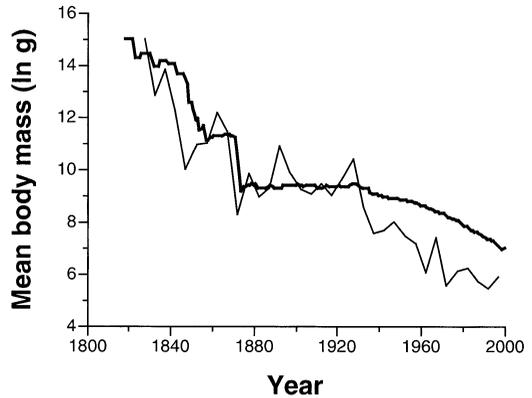


FIG. 3.—Mean natural log body mass of mammalian species recognized as valid in each historical year (thick line) and newly named within 5-year intervals (thin line).

cifically to ungulates; instead, consistent, population-level selection has been pushing medium-sized species up into the 100-kg range for much of the Tertiary.

Historical trends in description.—The enormous differences in taxonomic reliability among size classes turn out to have a predictable basis in paleontological history. The underlying cause can be illustrated by computing mean mass values for species recognized at any point in historical time, or for species newly named in particular 5-year cohorts (Fig. 3). The resulting time-versus-mass curves show remarkably strong trends, with earlier workers strongly favoring larger species (time versus mean mass of new species: $n = 29$ [1850–1999, 1860–1864 bin excluded], $r_s = -0.887$, $t = 9.965$, $P < 0.001$). Dates of naming of individual species (Fig. 4) also correlate with mass estimates ($n = 2,829$, $r_s = -0.407$, $t = 23.301$, $P < 0.001$). Indeed, the average new species was 61 kg in the late 1850s but consistently about 240–500 g throughout the 1970s, 1980s, and 1990s.

The overall pattern of decline is somewhat steplike, with averages for new species dropping in the 1870s, near the start of the epic feud between E. D. Cope and O. C. Marsh (Wallace 1999), and again around 1930. The average size of accepted species

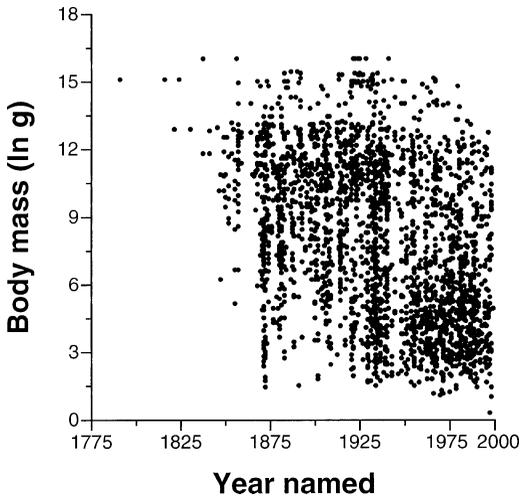


FIG. 4.—Correlation between natural log body mass estimates and dates of naming of fossil mammal species.

was remarkably stable during the 6 full decades between these 2 episodes. Ever since the 1930 shift, the decline in average size has if anything continued to accelerate, with averages for new species consistently being much lower than standing averages for all species (Fig. 3). This remarkable transition away from a stable distribution dominated by large mammals only followed the adoption of screen-washing techniques by workers like Hibbard, which allowed intense study of small taxa such as insectivorans and rodents. The shift was made possible by a straightforward technological innovation: Hibbard's independent 1928 rediscovery (Hibbard 1975) of screening methods briefly used by Wortman in 1891 and Loomis in 1904 (Gingerich 1980).

As noted earlier, the underlying proportion of valid names varies somewhat erratically through historical time, but results to be presented elsewhere demonstrate that current rates of taxonomic acceptance have increased steadily through the 20th century. Not surprisingly, this trend results in a strong among-cohort cross-correlation between currently accepted proportions and mean mass ($n = 27$ [1990s omitted because insufficient time has elapsed to quantify

flux rates], $r_s = -0.666$, $t = 4.462$, $P < 0.001$). However, there is only a weak cross-correlation for the underlying valid proportions and mean mass values ($r_s = -0.358$, $t = 1.916$, $P < 0.1$). This is something of a surprise because the 2 things do correlate so strongly across body mass bins, but it makes sense, given work in preparation that demonstrates an unpredictable historical trend in underlying quality. Put together, the 2 patterns fail to generate a correlation between size and current taxonomic quality ($r_s = -0.178$, $t = 0.906$, NS). Nonetheless, the steeper trend in current acceptance rates tends to confirm the idea that recently proposed, less-examined species are both smaller and less reliable. This would explain the general unreliability of names of small species, as visible in the binned data (Fig. 1).

Of course, all these correlations might be undesirably liberal because time series autocorrelation would lead to spurious cross-correlations (see Alroy et al. 2000 for a representative discussion of this well-known statistical artifact). However, although mean body mass is strongly autocorrelated ($n = 29$, $r_s = 0.769$, $t = 5.891$, $P < 0.001$), the taxonomic quality time series are not (flux ratios: $n = 26$, $r_s = 0.287$, $t = 1.467$, NS; current taxonomic quality: $n = 26$, $r_s = 0.115$, $t = 0.565$, NS). Furthermore, the point is somewhat moot because the correlations merely provide a broad-brush historical explanation for the undeniable biological problem highlighted by the body mass binning analysis (Fig. 1).

Worker effects.—Research in preparation shows that decadal-scale trends in name quality actually are largely a side effect of varying taxonomic practices among major workers. This raises the question of whether taxonomic differences among body mass classes also might be generated by worker effects. The question can be answered by jointly examining the quality and mean body mass of species proposed by individual researchers. There is plenty of variation in the latter variable because many taxon-

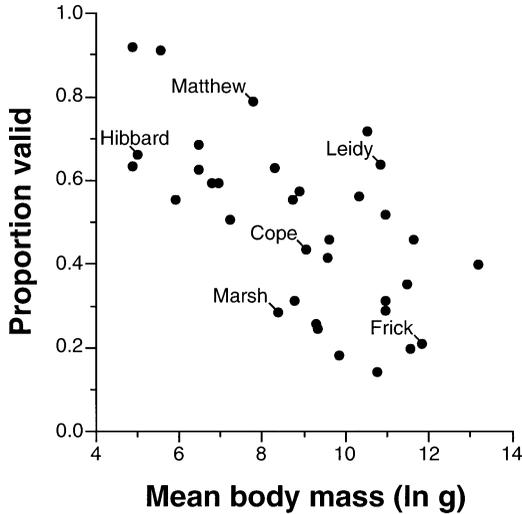


FIG. 5.—Correlation between average body mass and estimated valid proportion of mammalian species named by 34 productive taxonomists. Names indicate values for 6 taxonomists who are of particular historical importance.

omists do specialize narrowly on large mammals like ungulates or small mammals like rodents. Because the flux ratio does not vary systematically through time but both current taxonomic quality and mean body mass do, it is reasonable to predict that 1) an individual author's flux ratio should not correlate with the mean body mass of species named by that author, but 2) their names' current taxonomic quality should.

As it happens, there actually is a strong obvious correlation among productive workers between mean body mass and the flux ratio ($n = 34$, $r_s = -0.630$, $t = 4.585$, $P < 0.001$; Fig. 5)—so workers on larger-sized taxa have consistently produced less reliable names. In retrospect, this pattern is not really so surprising because large-mammal groups have tended to attract the interest of such well-known splitters as C. Frick, O. P. Hay, O. C. Marsh, W. B. Scott, and C. B. Schultz. It also turns out not to be so contradictory because of the unpredictable way that worker reliability scores are distributed through time. As one might expect based on the strong trend through time in

body mass (Fig. 3), there is a weak rank-order correlation between the average year that a major author's species names were proposed and the geometric mean body mass of these species ($r_s = -0.299$, $t = 1.771$, $P < 0.10$). However, there is no correlation at all between the average year of naming and the author's flux ratio ($r_s = 0.101$, $t = 0.576$, NS). Thus, splitters always have been strongly inclined to work on large mammals, and earlier workers were weakly inclined to work on large mammals, but earlier and later workers were equally likely to be splitters. All this is consistent with the important biological pattern: large-mammal names tend to be of lower underlying quality regardless of when they were proposed (Fig. 2).

Variation among workers in current taxonomic quality is even less predictable. Because both body mass and current quality fall through historical time, and because there is a clear historical cross-correlation between the 2, workers proposing small species also should have poorly evaluated names (a good example being C. W. Hibbard), whereas large-mammal workers should have well-evaluated names (examples include the oreodont specialists C. H. Falkenbach, C. B. Schultz, and M. R. Thorpe). However, the among-worker correlation between quality and average body mass turns out to be vanishingly weak ($n = 34$, $r_s = +0.086$, $t = 0.488$, NS). The pattern surely is real, at least for this restricted set of 34 authors: the most inadequately re-evaluated names actually were proposed by C. Frick, who mainly studied artiodactyls, and E. M. Schlaikjer, who also tended to favor large mammals. Meanwhile, G. L. Jepsen, G. G. Simpson, and R. W. Wilson tended to work on small- or medium-sized species, but their names are well evaluated and reliable.

A possible explanation of this surprising result is that many of the least-evaluated names have been proposed by living small-mammal specialists who simply are too young to be included in the among-worker

TABLE 1.—Body mass statistics for major phylogenetic groupings of fossil mammals. Clades are listed only if they include ≥ 40 named species (with or without body mass estimates). Most species are grouped into orders, but 3 diverse families of “condylarths” (Arctocyonidae, Hyopsodontidae, and Periptychidae) are treated separately because the group as a whole is likely to be paraphyletic (Archibald 1998). Valid species in this table are currently accepted and have body mass estimates; invalid species are the ones currently treated as junior synonyms of species with body mass estimates.

Clade	Valid species			Invalid species		
	<i>n</i>	\bar{X} (ln g)	<i>SD</i>	<i>n</i>	\bar{X} (ln g)	<i>SD</i>
Arctocyonidae	44	8.04	1.17	31	8.14	1.19
Artiodactyla	215	10.28	1.87	237	10.84	1.31
Carnivora	228	9.32	1.91	124	10.02	1.84
Creodonta	43	8.51	1.41	22	8.63	1.19
Dinocerata	3	10.68	0.31	3	10.69	0.30
Hyopsodontidae	30	5.85	0.99	11	6.55	0.71
Insectivora	141	3.48	1.25	20	3.84	0.91
Lagomorpha	43	4.93	0.61	4	5.64	0.53
Marsupialia	41	3.15	1.09	18	3.67	1.69
Pantodonta	15	12.51	0.82	19	13.06	0.60
Periptychidae	26	7.05	1.48	20	8.34	1.74
Perissodactyla	178	11.54	1.63	174	11.73	1.62
Plesiadapiformes	67	5.65	1.30	15	6.73	0.55
Primates	73	6.29	1.11	22	6.97	1.14
Proboscidea	13	15.10	0.33	45	15.24	0.38
Rodentia	515	4.44	1.85	86	5.68	2.32
Xenarthra	8	11.55	1.16	15	12.62	0.34

analysis. Examples include R. C. Fox, P. D. Gingerich, W. W. Korth, and J. E. Storer, who respectively are credited with 37.5, 70.5, 85.5, and 40 species averaging 101, 740, 52, and 54 g (half credit is given for coauthored species). As will be shown in another paper, names coined by living workers do appear to be responsible for much of the substantial drop in current taxonomic quality over the last few decades. Thus, if it had been possible to determine the reliability of currently active workers, the statistics might have shown an among-worker correlation of name reliability and body mass.

TAXONOMIC QUALITY OF MAJOR GROUPS

The large differences among body size classes suggest the possibility of equally large differences among major phylogenetic groupings such as orders. Indeed, the more diverse orders do occupy very distinct parts of the size gradient; large- and small-mam-

mal orders such as Artiodactyla and Rodentia may differ in average body mass by >6 ln g units and within-order standard deviations are never >1.9 ln g units (Table 1).

More to the point, separate flux ratio analyses of the major orders do suggest very different levels of current validity, underlying quality, and current quality (Table 2). A large majority of rodent, lagomorph, and insectivoran names currently are accepted, whereas the same proportion is much lower for the ungulate orders Artiodactyla and Perissodactyla (mean body mass of valid species versus current acceptance: $n = 17$, $r_s = -0.919$, $t = 9.035$, $P < 0.001$). Underlying quality also correlates with body mass ($r_s = -0.718$, $t = 3.997$, $P < 0.001$), so the 2 factors cancel out and current taxonomic inflation does not ($r_s = -0.216$, $t = 0.855$, NS).

However, removing 2 outlying values for minor groups (Dinocerata and Xenarthra) greatly increases the mass-inflation corre-

TABLE 2.—Taxonomic quality and body mass of major phylogenetic groupings of mammals. Clades are as in Table 1. Total = total number of named species; now accepted = number of currently accepted species; proportion = proportion of species accepted; flux ratio = estimated underlying proportion of valid names; current quality = ratio of flux ratio to accepted proportion; predicted proportion = predicted accepted proportion based on body mass analysis; predicted flux = predicted flux ratio based on body mass analysis. Asterisks indicate probability level.

Clade	Total	Now accepted	Proportion	Flux ratio	Current quality	Predicted proportion	Predicted flux
Arctocyonidae	116	78	0.672	0.477	0.709	0.658	0.485
Artiodactyla	929	517	0.557	0.423	0.761	0.555	0.426
Carnivora	551	392	0.711	0.567	0.797	0.591***	0.445***
Creodonta	108	68	0.630	0.387	0.615	0.622	0.459
Dinocerata	44	8	0.182	0.029	0.160	0.518***	0.435***
Hyopsodontidae	54	42	0.778	0.563	0.724	0.770	0.544
Insectivora	237	213	0.899	0.648	0.721	0.887	0.664
Lagomorpha	88	77	0.875	0.527	0.602	0.862	0.662**
Marsupialia	94	73	0.777	0.516	0.664	0.886**	0.658**
Pantodonta	47	25	0.532	0.463	0.870	0.460	0.409
Periptychidae	55	35	0.636	0.397	0.624	0.683	0.453
Perissodactyla	659	355	0.539	0.482	0.894	0.477**	0.382***
Plesiadapiformes	114	95	0.833	0.652	0.782	0.799	0.561*
Primates	122	93	0.762	0.580	0.761	0.773	0.541
Proboscidea	102	37	0.363	0.315	0.867	0.357	0.161***
Rodentia	949	838	0.883	0.715	0.810	0.849**	0.629***
Xenarthra	62	29	0.468	0.207	0.442	0.481	0.413***

* $P < 0.05$, based on G -test of goodness of fit, ** $P < 0.01$, *** $P < 0.001$.

lation ($r_s = -0.511$, $t = 2.142$, $P < 0.1$). Indeed, although inflation is modest for the most important small-mammal group, the Rodentia, it is quite high for many of the others, such as Insectivora, Lagomorpha, and Marsupialia. Medium-sized groups also have enjoyed forgiving treatment, most notably the creodonts and “condylarths.” Hence, there does seem to be a broad-brush relationship between mass and taxonomic inflation for the major groups.

The extreme example of the mass–inflation trade-off is the order Perissodactyla; names in this large-mammal group appear to be almost fully reevaluated. However, this statistic seems to hide a large and virtually unaddressed problem with the species-level nomenclature of brontotheres (Mader 1998). Perhaps it is simply difficult to quantify the group’s invalidation and revalidation rates in the absence of an open ongoing debate.

All this variation would be of little interest if it could be attributed entirely to the

body size effect. Indeed, the steep decline in current validity with increasing body mass (Figs. 1 and 2) surely is responsible for the fact that in almost every order valid species have substantially lower average body masses than synonyms in the same groups (Table 1). The body mass hypothesis is tested here in a straightforward way: the observed current valid proportions and flux ratio–based underlying proportions for each 1.0–ln g unit bin are used as estimated values for each species, and the estimates then are summed and averaged across all species in each taxonomic group to yield predicted current validity proportions and flux ratios for the groups.

The body mass–based predictions are remarkably good, fully succeeding for most of the taxonomic groups (Table 2). However, there are some important exceptions. By either measure, taxonomic quality is significantly better than expected for 3 orders (Carnivora, Perissodactyla, and Rodentia) of the 5 most diverse, all 5 of which are

extant and have dominated the continent's diversity for the last 40 million years (Alroy 2000c). By contrast, observed flux ratios are worse than expected for 1 small-mammal group (Lagomorpha) and 2 large-mammal groups whose Pleistocene nomenclature is widely known to be oversplit (Proboscidea and Xenarthra). The only other exception is that marsupial names are somewhat less reliable by either measure than expected.

These generally optimistic results should be tempered. The unusually robust statistics for perissodactyls might be an artifact of the previously mentioned brontothere problem, and although rodents are so diverse that the differences are statistically significant, these differences are unimpressive in absolute terms. On the other hand, the high reliability of carnivoran names probably is real and relates to recent important revisions of canids by Wang (1994) and Wang et al. (1999); canids were the dominant group of North American carnivorans for much of the Tertiary. It also is noteworthy that names of primates and primate-like pliesiadapiformes are not particularly robust, even though these groups seemingly have been paid a disproportionate amount of attention.

CONCLUSIONS

Mammals are blessed with an extraordinarily dense and well-studied record that has served as a keystone for evolutionary theory ever since the time of G. Cuvier. However, much work remains to be done. Analyses in this article call attention to groups such as lagomorphs, condylarths, and Paleogene mammals in general that deserve more work, while showing that (say) carnivorans are in relatively good shape.

Most of these patterns are related to the key fact that taxonomic quality varies greatly with body size (Fig. 2). Large mammals are grossly oversplit, but historically older and therefore far better studied, so current counts of species in large-size categories seem to be reasonably accurate. However, species names of small mammals are highly

unreliable, so the apparent peak in species richness in the size range of rodents appears to be inflated. This is true even though specialists on small mammals propose more robust names (Fig. 5), because many small-mammal specialists are still living and few of them were working before the mid-20th century, so their names simply have not been around long enough to be evaluated. Indeed, small-mammal descriptions continue at such a rapid pace that the average size of all recognized species has been in a near free fall for decades (Fig. 3).

These patterns have important implications for macroevolutionary and macroecological hypotheses concerning body mass. Some results are unlikely to be affected. For example, the fact that descendant species are consistently larger than ancestral species could hardly be explained away by a proliferation of small-mammal names, because reevaluation of large species is relatively thorough (Fig. 2) and ancestor-descendant differences are only substantial for medium- and large-sized species (Alroy 1998a). Likewise, qualitative trends through the Cenozoic in the mean, standard deviation, skewness, and kurtosis of body mass (Alroy 1998a, 1999, 2000c) are unlikely to be artifactual: the pronounced bimodality of the distribution over the last 40+ million years is hardly a function of the lower mode being inflated, and removing small-mammal species names but retaining large-mammal names would only exaggerate the increase in mean body mass going into the Neogene.

One pattern that does seem likely to change is the overall size distribution (Fig. 1). At 1st glance, it seems to be strongly skewed but basically unimodal and therefore loosely consistent with the hypothesis of a single, 100-g body mass optimum for mammals (Brown et al. 1993). However, correcting the data using flux ratios shows that the distribution actually is bimodal, which is supported by earlier results concerning evolutionary dynamics (Alroy 1998a) and the broad-scale temporal trajec-

tory of body mass evolution (Alroy 1998a, 2000c). Thus, taking taxonomic error into account merely strengthens the case for 2 separate body mass optima separated by 3 orders of magnitude.

Although seemingly arcane, the methodological problems attacked by this article do suggest possible new directions in the study of Recent organisms. Methods of quantifying taxonomic error are not specific to paleobiology and could be applied easily to Recent mammals—if more details could be added to the current picture of the history of taxonomy (Alroy 2002) and, indeed, if more primary taxonomic research could be funded. This kind of reassessment might substantially change our view of major biological patterns, such as the latitudinal diversity gradient (Badgley and Fox 2000; Simpson 1964), the differences between clades in diversification rates (Purvis et al. 1995), the relationship between body size and diversity (Gardezi and da Silva 1999; Gittleman and Purvis 1998), or the body mass distributions of continental mammal faunas (Brown and Nicoletto 1991; Chown and Gaston 1997; Marquet and Cofré 1999). There never has been a better time to take up the study of taxonomic diversity in extant and extinct mammals.

ACKNOWLEDGMENTS

I thank an anonymous reviewer for helpful comments. This work was conducted when I was a Center Associate at the National Center for Ecological Analysis and Synthesis, a Center funded by the National Science Foundation (DEB 0072909), the University of California, and University of California Santa Barbara. This paper's grammar and style are not the responsibility of the nominal author. This is *Paleobiology Database* publication #16.

LITERATURE CITED

- ADRAIN, J. M., AND S. R. WESTROP. 2000. An empirical assessment of taxic paleobiology. *Science* 289:110–112.
- ALROY, J. 1992. Conjunction among taxonomic distributions and the Miocene mammalian biochronology of the Great Plains. *Paleobiology* 18:326–343.
- ALROY, J. 1994. Appearance event ordination: a new biochronologic method. *Paleobiology* 20:191–207.
- ALROY, J. 1996. Constant extinction, constrained diversification, and uncoordinated stasis in North American mammals. *Palaeogeography, Palaeoclimatology, Palaeoecology* 127:285–312.
- 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. Pp. 232–287 in *Biodiversity dynamics: turnover of populations, taxa, and communities* (M. L. McKinney and J. A. Drake, eds.). 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. 2000a. Understanding the dynamics of evolutionary trends within evolving lineages. *Paleobiology* 26:319–329.
- ALROY, J. 2000b. Successive approximations of diversity curves: ten more years in the library. *Geology* 28:1023–1026.
- ALROY, J. 2000c. New methods for quantifying macroevolutionary patterns and processes. *Paleobiology* 26:707–733.
- ALROY, J. 2001. A multispecies overkill simulation of the end-Pleistocene megafaunal mass extinction. *Science* 292:1893–1896.
- ALROY, J. 2002. How many named species are valid? *Proceedings of the National Academy of Sciences* 99:3706–3711.
- ALROY, J., P. L. KOCH, AND J. C. ZACHOS. 2000. Global climate change and North American mammalian evolution. *Paleobiology* 26(4), supplement:259–288.
- ARCHIBALD, J. D. 1993. The importance of phylogenetic analysis for the assessment of species turnover: a case history of Paleocene mammals in North America. *Paleobiology* 19:1–27.
- ARCHIBALD, J. D. 1998. Archaic ungulates (“Condylarthra”). Pp. 292–331 in *Evolution of Tertiary mammals of North America. Terrestrial carnivores, ungulates, and ungulatelike mammals* (C. M. Janis, K. M. Scott, and L. L. Jacobs, eds.). Cambridge University Press, Cambridge, United Kingdom 1:1–691.
- BADGLEY, C., AND D. L. FOX. 2000. Ecological biogeography of North American mammals: species density and ecological structure in relation to environmental gradients. *Journal of Biogeography* 27:1437–1467.
- BARNOSKY, A. D. 2001. Distinguishing the effects of the Red Queen and Court Jester on Miocene mammal evolution in the northern Rocky Mountains. *Journal of Vertebrate Paleontology* 21:172–185.
- BROWN, J. H., P. A. MARQUET, AND M. L. TAPER. 1993. Evolution of body size: consequences of an energetic definition of fitness. *American Naturalist* 142:573–584.
- BROWN, J. H., AND P. F. NICOLETTO. 1991. Spatial scaling of species composition: body masses of North American mammals. *American Naturalist* 138:1478–1512.
- CHOWN, S. L., AND K. J. GASTON. 1997. The species-body size distribution: energy, fitness and optimality. *Functional Ecology* 11:365–375.
- CLYDE, W. C., AND P. D. GINGERICH. 1998. Mammalian community response to the latest Paleocene thermal

- maximum: an isotaphonomic study in the northern Bighorn Basin, Wyoming. *Geology* 26:1011–1014.
- FOOTE, M. 1999. Morphological diversity in the evolutionary radiation of Paleozoic and post-Paleozoic crinoids. *Paleobiology* 25(2), supplement:1–115.
- FOOTE, M. 2000. Origination and extinction components of taxonomic diversity: general problems. *Paleobiology* 26(4), supplement:74–102.
- FOOTE, M., J. P. HUNTER, C. M. JANIS, AND J. J. SEPKOSKI, JR. 1999. Evolutionary and preservational constraints on origins of biologic group: divergence times of eutherian mammals. *Science* 283:1310–1314.
- GARDEZI, T., AND J. DA SILVA. 1999. Diversity in relation to body size in mammals: a comparative study. *American Naturalist* 153:110–123.
- GINGERICH, P. D. 1980. History of early Cenozoic vertebrate paleontology in the Bighorn Basin. *University of Michigan Papers on Paleontology* 24:7–24.
- GINGERICH, P. D. 1989. New earliest Wasatchian mammalian fauna from the Eocene of north-western Wyoming. *University of Michigan Papers on Paleontology* 28:1–97.
- GITTLEMAN, J. L., AND A. PURVIS. 1998. Body size and species richness in carnivores and primates. *Proceedings of the Royal Society of London, B. Biological Sciences* 265:113–119.
- GUNNELL, G. F., M. E. MORGAN, M. C. MAAS, AND P. D. GINGERICH. 1995. Comparative paleoecology of Paleogene and Neogene mammalian faunas: trophic structure and composition. *Palaeogeography, Palaeoclimatology, Palaeoecology* 115:265–286.
- HIBBARD, C. W. 1975. Letter from C. W. Hibbard to W. G. Kühne. *University of Michigan Papers on Paleontology* 12:135–138.
- HUNTER, J. P., AND J. JERNVALL. 1995. The hypocone as a key innovation in mammalian evolution. *Proceedings of the National Academy of Sciences* 92:10718–10722.
- JANIS, C. M., J. DAMUTH, AND J. M. THEODOR. 2000. Miocene ungulates and terrestrial primary production: where have all the browsers gone? *Proceedings of the National Academy of Sciences* 97:7899–7904.
- JANIS, C. M., AND P. B. WILHELM. 1993. Were there mammalian pursuit predators in the Tertiary? Dances with wolf avatars. *Journal of Mammalian Evolution* 1:103–125.
- JERNVALL, J., J. P. HUNTER, AND M. FORTELIUS. 1996. Molar tooth diversity, disparity, and ecology in Cenozoic ungulate radiations. *Science* 274:1489–1492.
- JOHST, K., AND R. BRANDL. 1997. Body size and extinction risk in a stochastic environment. *Oikos* 78:612–617.
- LILLEGRAVEN, J. A. 1972. Ordinal and familial diversity of Cenozoic mammals. *Taxon* 21:261–274.
- MAAS, M. C., M. R. L. ANTHONY, P. D. GINGERICH, G. F. GUNNELL, AND D. W. KRAUSE. 1995. Mammalian generic diversity and turnover in the late Paleocene and early Eocene of the Bighorn and Crazy Mountain Basins, Wyoming and Montana (USA). *Palaeogeography, Palaeoclimatology, Palaeoecology* 115:181–207.
- MACFADDEN, B. J. 1986. Fossil horses from “Eohippus” (*Hyracotherium*) to *Equus*: scaling, Cope’s Law, and the evolution of body size. *Paleobiology* 12:355–369.
- MADER, B. J. 1998. Brontotheriidae. Pp. 525–536 in *Evolution of Tertiary mammals of North America* (C. M. Janis, K. M. Scott, and L. L. Jacobs, eds.). Cambridge University Press, Cambridge, United Kingdom 1:1–691.
- MARQUET, P. A., AND H. COFRÉ. 1999. Large temporal and spatial scales in the structure of mammalian assemblages in South America: a macroecological approach. *Oikos* 85:299–309.
- MARQUET, P. A., AND M. L. TAPER. 1998. On size and area: patterns of mammalian body size extremes across landmasses. *Evolutionary Ecology* 12:127–139.
- MARTIN, P. S. 1967. Prehistoric overkill: the global model. Pp. 75–120 in *Pleistocene extinctions: the search for a cause* (P. S. Martin and H. E. Wright, eds.). Yale University Press, New Haven, Connecticut.
- MCSHEA, D. W. 1994. Mechanisms of large-scale evolutionary trends. *Evolution* 48:1747–1763.
- NIKLAS, K. J., AND B. H. TIFFNEY. 1994. The quantification of plant biodiversity through time. *Philosophical Transactions of the Royal Society of London, B. Biological Sciences* 345:35–44.
- PROTHERO, D. R. 1999. Does climatic change drive mammalian evolution? *GSA Today* 9:1–7.
- PROTHERO, D. R., AND T. H. HEATON. 1996. Faunal stability during the early Oligocene climatic crash. *Palaeogeography, Palaeoclimatology, Palaeoecology* 127:257–283.
- PURVIS, A., S. NEE, AND P. H. HARVEY. 1995. Macroevolutionary inferences from primate phylogeny. *Proceedings of the Royal Society of London, B. Biological Sciences* 260:329–333.
- RAUP, D. M. 1972. Taxonomic diversity during the Phanerozoic. *Science* 177:1065–1071.
- SIMPSON, G. G. 1953. *The major features of evolution*. Columbia University Press, New York.
- SIMPSON, G. G. 1964. Species density of North American recent mammals. *Systematic Zoology* 13:57–73.
- STANLEY, S. M. 1973. An explanation for Cope’s rule. *Evolution* 27:1–26.
- STUCKY, R. K. 1990. Evolution of land mammal diversity in North America during the Cenozoic. Pp. 375–432 in *Current mammalogy* (H. H. Genoways, ed.). Plenum Press, Mount Kisco, New York 2:1–596.
- VAN VALKENBURGH, B. 1985. Locomotor diversity within past and present guilds of large predatory mammals. *Paleobiology* 11:406–428.
- VAN VALKENBURGH, B. 1988. Trophic diversity in past and present guilds of large predatory mammals. *Paleobiology* 14:155–173.
- VERMEIJ, G. J. 1999. The accumulation of taxonomic knowledge: the history of species descriptions of some predatory gastropods. *Malacologia* 41:147–150.
- WAGNER, P. J. 1995. Diversity patterns among early gastropods: contrasting taxonomic and phylogenetic descriptions. *Paleobiology* 21:410–439.
- WALLACE, D. R. 1999. The bonehunters’ revenge: dinosaurs, greed, and the greatest scientific feud of the

- Gilded Age. Houghton Mifflin, Boston, Massachusetts.
- WANG, X. 1994. Phylogenetic systematics of the Hesperocyoninae (Carnivora: Canidae). *Bulletin of the American Museum of Natural History* 221: 1–207.
- WANG, X., R. H. TEDFORD, AND B. E. TAYLOR. 1999. Phylogenetic systematics of the Borophaginae (Carnivora: Canidae). *Bulletin of the American Museum of Natural History* 243:1–391.
- WEBB, S. D. 1969. Extinction-origination equilibria in late Cenozoic land mammals of North America. *Evolution* 23:688–702.

Special Feature Editor was Barbara H. Blake.