

Supporting Information

Figueirido et al. 10.1073/pnas.1110246108

SI Text

Taxonomic Units Used in the Q-Mode Factor Analysis. The number of genera present in each time interval (samples) within taxonomic units at subfamily level (SFTUs, variables) was collected whenever possible. If subfamilies were not described, the number of genera within families was collected. A dagger indicates an extinct taxon. Data from Janis and Jacobs (1) with updates from Janis et al. (2).

ORDER CREODONTA:[†] Oxyaenidae (Tytthaeninae, Oxyaeninae, Palaenictinae, Machaeroidinae); Hyaenodontidae (Provirriniinae, Hyaenodontinae); Limnocyoniidae.

ORDER CARNIVORA: Miacoidea[†] (Viverravidae, Miacidae); Canidae (Hesperocyoninae,[†] Borophaginae,[†] Caninae); Amphicyoniidae[†] (Daphoeninae, Temnocyoninae, Amphicyoniinae); Ursidae (Amphicyonodontinae,[†] Hemicyoniinae,[†] Ursinae, Tremarctinae); Felidae (Felinae, Machairodontinae[†]); Nimravidae[†] (Nimravinae, Barbourfelinae); Procyonidae (Procyoninae); Ailuridae (Ailurinae; Simocyoninae[†]); Mustelidae (Mustelavinae,[†] Oligobuninae,[†] Leparctinae,[†] Melinae, Mustelinae); Mephitidae.

ORDER CONDYLARTHRA:[†] Arctocyonidae (Arctocyonidae *incertae sedis*; Arctocyoninae); Hyopsodontidae; Mioclaenidae (Mioclaeninae, Apheliscinae); Periptychidae (Periptychidae *incertae sedis*, Periptychinae, Anisonchinae); Phenacodontidae (Tetraclaenodontinae, Phenacodontinae, Meniscotheriinae); Mesonychidae (Triisodontinae, Hapalodectinae, Mesonychinae).

ORDER DINOCERATA:[†] Prodinoceratidae; Uintatheriidae.

ORDER PANTODONTA:[†] Pantolambdidae; Titanodeidae; Barylambdidae; Coryphodontidae.

ORDER TAENIODONTA:[†] Conoryctidae; Stylinodontidae.

ORDER TILLODONTIA:[†] Estonychidae; Trogosidae.

ORDER ARTIODACTYLA: Agriocheridae[†] (Agriocherinae, Protoreodontinae); Anthracotheriidae[†] (Anthracotheriinae; Bothriodontinae); Antilocapridae (Merycodontinae,[†] Antilocaprinae); Bovidae; Cervidae; Camelidae (Poebrotheriinae,[†] Stenomyliinae,[†] Floridatragulinae,[†] Miolabinae,[†] Protolabinae,[†] basal Camelinae,[†] Camelinae); Dichobunidae[†] (Diacodexinae, Leptochoerinae, Antiacodontinae, Homacodontinae, Helohyinae); Dromomerycidae[†] (Aletomerycinae, Dromomerycinae, Cranio-ceratinae); Entelodontidae[†]; Hypertragulidae[†]; Leptomerycidae[†]; Gelocidae[†]; Moschidae (Blastomerycinae[†]); Oromerycidae[†]; Protoceratidae[†] (Leptotragulinae, Protoceratinae, Synthetoceratinae); Tayassuidae (Hesperhyinae, Tayassuinae); Merycoidodontidae[†] (Aclistomyerinae, Brachyrucinae, Eporeodontinae, Leptaucheninae, Merychinae, Merychochoerinae, Merycoidodontinae, Ticholeptinae, Ustatochoerinae, Merycoidodontinae *incertae sedis*).

ORDER PERISSODACTYLA: Arynodontidae[†]; Brontotheriidae[†] (Eotitanopinae, Palaeosynopinae, Dolichorhininae, Brontotheriinae, Brontotheridae *incertae sedis*); Chalicotheroidea[†] (Eomoropidae, Chalicotheriidae); Hyracodontidae[†]; Rhinocerotidae (basal rhinocerotids,[†] Diceratheriinae,[†] Menocerotinae,[†] Aceratheriinae,[†] Teleoceratinae[†]); Tapiroidea (Isectolophidae,[†] basal ceratomorphs,[†] basal rhinocerotoids,[†] Helaletidae,[†] Tapiiridae); Equidae (Hyracotheriinae,[†] Anchitheriinae,[†] Equinae).

ORDER PROBOSCIDEA: Mammutidae[†]; Gomphotheriidae[†]; Elephantidae.

ORDER XENARTHRA: Glyptodontidae[†]; Pamphathiidae[†]; Dasypodidae, Megalonychidae,[†] Megatheriidae,[†] Mylodontidae.[†]

Time Intervals Used in the Q-Mode Factor Analysis. Time intervals (North American land mammal ages) described for the Cenozoic

of North America following Woodburne (3). The duration of each time interval is provided between parentheses in million years before present. The number of biochrons is 62, with an average duration of 1.05 Ma and a standard deviation of 0.94 Ma.

PALEOCENE [*Puercan*: Pu1 (0.30), Pu2 (0.40), Pu3 (0.40); *Torresonian*: To1 (1.50), To2 (1.00), To3 (0.50); *Tiffanian*: Ti1 (0.90), Ti2 (1.10), Ti3 (1.10), Ti4 (0.40), Ti5 (1.00), Ti6 (0.40); *Clarkforkian*: Cf1 (0.20), Cf2 (0.40), Cf3 (0.40)].

EOCENE [*Wasatchian*: Wa0 (0.10), Wa1 (0.20), Wa2 (0.40), Wa3 (0.20), Wa4 (0.60), Wa5 (0.50), Wa6 (0.60), Wa7 (2.30); *Bridgerian*: Br0 (0.10), Br1 (0.80), Br2(2.00), Br3 (0.90); *Uintan*: Ui1 (0.60), Ui2 (2.90), Ui3 (2.70); *Duchesnean*: Du (3.20); *Chadronian*: Ch1 (0.30), Ch2 (0.80), Ch3 (1.00), Ch4 (1.10)].

OLIGOCENE [*Orellan*: Or1 (0.20), Or2 (0.50), Or3 (0.40), Or4 (0.60); *Whitneyan*: Wh1 (0.60), Wh2 (1.40); *Arikareean*: Ar1 (2.00), Ar2 (5.00)].

MIOCENE [*Arikareean*: Ar3 (3.60), Ar4 (0.60); *Hemingfordian*: He1 (1.30), He2 (1.50); *Barstovian*: Ba1 (1.20), Ba2 (2.20); *Clarendonian*: Cl1 (0.50), Cl2 (2.00), Cl3 (1.10); *Hemphillian*: Hh1 (1.50), Hh2 (0.80), Hh3 (0.90), Hh4 (1.00)].

PLIOCENE [*Blancan*: EBI (2.30), LBI (0.60)].

PLEISTOCENE [*Irvingtonian*: Irv I (1.05), Irv II (0.45), Irv III (0.25); *Rancholabrean*: RBL (0.10)].

Time intervals used in the Q-mode factor analysis resulted from grouping substages to create time bins of durations as similar as possible. The length of each time interval in million years before present is provided between parentheses. The total number of intervals was 26, with an average duration of 2.5 Ma and a standard deviation of 0.68 Ma.

PALEOCENE [PU + TO1 (2.6), TO2 + TO3 (1.5), TI1 + TI2 (2), TI3 + TI4 (1.4), TI5 + TI6 + CF (2.4)].

EOCENE [WA0-WA6 (2.6), WA7-BR0 (2.4), BR1 + BR2 (3), BR3 + UI1 (1.3), UI2 (2.9), UI3 (2.7), DU (3.2), CH (3.2)].

OLIGOCENE [OR (2.8), WH (2), AR1 (2), AR2 (5)].

MIOCENE [AR3 (3.6), AR4 + HE1 (1.9), HE2 + BA1 (2.7), BA2 + CL1 (2.7), CL2 + CL3 (3.1), HH1 + HH2 (2.3), HH3 + HH4 (1.9)].

PLIOCENE [EBI + LBI (2.9)].

PLEISTOCENE [IRV + RBL (1.85)].

Results Obtained in the Q-Mode Factor Analysis Performed Using the Original Timetable of Woodburne. Factor analysis performed using the 62 time intervals of the original time table of Woodburne (3) provided nine eigenvectors or factors with eigenvalues greater than one (Fig. S1A and B). However, the last three factors only explained minor faunal variation within other factors (Fig. S1B), on the one hand, and a distinct break in the slope between the first six factors and the last three ones was observed when the eigenvalues were represented against their ranks (Fig. S1A), on the other. A new factor analysis (FA) was performed extracting only six eigenvectors (Fig. S1C), which provided fairly similar results to the ones obtained from the modified timetable (Fig. 1B). This fact means that the pattern obtained from FA is robust and is not biased by unevenness in the duration of the temporal bins.

Evaluating the Completeness of the Mammalian Fossil Record. The age of fossils is relevant for estimating preservation completeness, because outcrop areas of older sedimentary rocks may be smaller than those of younger sediments overlying them, which results in lower sampling efficiency. Higher diversity is expected to be recorded when there is a greater amount of rock preserved

at outcrop simply because more fossil collections can be made by paleontologists: For example, Peters and Heim (4) estimated that Cenozoic time intervals have an average geological completeness that is approximately 40% greater than mean Paleozoic completeness in North America. For this reason, Cenozoic fossils have been more intensively studied than earlier ones, which explains in part the apparent trend of increase of diversity in marine animals during the Phanerozoic (5–7). However, despite the increasing knowledge of the Phanerozoic fossil record during the last few decades, with additions and deletions of families as well as improvements in low-resolution stratigraphic data, a comparison between compilations of times of origination and extinction of marine animal families published with ten years of difference showed that macroevolutionary patterns for the entire marine fauna remained essentially constant and that all major events of radiation and extinction were identical (8). Therefore, this study demonstrated that completeness bias in large paleontological databases and arbitrariness of included taxa are not necessarily impediments to the analysis of patterns of diversification and extinction in the fossil record.

Although compared to the diversity of marine animals the fossil record of continental tetrapods has a lower degree of preservational completeness, relative proportions of major taxa and familial durations suggest that the family-level continental fossil record of Europe and North America is reasonably representative (9). This situation seems to be the case of the extensive dataset used here for studying major patterns of diversity in North American Cenozoic mammals (1, 2), compiled by experts in each of the mammalian orders and covering a huge number of paleontological localities: In this compilation, 93.89% (123/131) of mammalian families have a continuous stratigraphic record (i.e., no Lazarus genera) and the mean number of paleontological localities sampled per stratigraphic interval is $47.76 + 24.49$, with a minimum of 14 (Tiffanian 1 + 2) and a maximum of 102 (Barstovian 2 + Clarendonian 1).

Fig. S2A shows the trends in the number of large mammals genera recorded through the North American Cenozoic and in the number of paleontological localities available for the stratigraphic intervals analyzed in this study, which is used here as a proxy of sampling efficiency (and thus a measure of observational completeness). As expected, both variables show a significant correlation with time ($r^2 = 0.593$ and 0.592 , respectively, $p < 0.00001$ in both cases), which suggests a temporal trend in observational completeness for the North American Cenozoic mammalian record. In fact, a least-squares regression of genus diversity on number of localities (Fig. S2B) shows a statistically very significant relationship between both variables ($r^2 = 0.720$, $p < 0.00001$), which indicates that the number of genera recorded per stratigraphic interval is in part a reflection of the number of fossil localities studied. However, it is worth noting that the residuals of this regression line (i.e., number of recorded genera minus number of expected genera) do not show a statistical relationship with the age of the intervals ($r^2 = 0.128$, $p > 0.075$). This result indicates that the increasing trend in diversity of North American mammals during the Cenozoic cannot be attributed to an increasing degree of observational completeness that could bias the major evolutionary patterns of these faunas (Fig. S2C). In fact, there are a number of intervals during both the Paleogene (e.g., Wasatchian 0–6 and Orellan) and the Neogene (e.g., Clarendonian 2–3 and Blancan) that show diversities significantly lower than those merely expected from sampling efficiency and, similarly, several intervals show numbers of genera in excess of those expected (e.g., Arikareean 3 in the Neogene and Uintan 3 in the Paleogene). In addition, Fig. S2D and E show that there is no significant relationship between the number of localities and the duration of stratigraphic intervals ($r^2 = 0.019$, $p > 0.5$) and that this lack of significance also applies

to the number of genera ($r^2 = 0.035$, $p > 0.36$). If we also take into account that there is no temporal trend in the duration of stratigraphic intervals ($r^2 = 0.038$, $p > 0.33$), this evidence implies that the diversity of large mammals recorded is not affected by major biases in sampling efficiency and/or stratigraphic resolution.

In sum, these results show that although there is an unavoidable bias in the degree of knowledge on the North American mammalian fossil record, which relates to differences in sampling effort and observational completeness, there is no systematic trend in completeness through the Cenozoic that could be masking the results obtained on the patterns of origination, expansion and decline of the major evolutionary faunas of large mammals.

The Method of Generalized Differences. In this section we were interested in testing if major climatic changes through the Cenozoic influenced the pattern of mammalian biodiversity dynamics. For doing so, we regressed the positive factor loadings (PFLs) of each evolutionary fauna and the stacked deep-sea benthic foraminiferal oxygen isotopic values ($\delta^{18}\text{O}$) from Zachos et al. (10). We used the PFLs of each fauna for estimating its diversity dynamics because the intervals with positive loadings reflect the fauna's timespan. In contrast, negative PFLs evidence the time intervals where the faunas were not present. Thus, the PFLs of each fauna was used as a proxy for its paleodiversity. Accordingly, average values of $\delta^{18}\text{O}$ were used as a proxy of climate for the intervals where the faunas showed positive loadings. In the same way, the number of genera (NG) present in each stratigraphic interval of each fauna (i.e., the diversity curve of each fauna) was regressed against the average values of $\delta^{18}\text{O}$ for the timespan of each fauna. For doing this test, we first selected those mammalian SFTUs scoring >1.0 in each fauna and then we summed their diversity curves. Later, we regressed the diversity curve of each fauna against the average values of $\delta^{18}\text{O}$ for their respective time intervals.

However, the fact that the successive observations in a time series variable usually depend to some extent on the preceding values (i.e., they used to be serially correlated) precluded us from directly performing an ordinary least-squares regression analysis between the PFLs of each fauna (i.e., a proxy for its diversity pattern) and the oxygen isotopic values ($\delta^{18}\text{O}$) for their corresponding time intervals (i.e., a proxy for paleotemperature). Therefore, prior to analyzing the data by an ordinary least-squares regression analysis, the serial correlation component of the data was removed. To perform this analysis we used the method of generalized differences (GDRA) developed by Wonnacott and Wonnacott (11) and applied to fossil data by McKinney and Oyen (12), which allows the extraction of the serially correlated components of both PFLs' values or the NG of each fauna and the $\delta^{18}\text{O}$ values of their time intervals. The following transformed values were regressed:

$$\Delta Y_t = Y_t - \rho Y_{t-1} \quad \Delta X_t = X_t - \rho X_{t-1},$$

where X is the midpoint of each time interval, Y is the PFL or the NG and $\delta^{18}\text{O}$ values of each fauna and ρ is the serial correlation coefficient. For the first two observations, we performed a special transformation following ref. 11:

$$\Delta Y_1 = (1 - \rho^2)^{0.5} (Y_1) \quad \Delta X_1 = (1 - \rho^2)^{0.5} (X_1).$$

The values of ρ were estimated as follows, following Wonnacott and Wonnacott (11):

$$Y_t = \alpha(1 - \rho) + \rho Y_{t-1} + \beta X_t - \beta \rho X_{t-1} + (e_t - \rho e_{t-1}).$$

