

Twenty-million-year relationship between mammalian diversity and primary productivity

Susanne A. Fritz^{a,b,1}, Jussi T. Eronen^{a,c,2}, Jan Schnitzler^{a,d}, Christian Hof^a, Christine M. Janis^{e,f}, Andreas Mulch^{a,g}, Katrin Böhning-Gaese^{a,b}, and Catherine H. Graham^{a,b,h}

^aSenckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberg Gesellschaft für Naturforschung, 60325 Frankfurt, Germany; ^bInstitute of Ecology, Evolution and Diversity, Goethe University, 60438 Frankfurt, Germany; ^cDepartment of Geosciences and Geography, University of Helsinki, 00014 Helsinki, Finland; ^dInstitute of Biology, Leipzig University, 04103 Leipzig, Germany; ^eDepartment of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912; ^fSchool of Earth Sciences, University of Bristol, Clifton, Bristol BS8 1RJ, United Kingdom; ^gInstitute of Geosciences, Goethe University, 60438 Frankfurt, Germany; and ^hDepartment of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794

Edited by David Jablonski, The University of Chicago, Chicago, IL, and approved August 2, 2016 (received for review February 9, 2016)

At global and regional scales, primary productivity strongly correlates with richness patterns of extant animals across space, suggesting that resource availability and climatic conditions drive patterns of diversity. However, the existence and consistency of such diversity–productivity relationships through geological history is unclear. Here we provide a comprehensive quantitative test of the diversity–productivity relationship for terrestrial large mammals through time across broad temporal and spatial scales. We combine >14,000 occurrences for 690 fossil genera through the Neogene (23–1.8 Mya) with regional estimates of primary productivity from fossil plant communities in North America and Europe. We show a significant positive diversity–productivity relationship through the 20-million-year record, providing evidence on unprecedented spatial and temporal scales that this relationship is a general pattern in the ecology and paleo-ecology of our planet. Further, we discover that genus richness today does not match the fossil relationship, suggesting that a combination of human impacts and Pleistocene climate variability has modified the 20-million-year ecological relationship by strongly reducing primary productivity and driving many mammalian species into decline or to extinction.

macroecology | paleontology | mammals | net primary production

One ubiquitous pattern in ecology is the positive relationship between the diversity of terrestrial organisms and primary productivity (1–4). For consumers, this relationship is thought to arise because primary productivity limits energy flow to and total biomass at higher trophic levels (5). Because primary productivity depends largely on climatic conditions (1), and spatial richness patterns of extant species are often strongly correlated with climate at global and continental scales, the productivity hypothesis has been successful in explaining spatial patterns of diversity (1–5). However, the present-day diversity–productivity relationship may not be representative for Earth’s history, because present-day conditions have been strongly shaped by human activity (6, 7). Exponential increases in human population size and in biomass of a few domesticated species, such as cattle, pigs, and poultry, have resulted in increasing appropriation of the net primary production of biomass (NPP) since the beginning of the Holocene (8, 9). Today, human activity removes up to 30% of the global NPP from natural ecosystems, mostly through harvesting, deforestation, and grazing (10). Increasing human impact and strong glacial–interglacial climate oscillations superimposed on Pleistocene environmental changes have dramatically reduced the number of extant large mammal species (7, 11). Here, we test the diversity–productivity relationship in large mammals by analyzing the Neogene fossil record, which precedes Pleistocene climate change and human dominance of natural ecosystems.

To date, the generality of the terrestrial diversity–productivity relationship over long geological timescales remains elusive. Although temporal changes in terrestrial fossil diversity have been linked to changing productivity and temperature (12–14), the few quantitative analyses to date have been performed at highly disparate spatial

scales, either global to continental or for single fossil locations (15–17). The evidence for terrestrial diversity–climate relationships from these studies is equivocal, calling into question the universality of the diversity–productivity relationship. Some of the discrepancies may arise because quantitative studies on large spatial scales have used global paleo-climate reconstructions based on marine records (13, 15, 17), which are unlikely to represent terrestrial climatic conditions adequately. Temperature also could be an indirect or secondary driver of terrestrial diversity, because present-day spatial diversity patterns are often better explained by combinations of proxy variables for energy and water availability than by temperature alone (2–4).

To evaluate the mammalian diversity–productivity relationship through the Neogene, we combine Northern hemisphere mammalian fossil data for stratigraphic stages covering the Miocene and Pliocene epochs ~23–1.8 Mya (Table S1) with regional terrestrial NPP estimates derived from fossil plant communities (18), covering 23–2.6 Mya in Europe and 17–2.6 Mya in North America (Fig. 1). Our mammalian dataset contains 14,083 fossil occurrence records for 690 genera (orders Artiodactyla, Carnivora, Perissodactyla, Primates, and Proboscidea) in 1,567 locations, divided into three North American and three European regions (Fig. 1) based on biogeographic history (12, 19). We focus on large

Significance

Our study links diversity dynamics of fossil large mammals through time to primary productivity, i.e. net production of plant biomass. Spatial diversity patterns of extant terrestrial animals are often correlated with present-day primary productivity, but it is unclear whether the relationship holds throughout the geological past. Here we show that higher primary productivity was consistently associated with higher mammalian diversity throughout the geological period of the Neogene, supporting the hypothesis that energy flow from plants to consumers is a key factor determining the level of biodiversity. Our comparison of the fossil diversity–productivity relationship with present-day data suggests that human activity and Pleistocene climate change have conspired to dissolve the relationship that has characterized our planet over 20 My.

Author contributions: S.A.F., J.T.E., J.S., C.H., C.M.J., A.M., K.B.-G., and C.H.G. designed research; S.A.F., J.T.E., J.S., C.H., and C.H.G. performed research; J.T.E. and C.M.J. contributed new reagents/analytic tools; S.A.F., J.T.E., J.S., and C.H. analyzed data; and S.A.F., J.T.E., J.S., C.H., C.M.J., A.M., K.B.-G., and C.H.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Data and R codes for analysis are available at dataportal-senckenberg.de/database/metacat/bikf.10018.1/bikf.

¹To whom correspondence should be addressed. Email: sfritz@senckenberg.de.

²Present address: BIOS Research Unit, 00510 Helsinki, Finland.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1602145113/-DCSupplemental.

terrestrial mammals because of their comparatively well-resolved taxonomy, their high preservation rates, and their well-sampled and comprehensive Neogene fossil record. To account for preservation and spatial sampling biases still present in the record, we estimate regional and continental mammalian γ diversity on the genus level with a first-order Jackknife approach (20) separately within each global stratigraphic stage and each continent-specific land mammal age or unit (Table S1). Terrestrial NPP is estimated from paleobotanical data accounting for temporal uncertainty, uncertainty of climatic reconstruction, and spatial structure. We evaluate the fossil relationship between NPP and γ diversity through time (*i*) across the two continents and (*ii*) across focal regions. Finally, we compare predictions from this fossil diversity–productivity relationship with observed present-day diversity and NPP to test whether the Neogene relationship has persisted into the present despite Pleistocene climate change and increasing human impact.

Results and Discussion

The temporal dynamics of mammalian γ diversity, i.e., of the estimated regional diversity of genera, differ strongly between the two continents and across our focal regions (Fig. 2 *A–H*). Miocene γ diversity peaked earlier in North America (stratigraphic stage Burdigalian) than in Europe (Tortonian), a difference that has been linked to earlier drying and cooling in North America (14, 21, 22). In our terrestrial plant datasets, North America shows consistently lower NPP than Europe in the Miocene but not in the Early Pliocene (Fig. 2 *I* and *J*). Because terrestrial NPP data are available only at the resolution of stratigraphic stages, we use the stage-level mammalian diversity estimates in the following analyses. These stage-level diversity estimates generally track estimates in the more finely resolved land mammal ages (Pearson’s correlation coefficients between diversity estimates for stages and diversity estimates for the contemporary land mammal ages: $r = 0.701$, $t = 5.29$, $df = 29$, $P < 0.001$ for continents, $r = 0.688$, $t = 8.31$, $df = 77$, $P < 0.001$ for regions) (Fig. 2), although diversity is elevated in long stratigraphic stages compared with the corresponding land mammal ages (e.g., Tortonian in Europe; Fig. 2*B*). Because diversity estimates may partly reflect temporal turnover of genera within a stratigraphic stage, we assess the effects of temporal resolution by repeating analyses with the diversity estimates in land mammal ages, averaged within each stage.

Our analyses show a significant positive relationship of fossil mammalian γ diversity with NPP across the continents and stratigraphic stages (Fig. 3*A* and Table 1). We fit generalized linear mixed-effects models (GLMMs) with Poisson-distributed errors and account for the temporal and spatial data structure

through random effects. Further, we account for covariates describing known effects on richness by fitting the area of the region or continent and the duration of the stratigraphic stage as fixed effects. Because of the relatively low availability of paleobotanical locations where the taxonomic composition has been analyzed and NPP could be inferred, the regional analyses are restricted to three focal regions with highest data coverage and best spatial and temporal match of mammalian and paleobotanical locations: Western North America and Western and Eastern Europe (Figs. 1 and 2). These regional analyses confirm the continental-scale results (Fig. 3*B* and Table 1). All patterns reported here are robust to the well-known limitations associated with the analysis of fossil data (6), because we find no or little effect (*SI Materials and Methods*) of range-through genera (Figs. S1 and S2), diversity estimator algorithm (Fig. S2), location definition (Fig. S3), and temporal resolution (for analyses using diversity in land mammal ages, see Table S2; for different methods of allocating paleobotanical data to stratigraphic stages, see Fig. 2 *I* and *J*). Supplemental simulations based on present-day data indicate that first-order Jackknife estimation performs well in the parameter space likely to be important for our high-quality mammalian fossil record (Fig. S4). We also estimate fossil NPP taking climatic uncertainty into account (Fig. S5) and validate the NPP model through comparisons with present-day data (Fig. S6).

Our results provide strong support for the hypothesis that the terrestrial diversity–productivity relationship is a general pattern in ecology and paleo-ecology that persists in time and in space, at least in the Neogene across the two continents analyzed here. Our study might reconcile previous large-scale paleontological studies, e.g., those reporting that North American diversity of mammals is not consistently related to global temperature through the Cenozoic (15), even though major transitions between evolutionary faunas match periods of climate change in the same region (17, 22). This inconsistency, combined with the significant diversity–productivity relationship found here, could suggest that primary productivity is a more important or more direct driver of terrestrial mammalian diversity than temperature (3), although we do not directly compare temperature and productivity effects. In addition, the plant records in our study recovered regional variation in terrestrial NPP that could not be captured by the single global temperature curve from marine isotope data used in previous work, even though the terrestrial records have patchy spatial coverage and lower temporal resolution (18).

Next, we visually compare whether present-day diversity and NPP estimates were in agreement with the Neogene diversity–productivity relationship. We observe that present-day genus richness of large mammals in North America and Europe falls

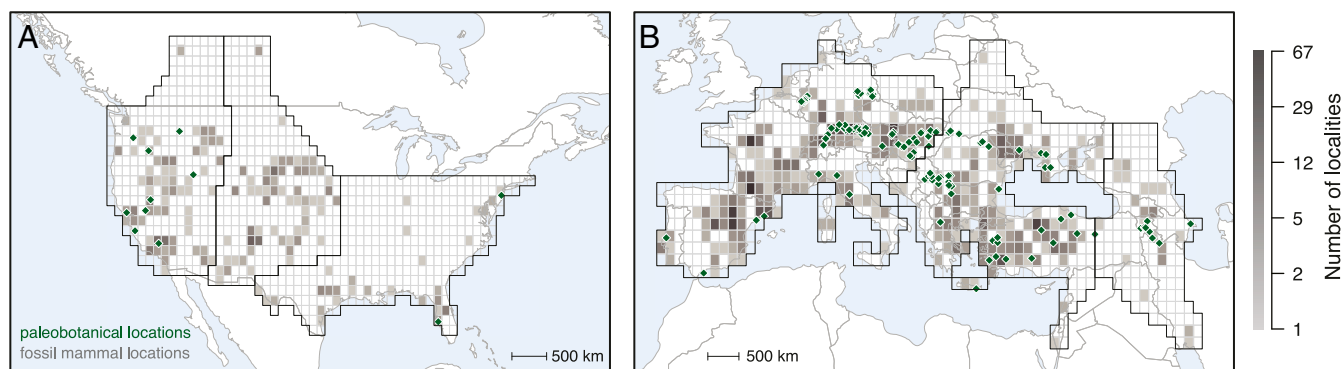


Fig. 1. Spatial coverage of the Neogene paleobotanical and mammalian fossil records in focal regions (black outlines) in (*A*) Western, Central, and Eastern North America and (*B*) Western, Eastern Europe, and the Caucasus. Based on 145 paleobotanical locations (green diamonds), we estimate the NPP across each continent and within each of the three best-covered regions (Western North America and Western and Eastern Europe). The coverage of 1,567 fossil locations for large terrestrial mammals is shown as the number of localities (unique combinations of spatial location and stratigraphic stage, gray shading) in 1° latitudinal–longitudinal grid cells.

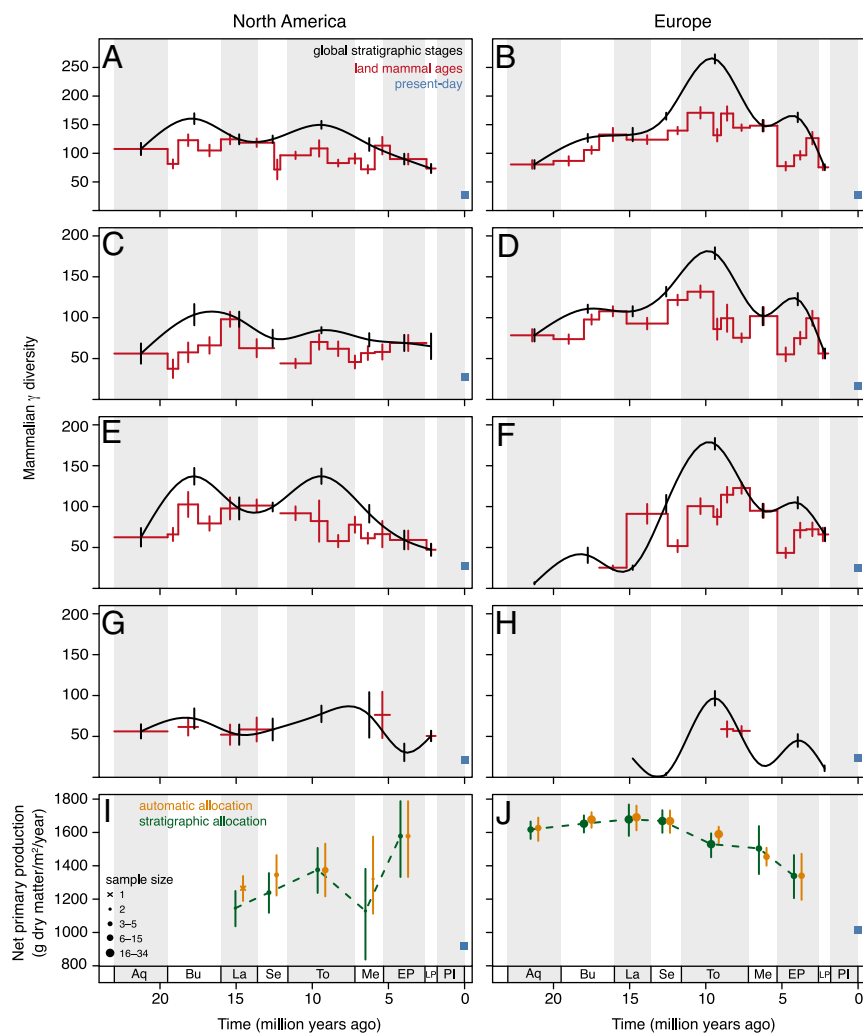


Fig. 2. Temporal dynamics of Neogene mammalian diversity (A–H) and net primary production (J–K) in North America and Europe. (A and J) North America continent-wide. (C) Western North America. (E) Central North America. (G) Eastern North America. (B and J) Europe continent-wide. (D) Western Europe. (F) Eastern Europe. (H) Caucasus. Patterns of γ diversity for large terrestrial mammals (genus-level, first-order Jackknife estimation) are largely consistent in global stratigraphic stages (black trend line shows natural cubic spline interpolation; vertical bars indicate SEs) and continent-specific land mammal ages (red stepped line and error bars). Only time intervals with more than five mammalian locations are shown. Present-day observed genus richness (blue squares) is markedly lower than fossil diversity. The fossil NPP estimates in the two continents within stratigraphic stages (orange and green symbols; symbol size indicates the number of grid cell values underlying the estimate; error bars indicate the entire range between average minimum and maximum values across the grid cells) were very similar with two approaches to allocating paleoclimatic estimates to stratigraphic stages, i.e., whether paleobotanical records were assigned automatically following absolute dates given in source datasets (orange) or were assigned manually according to stratigraphic information in source datasets (green). Neogene estimates were generally much higher than the present-day estimates (potential NPP, blue squares with SEs too small to see). Stratigraphic stages (see Table S1): Aquitanian (Aq); Burdigalian (Bu); Langhian (La); Serravallian (Se); Tortonian (To); Messinian (Me); Early Pliocene (EP); Late Pliocene (LP); Pleistocene (PI).

far below the predictions from the fossil relationship that has prevailed over 15–20 My (Fig. 3, gray symbols). Additionally, adjusting present-day values for both human appropriation of NPP (HANPP) (Fig. 3, blue symbols) and end-Pleistocene and Holocene extinctions (red symbols) would seem to reconcile present-day values with the fossil relationship, suggesting that increasing HANPP (8–10) and the end-Pleistocene and Holocene extinctions (11, 23) have impacted the temporal diversity–productivity relationship in large mammals since the end of the Neogene. However, conclusions from these comparisons must be taken cautiously. We could not fit a combined model across fossil and present-day data points because of substantial differences, particularly in the underlying timescale: The average stratigraphic stage in the Neogene lasted 2.6 My, whereas the present-day data are a snapshot of the last 10,000 y at most. The large differences in diversity and NPP between fossil and present-day data could result from this differing timescale and mean that present-day data must be compared with fossil model predictions that are made outside the range of diversity and NPP values ever recorded in the Neogene (dashed line in Fig. 3). Nevertheless, we observe that the differences between the fossil diversity–productivity relationship and the observed present-day data points are striking (Fig. 3) and might reflect a fundamental change in the diversity–productivity relationship that occurred between the Neogene and today.

If the diversity–productivity relationship has been changed since the Neogene, we would expect the present-day relationship in space

to be weakest in those regions most impacted by climatic oscillations and mammalian extinctions, such as North America and Europe. Across the globe, we find a significant present-day spatial relationship of mammalian diversity with terrestrial NPP (adjusted for human appropriation; see *SI Materials and Methods* and Fig. S7), in agreement with previous studies (2–4). In contrast, we show that the present-day spatial relationship within the focal regions Western North America, Western Europe, and Eastern Europe is much weaker (Fig. S7), as could be expected because of climatic and anthropogenic impacts since the end of the Pliocene. Presumably, increasing HANPP in these regions has prevented a recovery from the numerous mammalian declines and extinctions that occurred in the Pleistocene and Holocene and are ongoing (8, 9) and has changed the diversity–productivity relationship through time. These results could be specific to large mammals, because they have been most strongly affected by past extinctions and experience high extinction risk today (7, 24). The applicability of our fossil and present-day diversity–productivity relationships to small mammals is unclear, because small mammals may be less susceptible to climate oscillations and have experienced fewer end-Pleistocene and Holocene extinctions (7, 13). Future studies could test the prediction that the diversity–productivity relationship through time is consistent with present-day patterns in other taxa, including those less affected by climate oscillations and human impact.

Because of the large spatial and temporal scales of our diversity–productivity analysis, we cannot fully disentangle the ultimate underlying ecological and evolutionary mechanisms: Because

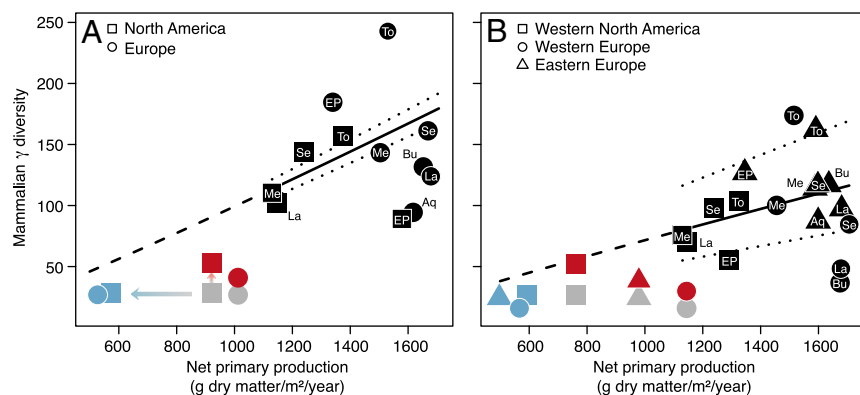


Fig. 3. Models of the fossil mammalian diversity-productivity relationship in (A) continents and (B) focal regions across stratigraphic stages in the Neogene (black symbols) and visual comparison with present-day data (gray and colored symbols). GLMMs (black continuous lines) account for temporal and spatial data structure with random effects (dotted lines) and show consistent effects of NPP on fossil γ diversity. Black symbols represent mean conditional response values for stratigraphic stages (as in Fig. 2) predicted for median values of the fixed-effect covariates (Table 1). Present-day observed data (gray symbols), data adjusted for human appropriation of NPP (blue symbols), and data adjusted for end-Pleistocene and Holocene extinctions (red symbols) fall below the fossil model predictions (dashed lines).

resources drive consumer abundances and biomass, productivity could directly limit the diversity that can exist in a given region or it could influence extinction and speciation processes (5, 25). It is clear from our fossil results that productivity is not the only factor influencing diversity and that mammalian diversity does not perfectly track productivity through time. In our Neogene models, the effects of area are stronger than the effects of productivity, and the duration of the stratigraphic stage is also a significant covariate in most models. Additionally, there is a surprisingly large amount of scatter in the global present-day diversity-productivity relationship (Fig. S7). Presumably, our Neogene relationship captures the large-scale temporal transition from tropical and subtropical wet environments to much drier and colder temperate systems today (14) rather than a fine-scale temporal correlation between diversity and productivity. Also, the variability in primary productivity might have a cumulative effect, so that regions with a stable paleo-climatic history accumulate high diversity over long timespans (26). Although we did not test this possibility explicitly, the weak spatial diversity-productivity relationship in our focal regions today in comparison with the stronger global spatial relationship could support this idea, because the focal regions were influenced by glaciations until relatively recently.

Conclusions

There has been increasing interest in reconciling paleontological and neontological perspectives on diversity (27), but this integration has been challenging because of the inherent differences in sampling, timescale, and taxonomy (6). Here we successfully use the fossil record to test an ecological pattern over geological timescales and pioneer large-scale quantitative analyses that directly link fossil occurrence datasets to terrestrial paleo-environmental proxy data. Our results suggest that general ecological rules cannot be inferred exclusively either from the geological past or from present-day data alone. Mammalian diversity and terrestrial primary production are currently much lower than over the last 23 My and seem to be inconsistent with the universal diversity-productivity relationship we find through the Neogene. This difference renders predictions of future diversity dynamics based on knowledge of past and present relationships more challenging than previously thought. In fact, accelerating human impacts strongly decrease the probability of a rebound of diversity (8, 9, 28), supporting the hypothesis that an irreversible anthropogenic state shift of the biosphere has already taken place (29).

Materials and Methods

Mammalian Fossil Data. We extracted geo-referenced and dated fossil species and genus occurrences of nonmarine members of the mammalian orders Artiodactyla, Carnivora, Perissodactyla, Primates, and Proboscidea throughout the Miocene and Pliocene for North America (30, 31) and for Eurasia (NOW, the New and Old Worlds Database of Fossil Mammals, www.helsinki.fi/science/now/).

Original data will be publicly available through the NOW database during 2016, and our cleaned datasets, processed data for analyses, and R scripts are available online (dataportal-senckenberg.de/database/metacat/bikf.10018.1/bikf). Fossil locations were only included if they could be unambiguously assigned to one time interval. The sources used two different chronologies (Table S1): the North American Land Mammal Ages (NALMA) (32) and the Mammal Neogene (MN) units (33). We evaluated mammalian diversity within these land mammal ages or units to gain a detailed view of temporal diversity dynamics, but we also combined occurrence data into a set of broader, global stratigraphic stages (34) (Table S1). Although these global stages were less well resolved in time, they were comparable across continents and matched the temporal resolution of the paleobotanical data.

We followed the taxonomy of our sources for fossil (NOW database and refs. 30 and 31) and extant species (35). The raw data were corrected on the species level for taxonomic errors, and we unified the taxonomy across the data sources (using a taxonomic look-up table available at dataportal-senckenberg.de/database/metacat/bikf.10018.1/bikf) to avoid biases in genus counts arising from synonyms. We performed all analyses at the genus level because the sampling bias inherent in the fossil record should be less

Table 1. Model results for the fossil mammalian diversity-productivity relationship

	Estimate	Std. error/std. dev.	z	P
Continents				
NPP	1.109	0.373	2.97	0.003
Area	0.535	0.081	6.59	<0.001
Stage duration	0.237	0.084	2.83	0.005
Continent identity	0.062	0.249		
Stage midpoint	0.001	0.026		
Regions				
NPP	0.912	0.318	2.87	0.004
Area	0.628	0.089	7.04	<0.001
Stage duration	0.251	0.083	3.03	0.002
Region identity	0.346	0.588		
Stage midpoint	0.004	0.062		

Models were GLMMs with Poisson error functions, fitted across stratigraphic stages in either the two continents or the three focal regions. The response variable was fossil mammalian diversity (γ diversity estimates rounded to integers). Variables fitted as fixed effects were NPP in grams of dry matter per square meter per year, log-transformed; area (minimum convex hull around all mammalian locations in each stage, in square kilometers log-transformed); and stage duration (My, log-transformed). Variables fitted as random effects were continent or region identity (random intercept) and stage midpoint (age in Mya, random slope). The random effects therefore accounted for temporal trends and spatial structure in the data (see *SI Materials and Methods* and Table S2). Model statistics were as follows: model across continents: $n = 12$ observations; residual $df = 5$; marginal R^2 (variance explained by the fixed effects) = 0.68; conditional R^2 (variance explained by entire model, i.e., both the fixed and random effects) = 0.89; Akaike Information Criterion (AIC) = 116.6; model across regions: $n = 17$; residual $df = 10$; marginal $R^2 = 0.39$; conditional $R^2 = 0.82$; AIC = 183.3. Std. dev., SD of the variance estimate for random effects; Std. error, SE of the slope estimate for fixed effects.

influential on diversity estimates calculated at higher taxonomic levels (6). Additionally, morphological disparity at the genus level in fossil mammals has been shown to approximate disparity at the species level in extant mammals (36, 37). The final dataset contained a total of 1,688 unique species in 663 genera, plus 27 genera for which we had only genus-level occurrences (the full dataset is available at dataportal-senckenberg.de/database/metacat/bikf.10018.1/bikf). We performed analyses at two spatial extents. Continental datasets of North America and Europe included all their respective locations. For regional analyses, focal regions defined based on existing knowledge of biogeographic history (12, 19) were small enough to capture biogeographically meaningful units but were large enough to contain a sufficient number of mammalian fossil locations within the stratigraphic stages. Fig. 1 shows the final regions delimited on a 1° latitudinal-longitudinal grid.

Estimation of Mammalian γ Diversity. The number of genera varied considerably across time intervals (Table S1) and was significantly correlated with the number of locations ($r = 0.83$, $t = 9.8$, $df = 45$, $P < 0.001$ across all time intervals for continents; $r = 0.76$, $t = 13.3$, $df = 130$, $P < 0.001$ for regions) (Fig. S1). We corrected for this sampling bias with algorithms to estimate γ diversity (i.e., the region- or continent-wide genus richness) based on the occurrence of genera (20, 38–40). We applied the richness estimators Chao, Jackknife, and Bootstrap (20, 40) to a genus-by-location matrix of presences and absences for each time interval within each focal region and each continent. From these matrices, we also calculated genus-level occupancy for each subset, i.e., the number of locations where a genus was present (36). We applied the site-specific, abundance-based richness estimators of Chao1 (the unbiased variant of the Chao estimator) and Abundance-Coverage Estimator (ACE) to these occupancy data, treating a region or continent as one site (20, 39). Analyses were performed in R with the vegan package (41, 42), and estimates based on fewer than six locations were excluded.

Values of γ diversity from different estimators were strongly correlated (Fig. S2 A–J), so we present results with first-order Jackknife here (see *SI Materials and Methods* for details of estimator selection and results with different estimators). One central issue is that diversity in the relatively long stratigraphic stages is likely to represent signals of both standing diversity and temporal turnover of genera within a stage. We were restricted to global stratigraphic stages for comparisons between the two continents and because terrestrial NPP data were available only at that temporal resolution. To assess the effect of temporal resolution directly, we repeated analyses with the diversity estimates in the more finely resolved land mammal ages, which were then averaged within stratigraphic stages. Further, we assessed key assumptions and the performance of diversity estimation in supplemental analyses and simulations (following refs. 36, 38, and 43); see *SI Materials and Methods* and Figs. S2–S4.

Estimation of Present-Day Diversity and End-Pleistocene and Holocene Extinctions. To estimate present-day γ diversity for the same five orders of large mammals, we extracted occurrences by overlaying species' range maps with our 1° grid (Fig. 1). We edited the range maps from the International Union for the Conservation of Nature (IUCN) Red List Global Mammal Assessment 2008 (www.iucnredlist.org/initiatives/mammals) to match our taxonomy (35) as described previously (44), excluding humans, domesticated and marine species, and uncertain, historical, and introduced ranges (see *SI Materials and Methods* for details). The dataset included a total of 861 extant species in 267 genera across the globe and 86 species in 44 genera in our regions (Fig. S7A). To adjust for the effects of end-Pleistocene and Holocene extinctions, we compiled available lists of extinct species (11, 23), selected the species recorded for our focal regions, and cross-checked them with our extant dataset (see *SI Materials and Methods* for the final list). We then adjusted present-day mammalian diversity in each continent and focal region by adding the number of extinct genera to the present-day observed genus richness.

Paleobotanical Data. Paleo-climatic data were obtained from several public sources that covered the Neogene as a whole (18), exclusively the Miocene (45), or exclusively the Pliocene (46, 47). We used terrestrial estimates of mean annual temperature (MAT) and precipitation (MAP) inferred from fossil plant communities which allowed us to calculate spatially explicit values of terrestrial NPP for each region or continent. We consider these datasets appropriate for the large temporal and spatial scales addressed here (48) and accounted for the temporal and climatic uncertainties associated with paleobotanical climate reconstructions as follows (see *SI Materials and Methods* and Fig. S5 for details). We allocated paleo-climatic records to our stratigraphic scheme following two different approaches to account for temporal assignment uncertainty (*SI Materials and Methods*) but found no substantial differences between the resulting NPP datasets (Fig. 2 I and J). To account for the spatially clumped data structure (Fig. 1), we summarized the

paleo-climatic records that fell into our set of focal regions (344 records in 182 locations, or 439 location-by-stratigraphic-stage combinations, available at dataportal-senckenberg.de/database/metacat/bikf.10018.1/bikf) into the 1° grid (Fig. S5). Because paleo-climatic records often provided ranges between minimum and maximum estimates that reflect climatic uncertainty for each fossil plant community (18), we took the entire distribution of climatic estimates into account when calculating mean estimates (and 50% credibility intervals) for each grid cell (*SI Materials and Methods* and Fig. S5). The great majority of paleo-climatic records were from Europe (Fig. 1). The sparseness of records in North America results from the known rarity of settings suitable for the preservation of paleobotanical material in the arid Neogene there (49), and hardly any alternative terrestrial paleo-climatic records exist for our spatial and temporal scales (50). We excluded paleo-climatic data derived from Neogene paleosols in North America (49) because these showed very low spatial and temporal congruence with our data (most paleosol data were for Central North America) and because similar paleosol compilations are lacking in Europe.

Calculation of NPP from Paleo-Climatic Estimates. We calculated NPP (in grams of dry matter per square meter per year) with the Miami model equation (Fig. S5B) (51) within each of the 1° grid cells that contained an estimate of MAT (in degrees centigrade) and an estimate of MAP (in millimeters per year). The Miami model is commonly applied to fossil data when no other NPP estimates or environmental drivers for more complex modeling are available and is considered robust at large spatial scales (52). We further demonstrated the robustness of NPP estimates from the Miami model with present-day data (see below). Our methods assume no effects of temporal changes in atmospheric CO₂ levels on paleo-climatic estimation from plant fossils or on conversion of paleo-climatic values to NPP estimates, because past CO₂ levels are still under debate, and recent vegetation models suggest that they are likely comparable to preindustrial levels since at least the late Miocene (50). Additionally, the influence of CO₂ fertilization on paleo-climatic reconstruction is considered negligible, particularly in areas where water is not the main limiting factor (18). For each stratigraphic stage and each region and continent, we calculated weighted mean NPP based on all grid cells with both a MAT and a MAP estimate (excluding stages with only one cell). To account for uncertainty in underlying climatic estimates, we used our measure of the paleo-climatic variance within grid cells as weights, i.e., we calculated a mean that was weighted with the inverse values of the width of the 50% credibility interval from the binned distribution of original paleo-climatic estimates (*SI Materials and Methods* and Fig. S5C).

Present-Day NPP Data and Human Appropriation. To obtain comparable NPP estimates for the present day, we calculated NPP with the Miami model (51) from contemporary climate records. Data on MAT and total annual precipitation from the Climate Research Unit (CRU) time-series (TS) dataset (version 3.21) for the years 1960–2010 (53) were resampled to our 1° grid. We calculated average present-day NPP within grid cells based on the arithmetic means across the 50 y (Fig. S6A) and regional and continental estimates as the average across all respective grid cells. We did not use remote-sensing data because these show actual NPP (including human impact), whereas the NPP estimated from potential vegetation is more appropriate for comparison with the fossil record. To investigate the robustness of NPP estimates, we showed that the potential NPP values derived with the Miami model correlated strongly with potential NPP estimated from a dynamic global vegetation model (DGVM) (Fig. S6 B–E). DGVMs are sophisticated models of plant population dynamics in response to abiotic parameters and perform well in the biomes covered by our focal regions (54); the DGVM used here was based on plant physiology, atmospheric CO₂, climate, hydrology, and soil (10). Our comparison (Fig. S6E) showed that NPP estimates derived with the Miami model provided a realistic picture of present-day potential NPP in the absence of human impact at the global scale. Finally, we estimated the actual primary productivity available in natural ecosystems today by adjusting NPP values for HANPP with a correction factor (Fig. S6C), which was the proportion of potential NPP (modeled by the DGVM, Fig. S6B) that remains after human modification and harvest (10). Remaining NPP adjusted for human appropriation (Fig. S6D) was calculated for each grid cell by multiplying potential NPP from the Miami model with the HANPP factor.

Analyses of the Mammalian Diversity–Productivity Relationship. We analyzed the temporal relationship of fossil γ diversity with NPP separately on the continental and regional scales and across stratigraphic stages for which we had sufficient data (more than five mammalian locations and more than one grid cell with NPP estimate), from the Aquitanian (starting 23 Mya, Europe only) or Langhian (starting 17 Mya, both continents) to the Early Pliocene

(ending 2.6 Mya) (datasets and R scripts are available at dataportal-senckenberg.de/database/metacat/bikf.10018.1/bikf). We fitted GLMMs with Poisson-distributed errors using Maximum Likelihood with the lme4 package for R (55). We chose a particular model structure because it best represented the hypothesis we wanted to test, i.e., whether γ diversity was related to NPP when accounting for effects of area and duration of the time interval (38) as well as for the temporal and spatial structure in the data (see Table 1, *SI Materials and Methods*, and Table S2 for details). These models were the best GLMMs from a selection of possible model specifications we tested (Table S2) following a standardized protocol (56). Marginal and conditional R^2 values for GLMMs were calculated with the MuMIn package (57, 58).

ACKNOWLEDGMENTS. We thank A. Barnosky, D. Currie, T. Hickler, T. Müller, M. Lawing, S. Pauls, R. O'Hara, E.-M. Gerstner, U. Salzmann, and B. Schmid for providing data, advice, and discussion. Our work was supported by German Research Foundation Emmy Noether Fellowship DFG FR 3246/2-1 (to S.A.F.), by the sFossil workshop at the Synthesis Centre for Biodiversity Sciences (iDiv) DFG FZT 118 (to J.S. and A.M.), by European Commission Marie Skłodowska-Curie Grant FP7-PEOPLE-2012-IEF-329645 (to J.T.E. and A.M.), by the Landesoffensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz (LOEWE) funding program of Hesse's Ministry of Higher Education, Research, and the Arts, and by the Kone Foundation (to J.T.E.). C.H.G. holds German Research Foundation Mercator Guest Professorship DFG INST 161/723-1 at Goethe University Frankfurt. This article is a contribution to the integrative Climate Change Biology (iCCB) programme.

- Currie DJ (1991) Energy and large-scale patterns of animal- and plant-species richness. *Am Nat* 137(1):27–49.
- Badgley C, Fox DL (2000) Ecological biogeography of North American mammals: Species density and ecological structure in relation to environmental gradients. *J Biogeogr* 27(6):1437–1467.
- Hawkins BA, et al. (2003) Energy, water, and broad-scale geographic patterns of species richness. *Ecology* 84(12):3105–3117.
- Hortal J, Rodríguez J, Nieto-Díaz M, Lobo JM (2008) Regional and environmental effects on the species richness of mammal assemblages. *J Biogeogr* 35(7):1202–1214.
- Wright DH (1983) Species-energy theory: An extension of species-area theory. *Oikos* 41(3):496–506.
- Barnosky AD, et al. (2011) Has the Earth's sixth mass extinction already arrived? *Nature* 471(7336):51–57.
- Dirzo R, et al. (2014) Defaunation in the Anthropocene. *Science* 345(6195):401–406.
- Barnosky AD (2008) Colloquium paper: Megafauna biomass tradeoff as a driver of Quaternary and future extinctions. *Proc Natl Acad Sci USA* 105(Suppl 1):11543–11548.
- Doughty CE, Faurby S, Wolf A, Malhi Y, Svenning J-C (May 19, 2016) Changing NPP consumption patterns in the Holocene: From megafauna-'liberated' NPP to 'ecological bankruptcy'. *Anthropocene Rev*, 10.1177/2053019616650466.
- Haberl H, et al. (2007) Quantifying and mapping the human appropriation of net primary production in earth's terrestrial ecosystems. *Proc Natl Acad Sci USA* 104(31):12942–12947.
- Koch PL, Barnosky AD (2006) Late Quaternary extinctions: State of the debate. *Annu Rev Ecol Evol Syst* 37:215–250.
- Janis CM, Damuth J, Theodor JM (2004) The species richness of Miocene browsers, and implications for habitat type and primary productivity in the North American grassland biome. *Palaeogeogr Palaeoclimatol Palaeoecol* 207(3–4):371–398.
- Blois JL, Hadly EA (2009) Mammalian response to Cenozoic climate change. *Annu Rev Earth Planet Sci* 37:181–208.
- Fortelius M, et al. (2014) Evolution of Neogene mammals in Eurasia: Environmental forcing and biotic interactions. *Annu Rev Earth Planet Sci* 42:579–604.
- Alroy J, Koch PL, Zachos JC (2000) Global climate change and North American mammalian evolution. *Paleobiology* 24(suppl. 4):259–288.
- Badgley C, et al. (2008) Ecological changes in Miocene mammalian record show impact of prolonged climatic forcing. *Proc Natl Acad Sci USA* 105(34):12145–12149.
- Figueirido B, Janis CM, Pérez-Claros JA, De Renzi M, Palmqvist P (2012) Cenozoic climate change influences mammalian evolutionary dynamics. *Proc Natl Acad Sci USA* 109(3):722–727.
- Utescher T, et al. (2014) The Coexistence approach—theoretical background and practical considerations of using plant fossils for climate quantification. *Palaeogeogr Palaeoclimatol Palaeoecol* 410:58–73.
- Fortelius M, et al. (1996) Provinciality, diversity, turnover, and paleoecology in land mammal faunas of the later Miocene of Western Eurasia. *The Evolution of Western Eurasian Neogene Mammal Faunas*, eds Bernor RL, Fahlbusch V, Mittmann H-W (Columbia Univ Press, New York), pp 414–448.
- Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. *Philos Trans R Soc Lond B Biol Sci* 345(1311):101–118.
- Eronen JT, et al. (2012) Neogene aridification of the Northern Hemisphere. *Geology* 40(9):823–826.
- Eronen JT, Janis CM, Chamberlain CP, Mulch A (2015) Mountain uplift explains differences in Palaeogene patterns of mammalian evolution and extinction between North America and Europe. *Proc R Soc B* 282(1809):20150136.
- Turvey ST (2009) *Holocene Extinctions* (Oxford Univ Press, Oxford, UK).
- Turvey ST, Fritz SA (2011) The ghosts of mammals past: Biological and geographical patterns of global mammalian extinction across the Holocene. *Philos Trans R Soc Lond B Biol Sci* 366(1577):2564–2576.
- Rabosky DL (2009) Ecological limits on clade diversification in higher taxa. *Am Nat* 173(5):662–674.
- Fine PVA, Rees RH (2006) Evidence for a time-integrated species-area effect on the latitudinal gradient in tree diversity. *Am Nat* 168(6):796–804.
- Fritz SA, et al. (2013) Diversity in time and space: Wanted dead and alive. *Trends Ecol Evol* 28(9):509–516.
- Newbold T, et al. (2015) Global effects of land use on local terrestrial biodiversity. *Nature* 520(7545):45–50.
- Barnosky AD, et al. (2012) Approaching a state shift in Earth's biosphere. *Nature* 486(7401):52–58.
- Janis CM, Scott KM, Jacobs LL (1998) *Evolution of Tertiary Mammals of North America* (Cambridge Univ Press, Cambridge, UK), Vol 1.
- Janis CM, Gunnell GF, Uhen MD (2008) *Evolution of Tertiary Mammals of North America* (Cambridge Univ Press, Cambridge, UK), Vol 2.
- Woodburne MO (2004) *Late Cretaceous and Cenozoic Mammals of North America: Biostratigraphy and Geochronology* (Columbia Univ Press, New York).
- Steininger FF (1999) Chronostratigraphy, geochronology and biochronology of the Miocene European land mammal mega-zones (ELMMZ) and the Miocene mammal-zones. *The Miocene Land Mammals of Europe*, eds Rössner GE, Heissig K (Dr. Friedrich Pfeil, Munich), pp 9–24.
- Gradstein FM, Ogg JG, Schmitz MD, Ogg GM (2012) *The Geologic Time Scale 2012* (Elsevier, Boston).
- Wilson DE, Reeder DM (2005) *Mammal Species of the World: A Taxonomic and Geographic Reference* (John Hopkins Univ Press, Baltimore), 3rd Ed.
- Jernvall J, Fortelius M (2004) Maintenance of trophic structure in fossil mammal communities: Site occupancy and taxon resilience. *Am Nat* 164(5):614–624.
- Eronen JT, Evans AR, Fortelius M, Jernvall J (2011) Genera are often better than species for detecting evolutionary change in the fossil record: A reply to Salesa et al. *Evolution* 65(5):1514–1516.
- Barnosky AD, Carrasco MA (2002) Effects of Oligo-Miocene global climate changes on mammalian species richness in the northwestern quarter of the USA. *Evol Ecol Res* 4(6):811–841.
- O'Hara RB (2005) Species richness estimators: How many species can dance on the head of a pin? *J Anim Ecol* 74(2):375–386.
- Gotelli NJ, Colwell RK (2011) *Estimating Species Richness. Biological Diversity - Frontiers in Measurement and Assessment*, eds Magurran AE, McGill BJ (Oxford Univ Press, Oxford, UK), pp 39–54.
- R Development Core Team (2014) *R: A Language and Environment for Statistical Computing. Version 3.1.2* (R Foundation for Statistical Computing, Vienna).
- Oksanen J et al. (2015) *Vegan: Community ecology package. Version 2.3-1*. Available at CRAN.R-project.org/package=vegan.
- Saarinen J, Oikarinen E, Fortelius M, Mannila H (2010) The living and the fossilized: How well do unevenly distributed points capture the faunal information in a grid? *Evol Ecol Res* 12:363–376.
- Fritz SA, Purvis A (2010) Phylogenetic diversity does not capture body size variation at risk in the world's mammals. *Proc R Soc B* 277(1693):2435–2441.
- Pound MJ, Haywood AM, Salzmann U, Riding JB (2012) Global vegetation dynamics and latitudinal temperature gradients during the Mid to Late Miocene (15.97–5.33 Ma). *Earth Sci Rev* 112(1–2):1–22.
- Salzmann U, Haywood AM, Lunt DJ, Valdes PJ, Hill DJ (2008) A new global biome reconstruction and data-model comparison for the Middle Pliocene. *Glob Ecol Biogeogr* 17(3):432–447.
- Salzmann U, et al. (2013) Challenges in quantifying Pliocene terrestrial warming revealed by data-model discord. *Nat Clim Chang* 3(11):969–974.
- Grimm GW, Denk T (2012) Reliability and resolution of the coexistence approach — a reevaluation using modern-day data. *Rev Palaeobot Palynol* 172:33–47.
- Retallack GJ (2007) Cenozoic paleoclimate on land in North America. *J Geol* 115(3):271–294.
- Forrest M, et al. (2015) Climate-vegetation modelling and fossil plant data suggest low atmospheric CO₂ in the late Miocene. *Clim Past* 11(12):1701–1732.
- Lieth H (1975) Modeling the primary productivity of the world. *Primary Productivity of the Biosphere*, eds Lieth H, Whittaker RH (Springer, Berlin), pp 237–263.
- Zaks DPM, Ramankutty N, Barford CC, Foley JA (2007) From Miami to Madison: Investigating the relationship between climate and terrestrial net primary production. *Global Biogeochem Cycles* 21(3):GB3004.
- Harris I, Jones PD, Osborn TJ, Lister DH (2014) Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 dataset. *Int J Climatol* 34(3):623–642.
- Smith B, et al. (2014) Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model. *Biogeosciences* 11(7):2027–2054.
- Bates D et al. (2015) lme4: Linear mixed-effects models using 'Eigen' and S4. Version 1.1-8. Available at CRAN.R-project.org/package=lme4.
- Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM (2009) *Mixed Effects Models and Extensions in Ecology with R* (Springer, New York).
- Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol Evol* 4(2):133–142.
- Bartoň K (2016) MuMIn: Multi-model inference. Version 1.15.6. Available at CRAN.R-project.org/package=MuMIn.
- Alroy J (2010) Fair sampling of taxonomic richness and unbiased estimation of origination and extinction rates. *Quantitative Methods in Paleobiology*, eds Alroy J, Hunt G. The Paleontological Society Papers 16:55–80.
- Dunhill AM, Hannisdal B, Benton MJ (2014) Disentangling rock record bias and common-cause from redundancy in the British fossil record. *Nat Commun* 5:4818.
- Mix HT, Mulch A, Kent-Corson ML, Chamberlain CP (2011) Cenozoic migration of topography in the North American Cordillera. *Geology* 39(1):87–90.
- Dormann CF, et al. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography* 30:609–628.

Supporting Information

Fritz et al. 10.1073/pnas.1602145113

SI Materials and Methods

Estimation of Mammalian γ Diversity. We calculated mammalian γ diversity with five different estimation algorithms (Chao, first-order Jackknife, Bootstrap, Chao1, and ACE) as described in *Materials and Methods* in the main text and found that the resulting diversity estimates were strongly correlated across the different methods (Fig. S2 A–J). Nonetheless, we used two algorithms (first-order Jackknife and Chao) in all analyses because they represent two different treatments of occurrence data, are considered reliable estimators (40), and are comparatively weakly correlated in our analyses (Fig. S2 A–J). The first-order Jackknife estimates diversity based on the presence or absence of location matrices (*Materials and Methods*), and the unbiased variant of the Chao estimator (Chao1, presented below) estimates diversity based on genus occupancy within continents or regions.

Despite the differences in underlying occurrence or occupancy data, the diversity patterns of the Chao1 estimator did not differ substantially from those for the Jackknife estimator (median absolute difference in regional and continental γ diversity estimates: five genera, maximum: 65 in continents, 78 in regions; Fig. S2F). Absolute diversity had a slight tendency to be higher with Chao1 than with Jackknife (53% of time intervals). In addition, some estimated errors were extremely large with Chao1; however, differences between diversity estimates were nearly always smaller than the associated errors (in 88% of time intervals for Chao1 errors). We analyzed the diversity–productivity relationship at the continental scale and in the three regions with highest data coverage using the stratigraphic stages (because paleo-climatic estimates were available only at that temporal resolution). In the continental datasets and the selected regional datasets, differences in diversity estimates across stratigraphic stages were very small between the two methods (median difference in γ diversity was four genera). Therefore, we conclude that the γ diversity estimates included in our analyses were robust to the diversity estimator algorithm.

We evaluated how data characteristics influenced both richness estimators and quantified the performance of Jackknife diversity estimation if true genus richness is known (36, 38, 43). To this end, we present three sets of analyses to (i) assess the effect of range-through genera with the Jackknife and Chao1 richness estimators, (ii) determine any effects of the different location definition used in North American and European fossil datasets on diversity estimation with Jackknife, and (iii) assess the performance of Jackknife estimation with simulations based on present-day data.

Effects of range-through genera. Our γ diversity estimation within each stratigraphic stage and land mammal age implicitly assumes that time intervals are independent of each other. This assumption increases the likelihood of underestimating diversity especially in short time intervals with few locations, because these often miss genera that were recorded before and after the interval but not within it. Here, we define range-through genera as those that have an occurrence gap during one or more time intervals but are present in intervals either side of the gap (note that this definition differs from that of ref. 59 and others, in which such genera are termed “Lazarus taxa,” denoted “ N_{r-} ”). Given that occurrence gaps may reflect either true fluctuations in diversity or data biases, we counted the number of range-through genera for gaps of either one or two intervals in length (note that one-interval range-through genera as defined here are equivalent to “part-timers” in ref. 59, denoted “ N_{pt} ”). Then, as a simple test for underestimation of genus richness by the diversity estimation methods, we investigated whether the total genus number (the sum of observed

and range-through genera) was above our γ diversity estimates in each of the time intervals.

We found that the number of range-through genera for gaps of either one or two intervals in length were usually low relative to the number of observed genera (see Fig. S1 for range-through genera with gaps of a single interval). For gaps of a single stratigraphic stage, the median number of range-through genera was 1.5 (maximum = 13) in continents and 0.5 (maximum = 11) in regions. The median number of range-through genera with single-interval gaps in continent-specific land mammal ages was 3 (maximum = 20) for continents and was 1 (maximum = 23) for regions. For gaps lasting two time intervals, the numbers decreased markedly: the median number of range-through genera was 0 (maximum = 2) in stratigraphic stages for the continental data, and 0 (maximum = 3) for the regional data. The median number of range-through genera with gaps lasting two land mammal ages was 3 (maximum = 7) for continents and was 1 (maximum = 5) for regions. Virtually all γ diversity estimates with both Jackknife and Chao1 estimators within a given time interval were above its total genus number for both continental and regional datasets (Fig. S2 K and L). The only exceptions were the stratigraphic stage Langhian and the MN05 zone in Eastern Europe (Fig. S2 K and L), where the number of locations was low and the proportion of range-through was genera high (10 locations with 19 genera observed, and 10 range-through genera in the Langhian and 8 in MN05) (Fig. S1G). However, because the numbers of range-through genera were low, especially in the stratigraphic stages, and because our diversity estimates were greater than the total genus number for almost all time intervals and regions, we argue that our diversity estimation effectively took the relatively weak effects of range-through genera into account.

Effects of different spatial location definitions in North American and European datasets. Spatial locations in the Eurasian source dataset were the unique spatial fossil sites, but locations in North America were based on fossil sites that were grouped if they were close in space and belonged to the same stratigraphic formation (see ref. 30 for details). To assess the impact of this difference in location definition between our source datasets, we ran a sensitivity analysis across the best-sampled part of Europe (following refs. 36 and 43) and determined the extent to which diversity estimates with first-order Jackknife varied when locations are spatially grouped. We sampled all locations in Western and Eastern Europe into the 1° grid (Fig. 1). This grouping of locations is probably at a much larger spatial scale than the grouping inherent in the North American dataset, so it should represent a robust test case for the effects of spatial grouping on diversity estimation. Up to 31 locations (within a given time interval) were grouped into one grid cell (Fig. S3A), and 8,875 occurrences of 411 genera in 1,222 locations across Western and Eastern Europe in the original location dataset were transformed into 6,208 occurrences in 257 grid cells (of the same genera).

There was a strong and significant correlation between Jackknife γ diversity estimates calculated using the original and grouped locations (Pearson’s $r = 0.996$, $t = 71.9$, $df = 40$, $P < 0.001$) (Fig. S3B). Temporal diversity patterns for the two European regions (Fig. S3 C and D) were highly robust to spatial grouping (median difference between estimates across stratigraphic stages and MN units: 2 genera; maximum: 12), although the confidence intervals were slightly larger with spatial grouping (mean error: 8 genera with locations, 10 genera with grid cells). Given these results, we conclude that our comparisons of North American and European diversity estimates are not affected by differences in spatial grouping of locations between the source datasets.

Performance assessment with simulations based on present-day data. We conducted a sensitivity analysis of Jackknife estimator performance with the present-day mammalian occurrence data (see *Materials and Methods* in the main text and discussion below for data description). To maintain the nonrandom spatial sampling structure of our fossil data in the present-day simulation dataset, we extracted present-day species lists by overlaying the species' range maps with the actual locations from the Neogene fossil dataset (298 locations in North America and 1,225 in Europe which could be assigned to a single continent-specific land mammal age; compare with Fig. 1). With this dataset, we assessed the effects of varying (i) the number of locations and (ii) the preservation of genera within locations on the estimation of γ diversity. To simulate the variation in number of locations observed in our fossil record, we randomly picked 100 sets of locations within each continent consisting of 10, 50, 100, or 200 locations each. These location numbers reflect typical values from the Neogene fossil data: Stratigraphic stages contained 166 locations on average (minimum: 38 locations; maximum: 485 locations) in Europe, and 47 locations on average in North America (minimum: 21 locations; maximum: 93 locations), whereas the more finely resolved continent-specific time intervals contained fewer locations (European MN units contained 1–190 locations and an average of 86, and North American NALMAs contained 11–48 locations and 25 on average). We also explored preservation of genera, because fossil locations preserve only a subset of the genera that actually lived there. For this exploration, we randomly deleted 20, 40, or 60% of the genus occurrences in each location from our randomized subsets of 100 locations. We then calculated γ diversity with the first-order Jackknife algorithm across the 100 randomizations in each location and preservation set and compared the results with the true number of genera from the present-day range maps.

Our simulations showed a robust performance of γ diversity estimation (Fig. S4). With only 50 locations, a median of 93% of the genus richness in North America and 78% in Europe was estimated (Fig. S4A). Diversity in North America was estimated well with as few as 10 locations (median 81%); in Europe performance with 10 locations was lower (median 58%) but increased to median values of 85% and 89% with 100 and 200 locations, respectively. Random deletion of genus occurrences within each location had comparatively little effect (Fig. S4B). In North America, reducing the genus occurrences by as much as 60% led to virtually the same median estimate as that of the full set (25.98 vs. 25.99 genera). In Europe, a decrease in performance was observed with reduction of genus occurrences, but it was weak (median estimate 74% of genera estimated with 60% of occurrences deleted, vs. 85% performance with all occurrences) (Fig. S4B). Our simulations also showed that even though not all genera known to be present in the focal regions were sampled in the complete location set, we were able to approximate the correct number of genera with a reasonable number of locations (compare the ranges of estimates and the horizontal dotted lines in Fig. S4 C and D).

These results indicate that γ diversity can be well estimated, even when data are much more incomplete than our mammalian fossil data are likely to be. Our analyses of the fossil diversity–productivity relationship focused on the stratigraphic stages, which on average contained more locations than necessary to achieve a performance of above 85% in our simulations. Further, provided enough locations were selected, the effect of undetected or nonpreserved genera on diversity estimation was low, assuming random loss of genera and unbiased preservation rates. Recent evidence shows that these assumptions are reasonable for large datasets such as our mammalian fossil record (60). Although for practical and philosophical reasons we can never be sure that we are estimating the true diversity (39, 40), our evaluation of the diversity estimation clearly indicated that

estimating mammalian γ diversity from the fossil occurrence data with a first-order Jackknife approach constituted a robust and appropriate way to quantify regional and continental genus richness.

Estimation of Present-Day Diversity and End-Pleistocene and Holocene Extinctions. To calculate present-day mammalian diversity, we edited the range maps from the IUCN Red List Global Mammal Assessment 2008 (www.iucnredlist.org/initiatives/mammals) to match our taxonomy (35), as described previously (44). We excluded marine species (the pinniped families Phocidae, Otariidae, and Odobenidae, the sea otter *Enhydra lutris*, and the polar bear *Ursus maritimus*), ranges classified as “presence uncertain,” “historical range,” and “introduced,” and humans and domesticated species (*Equus caballus*, *Bos taurus*, *Capra hircus*, *Ovis aries*, *Felis catus*), but not wild boar (*Sus scrofa*), wildcat (*Felis silvestris*), or wolf (*Canis lupus*).

We compiled lists of extinct genera for our focal regions and mammalian orders from literature (11, 23). The following 14 genera went extinct within Europe (listed by order): Artiodactyla: *Myotragus*, *Ovibos*, *Spirocerus*, *Bos*, *Megaloceros*, *Praemegaceros*, *Hippopotamus*, and *Phanourios*; Carnivora: *Crocota*; Perissodactyla: *Coelodonta*, and *Stephanorhinus*; Proboscidea: *Elephas*, *Mammuthus*, and *Paleoloxodon*. In North America, we counted 25 extinct genera: Artiodactyla: *Stockoceros*, *Tetrameryx*, *Bootherium*, *Bos*, *Euceratherium*, *Saiga*, *Camelops*, *Hemiauchenia*, *Paleolama*, *Bretzia*, *Cervalces*, *Navahoceros*, *Torontoceros*, *Mylohyus*, and *Platygonus*; Carnivora, *Homotherium*, *Miracinonyx*, *Smilodon*, *Arctodus*, *Tremarctos*, and *Platygonus*; Perissodactyla: *Equus* and *Tapirus*; Proboscidea: *Cuvieronius*, *Mammut*, and *Mammuthus*. Because our focal taxa also contained smaller-bodied carnivores that are generally not counted in estimates of megafaunal extinctions (11), these numbers are likely to underestimate the true number of extinctions for our focal taxa.

Paleobotanical Data.

Accounting for temporal uncertainty in allocation of stratigraphic stages. We obtained paleo-climatic data for temperature and precipitation from several public datasets, all of which were compilations across many original source studies that derived paleo-climatic estimates from fossil plant communities (18, 45–47). We identified and removed duplicate climatic locations (i.e., those included in several of these compilation datasets) based on the original source literature citations and on spatial and temporal proximity and overlap. All climatic records for the Pliocene were allocated to the stratigraphic stage of the Early Pliocene following previous work (46, 47). Age estimates for the other climatic records were usually given as absolute start and end times of stratigraphic stages (18, 45). Because absolute age definitions for any given stratigraphic time interval can change substantially depending on the stratigraphic reference scheme, we had to assign records from these datasets to our stratigraphic stages in a consistent fashion rather than following the original, disparate stratigraphic schemes in the different compilations, which were often not consistent with each other. We therefore allocated the non-Pliocene climatic records to our stratigraphic stage scheme (Table S1) following two different approaches to account for the uncertainty of date assignment. In both approaches, all records allocated to more than two stratigraphic stages were removed to reduce the temporal uncertainty of paleo-climatic estimates.

Our first approach (hereafter “automatic allocation”) was to assign stages automatically based on the minimum and maximum absolute age estimates given in the source data using a temporal accuracy of 0.3 Mya. To do so, we identified the uppermost and lowermost stage assigned to a climatic record as the most recent stages with start dates smaller or equal to the maximum and minimum estimated age ± 0.3 Mya, respectively. The upper- and lowermost stage and all stratigraphic stages between these then

were assigned to the climatic record. This first approach assigned many stages to most climatic records, so when we excluded records that were allocated to more than two stages, the automatic allocation dataset was severely reduced. As an alternative second approach (hereafter “stratigraphic allocation”), we assigned the nearest equivalent stage manually based on the stratigraphic reference scheme (or more detailed stratigraphic information if given) in the compilation datasets (18, 45). A table showing the allocated stages for each age range of paleo-climatic records is available online; a link to the files is given in *Materials and Methods* in the main text. In both approaches we excluded climatic records that had no age information or that were allocated to stages older than Aquitanian or younger than Early Pliocene.

After deriving the NPP from the paleo-climatic estimates (see *Materials and Methods* in the main text and details below), we evaluated the influence of temporal uncertainty in the assignment of stratigraphic stages on the final NPP estimates for continents and regions. We found that the temporal allocation of stratigraphic stages had little effect on the final NPP estimates, which were very similar when the automatic and stratigraphic allocation methods were compared (correlation across the stratigraphic stages: Pearson’s $r = 0.96$, $t = 10.5$, $df = 10$, $P < 0.001$) (see Fig. 2 I and J for continental estimates). The mean absolute difference in continental NPP between the two datasets was only $48.1 \text{ g dry matter}\cdot\text{m}^{-2}\cdot\text{y}^{-1}$ (maximum $191.8 \text{ g dry matter}\cdot\text{m}^{-2}\cdot\text{y}^{-1}$ in the North American Messinian, where estimates were based on two grid cells) (Fig. 2I). We therefore used the dataset based on manual stratigraphic allocation in the diversity–productivity analyses, because it was based on more original paleo-climatic data points.

Accounting for climatic uncertainty in deriving paleo-climatic estimates for each grid cell. Averaging paleo-climatic records across whole continents or regions directly does not account for spatial variation in sampling or uncertainty of climatic values. Instead, we averaged paleo-climatic estimates across locations within 1° grid cells taking climatic uncertainty into account, calculated the NPP and its uncertainty within these grid cells, and then averaged NPP estimates across grid cells weighted by the uncertainty (the latter two steps are described in the next section). Almost all paleo-climatic records provided ranges between a minimum and a maximum climatic estimate for a given fossil plant community, and the distribution of possible climatic values usually is assumed to be uniform between these endpoints (18). We used this distribution of climatic estimates, or uncertainty range, to calculate mean climatic estimates within each grid cell (Fig. S5). We generated a distribution of climatic values for a given grid cell by allocating each climatic range between a given minimum and maximum value into bins (bin size for temperature: 0.01°C ; bin size for precipitation: 1 mm/y). The mean climatic value assigned to the cell was the highest-density point of the resulting frequency distribution of binned values (Fig. S5B). To incorporate the variation within the grid cell and the uncertainty associated with the climatic estimates, we also extracted the upper and lower limits of the credibility interval containing 50% of the binned values (Fig. S5B).

In the two datasets, i.e., automatic and stratigraphic stage allocation, between 103 and 129 grid cell-by-stratigraphic stage combinations (in 64–82 grid cells) contained only one record for either temperature or precipitation. In these cases, the mean for the grid cell was calculated as the average of the given minimum and maximum climatic value. To derive a credibility interval around this mean, the limits of the 50%-interval of climatic values were approximated as the mean $\pm 1 \text{ SD}$ (this interval should contain 68.2% of values if data are normally distributed), with the SD calculated as the total range (maximum minus minimum) divided by 4 (because in a normal distribution 95% of values fall into the interval of 4 SD units around the mean). Further, a few climatic records did not supply a range of values

but instead supplied point estimates: Between 13 and 16 grid cell-by-stratigraphic stage combinations (in 12–14 grid cells) contained only a single record with a climatic point estimate. In these cases, the SD was set to 1°C or 120 mm/y to derive credibility intervals as above, based on the approximate median values across all climatic estimates given as ranges in the two datasets.

Calculation of NPP from Paleo-Climatic Estimates. Within each grid cell that contained paleo-climatic data, we calculated the NPP with the Miami model formula (Fig. S5B). The Miami model assumes positive relationships of MAT and MAP with the NPP and that the NPP is limited by either temperature or precipitation (Fig. S5B) (51). In effect, NPP values are calculated for each of MAT and MAP separately, and the smaller value then is chosen to reflect the final NPP. In our case, final mean NPP values were based on precipitation in the majority of grid cells. The credibility interval values for NPP within a grid cell were always derived from the same variable as the mean values. Finally, regional and continental values were obtained using the mean across grid cells weighted by the climatic uncertainty (inverse of the size of the 50% credibility interval; see *Materials and Methods* in the main text and Fig. S5C).

Analyses of the Mammalian Diversity–Productivity Relationship.

Main models of the fossil relationship. We ran separate analyses for continents and regions (see also *Materials and Methods* in the main text). These spatial units were not simple replicates at different spatial aggregation levels but instead represented the best-available current knowledge at a given spatial scale. Because of the lower data coverage in some regions, such as Caucasus and Eastern North America, we excluded these regions in regional analyses. We used GLMMs with Poisson-distributed errors because the response was a count variable (γ diversity estimates were rounded to integers to meet the model assumptions). We tested different model specifications (following ref. 56) to account for the temporal and spatial data structure (Table S2) and determined the best model based on comparison of the AIC as a random slope and intercept GLMM. This kind of model assumes that random effects act on the variance of fixed-effect estimates but not on their mean values (see Table S2 for model comparisons and formula).

We fitted random slope effects of the midpoint of the stratigraphic stage (age in Mya) within each continent or region and random intercept effects for continent or region identity. Preliminary analyses with generalized least-square models containing autoregressive terms indicated that temporal autocorrelation had no significant effect of on variance heterogeneity and that the fitted random effects in the GLMM should account for the temporal trend and the spatial structure in the data (56). The variable of interest (NPP, log-transformed), as well as the two covariates area (in square kilometers, log-transformed) and duration of the stratigraphic stage (in My, log-transformed), were fitted as fixed effects. To approximate the area for which γ diversity was estimated in each stratigraphic stage, area was measured with a convex hull around all the mammalian locations occurring in each stage (using a Mollweide projection). We accounted for area and stage duration because these parameters are known to influence diversity in fossil samples significantly even after controlling for spatial and temporal sampling biases. Correlation coefficients between the variables fitted as fixed effects were low (all absolute $r < 0.34$ for continents and $r < 0.24$ for regions), indicating that collinearity was unlikely to affect the models.

Supplemental models of the fossil relationship. We fitted the same GLMMs to γ diversity estimates averaged across land mammal ages within each stratigraphic bin to test the effect of estimating fossil mammalian diversity in the relatively long stratigraphic

stages as opposed to the finer temporal resolution of the land mammal ages (supplemental models in Table S2). Because of the strong effects of area in our models, we also ran additional models to check that the tectonic extension of Western North America during the Miocene did not drive our results. Between 15 and 5 Mya, Western North America experienced a major extension in area that is estimated to have doubled its east–west extent over that timeframe (61). The area estimates for Western North America were rescaled to approximate the extension process with the following multiplication factors: Aquitanian and Burdigalian 0.5 \times , Langhian 0.6 \times , Serravallian 0.7 \times , Tortonian 0.8 \times , Messinian 0.9 \times . We then fitted the GLMMs across stratigraphic stages as before and found that results of the continental analysis did not change qualitatively when area values were adjusted for the tectonic extension of North America (supplemental models in Table S2). The NPP remained as a significant predictor of mammalian γ diversity in the continental model but was nonsignificant in the regional model that accounted for tectonic extension (Table S2). We conclude that the diversity–productivity relationship found in the continental models was robust to the influence of North American tectonic changes on the covariate area, but the effect of area extension on the regional level warrants further study.

Models of the present-day spatial relationship. The substantial differences between fossil and present-day data, especially in their

sampling and underlying timescale, precluded a joint statistical analysis of the diversity–productivity relationship (see *Results and Discussion* in the main text). Therefore, we compared the present-day estimates with the model predictions from the fossil record only visually (Fig. 3). To evaluate the present-day diversity–productivity relationship across the globe, we followed a purely spatial approach, as commonly applied in macroecology, fitting generalized linear models with a Poisson error distribution across the 1° grid cells (Fig. S7). To check the influence of spatial autocorrelation, we refitted these models including a spatial autocovariate as additional predictor variable, which was calculated from the distance-weighted spatial correlation matrix of the residuals from the initial models that included only the NPP as predictor variable (62). The models controlling for spatial autocorrelation were virtually identical to the models presented in Fig. S7: The present-day diversity–productivity relationship across all grid cells globally was significant and strongly positive ($n = 12,730$, residual deviance = 6,455, slope estimate = 0.0005, $z = 149.6$, $P < 0.001$), as was the relationship across all cells falling within the fossil NPP gradient ($n = 4,531$, residual deviance = 3,553.9, slope estimate = 0.0006, $z = 75.2$, $P < 0.001$). The relationship across only the grid cells within our focal regions was significant but much weaker ($n = 1,460$, residual deviance = 112, slope estimate < 0.0001, $z = 3.2$, $P < 0.01$).

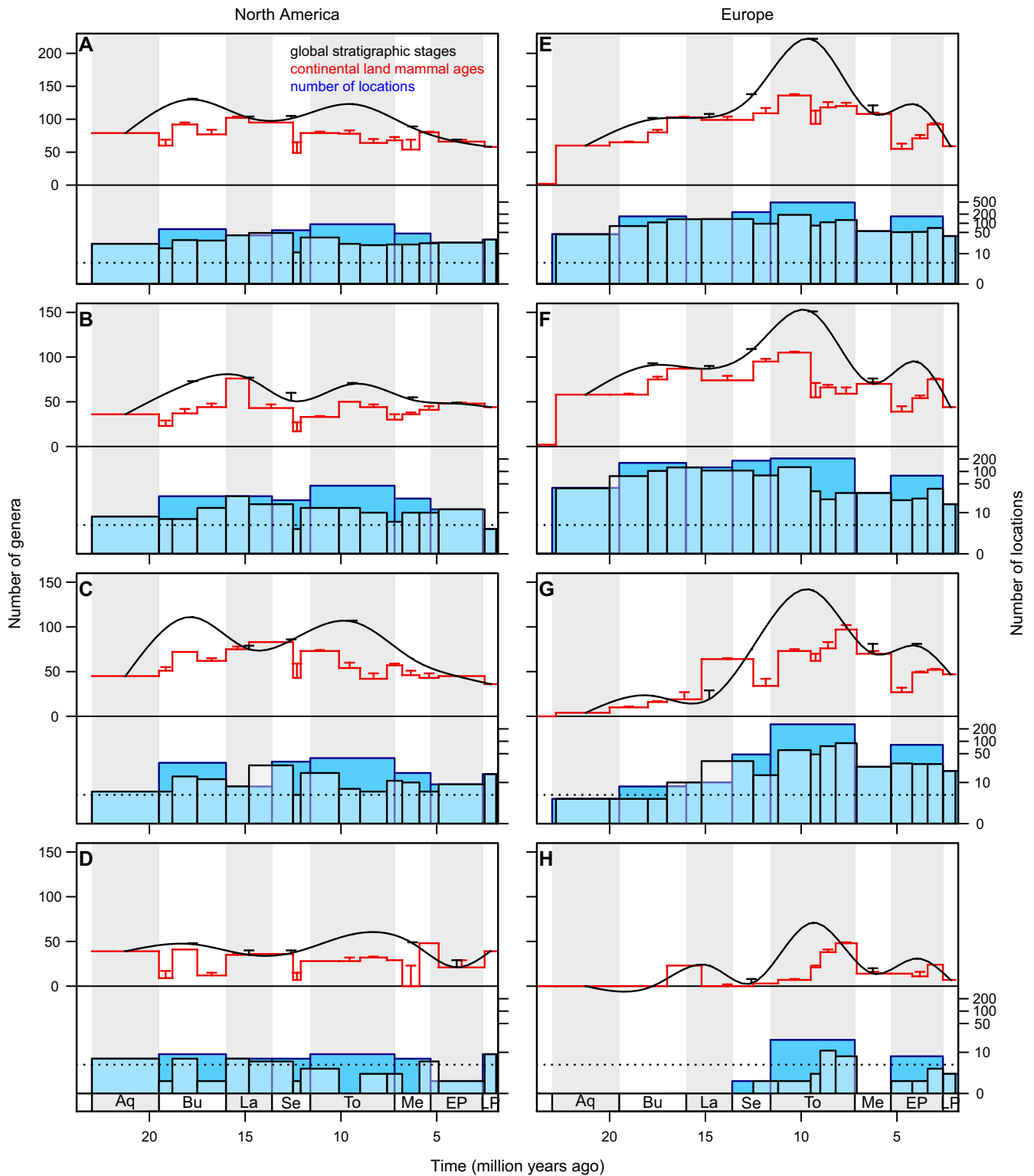


Fig. S1. Observed genus richness, number of range-through genera, and number of locations in the Neogene mammalian fossil record in North America (A–D) and Europe (E–H). Temporal dynamics of the number of observed mammalian genera (top of each graph, left axis) and the number of locations with mammalian fossils (bottom of each graph, right axis) were very similar. Numbers of genera are shown in global stratigraphic stages (black symbols, trend line fitted by natural cubic spline interpolation) and continent-specific land mammal ages (red stepped line), with error bars denoting the number of range-through genera (those genera with an occurrence gap for that single time interval). Numbers of locations are shown on a logarithmic scale for visual clarity, in stratigraphic stages (dark blue bars, blue outline) and in land mammal ages (light blue bars, black outline); the horizontal dotted line shows the cutoff of five locations (diversity estimation was performed only with at least six locations). Stratigraphic stages for land mammal ages (Table S1) are Aquitanian (Aq) (23.0–19.5 Mya), Burdigalian (Bu) (19.5–16.0 Mya), Langhian (La) (16.0–13.6 Mya), Serravallian (Se) (13.6–11.6 Mya), Tortonian (To) (11.6–7.2 Mya), Messinian (Me) (7.2–5.3 Mya), Early Pliocene (EP) (5.3–2.6 Mya), and Late Pliocene (LP) (2.6–1.8 Mya). (A) Continent-wide North America. (B) Western North America. (C) Central North America. (D) Eastern North America. (E) Continent-wide Europe. (F) Western Europe. (G) Eastern Europe. (H) Caucasus.

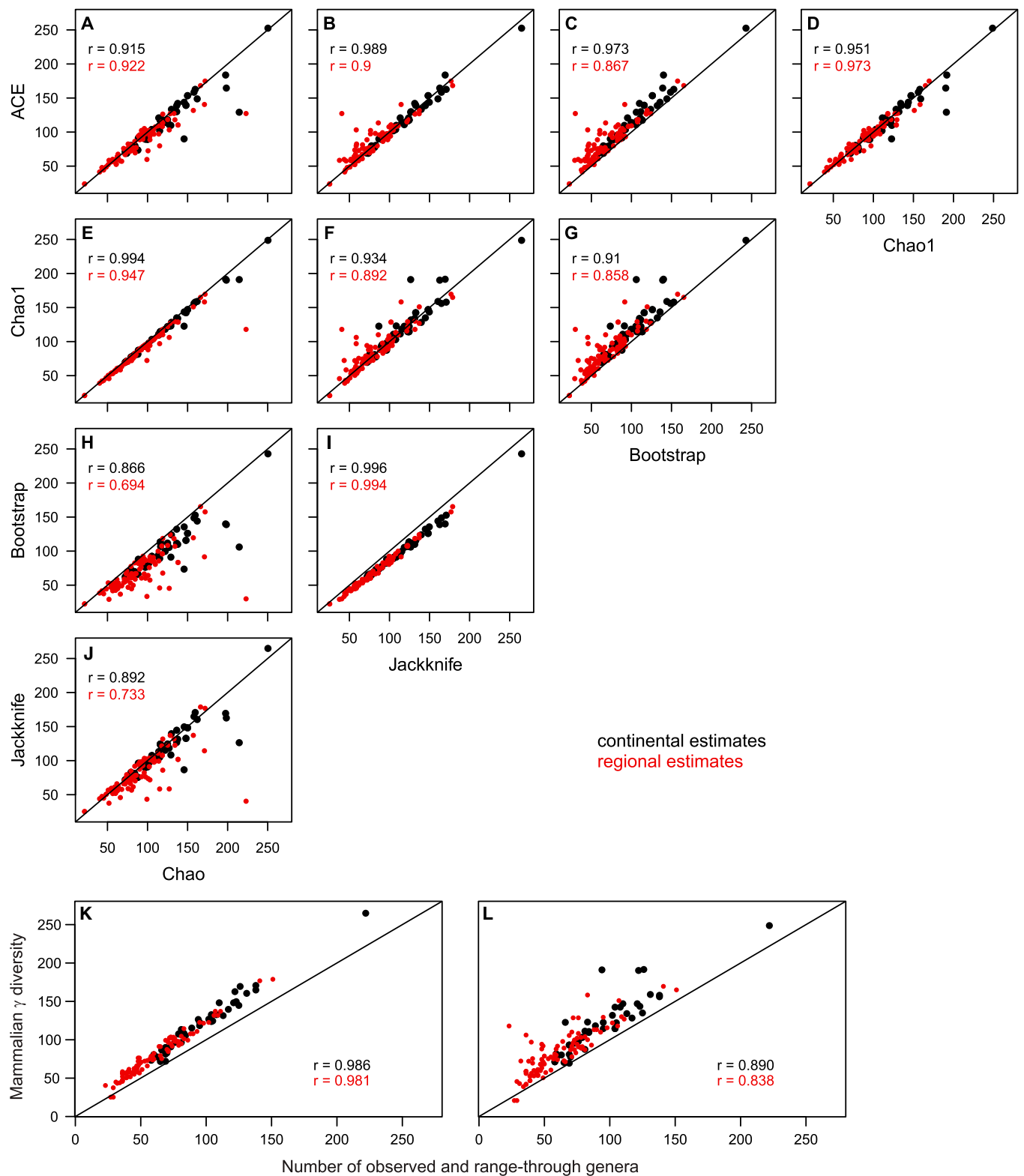


Fig. S2. Estimation of mammalian diversity with different richness estimators. (A–J) Matrix of scatter plots comparing γ diversity estimates for large terrestrial mammals through the Neogene, generated with five different richness estimators: the two richness estimators based on occupancy within continents or regions (A–D, ACE; D–G, Chao1) and the three estimators based on presence–absence location matrices for each continent or region (C and G–I, Bootstrap; B, F, I, and J, first-order Jackknife; A, E, H, and J, Chao). (K and L) Scatter plots of the relationship of γ diversity values, which were estimated with first-order Jackknife (K) and Chao1 (L), with the sum of observed genera and range-through genera (i.e., genera with an occurrence gap for that single time interval). Each plot shows all continental (black) and regional (red) diversity estimates within each time interval with more than five mammalian locations (stratigraphic stages and land mammal ages), Pearson's correlation coefficients across continental and regional estimates, and the 1:1 line.

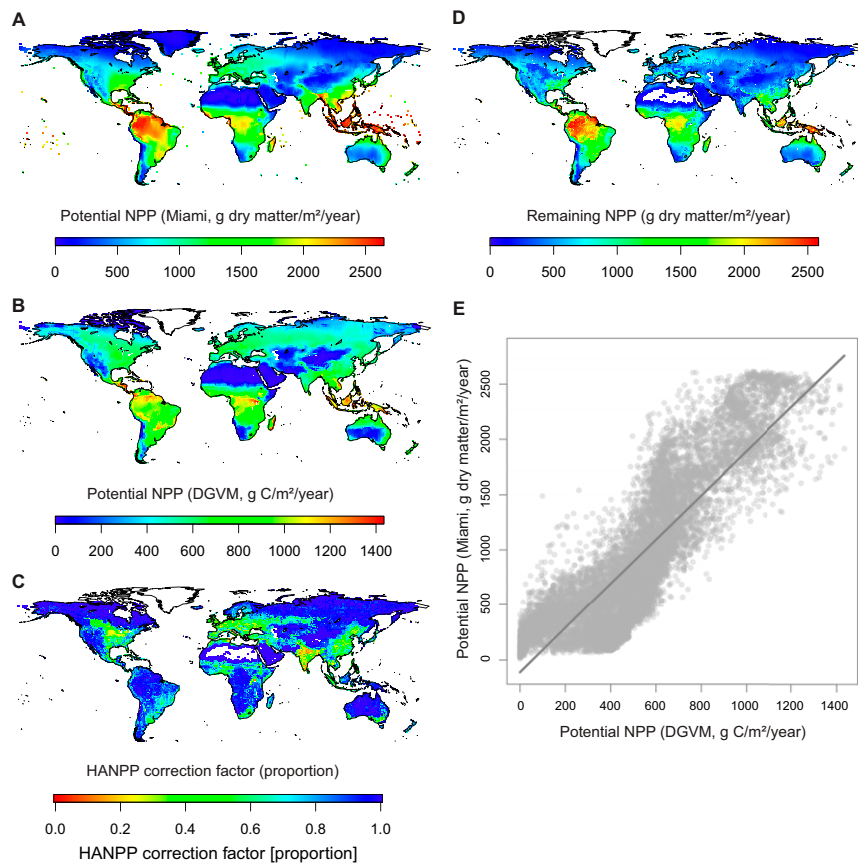


Fig. S6. Global variation of present-day NPP and its human appropriation and comparison of the Miami model with a DGVM. (A–D) Global maps showing different measures of present-day NPP in 1° latitudinal–longitudinal grid cells. (A) Potential NPP calculated from temperature and precipitation with the Miami model. (B) Potential NPP derived from a DGVM. (C) Correction factor for HANPP, calculated as the proportion of potential NPP (estimated with the DGVM, *B*) that remains after human modification and harvest. (D) Remaining NPP adjusted for human appropriation, derived by multiplying the Miami-model NPP values by the HANPP correction factor. (E) Scatter plot showing a strong global relationship across 1° grid cells (linear regression model: adjusted $R^2 = 0.77$, slope estimate = 1.96, $P < 0.001$) between present-day potential NPP values derived from the Miami model (A) and the DGVM (B). Symbols are semitransparent to show overlapping data points.

Table S1. Definitions of time intervals across North America and Europe and summary statistics of mammalian and paleobotanical Neogene fossil records

Land mammal ages or units			Stratigraphic stages								Order						Climate			
Europe			America		Global		Artiodactyla		Carnivora		Perissodactyla		Primates		Proboscidea		Climate			
Name	Start, Mya	Name	Start, Mya	Name	Start, Mya	Genus no.	Locations	Genus no.	Locations	Genus no.	Locations	Genus no.	Locations	Genus no.	Locations	Genus no.	Locations	Stratigraphic	Automatic	
MN01	23.80	Ar3	23.00	Aquitainian	23.00	65	47	44	35	25	39	0	0	1	1	1	26	20		
MN02	22.80																			
MN03	20.00	Ar4	19.50	Burdigalian	19.50	93	179	77	119	44	131	0	0	5	63	5	26	26		
MN04	18.00	He1	18.80																	
		He2	17.50																	
MN05	17.00	Ba1	16.00	Langhian	16.00	76	137	69	86	40	116	2	24	8	78	8	61	49		
MN06	15.20	Ba2	14.80	Serravallian	13.60	88	190	78	97	38	149	8	32	12	121	12	135	123		
MN07	12.50	Cl1	12.50																	
MN09	11.20	Cl2	12.10	Tortonian	11.60	131	360	113	241	48	395	13	58	18	267	18	68	93		
MN10	9.50	Cl3	10.10																	
MN11	9.00	Hh1	9.00																	
MN12	8.20	Hh2	7.60																	
MN13	7.10	Hh3	6.80	Messinian	7.20	74	68	69	60	27	81	3	13	9	35	9	53	39		
		Hh4	5.90																	
MN14	5.30	Bl1-4	4.90	Early Pliocene	5.32	81	137	60	87	16	115	4	25	11	100	11	22	23		
MN15	4.20																			
MN16	3.40																			
MN17	2.60	Bl5	2.50	Late Pliocene	2.60	45	54	41	42	8	55	2	7	8	44	8	—	—		
	1.95		1.90		1.80															
				Total dataset		320	1,075	239	702	86	991	21	156	24	679	24	306	240		

Time intervals were the MN units in Europe and NALMA in North America: Arikareean (Ar); Barstovian (Ba); Blancan (Bl); Clarendonian (Cl); Hemphillian (Hh). End dates of the youngest time interval are given in the second-last row. For each order of large terrestrial mammals and in each stratigraphic stage, the total number of genera observed and the number of spatially unique locations are given; climate data are summarized as the final number of paleo-climatic records from paleobotanical locations that could be used to calculate the NPP for each of the two datasets (*SI Materials and Methods*), one in which records were allocated manually to stratigraphic stages and one in which allocation was automatic based on the absolute age range given for the paleo-climatic record. The final row gives counts of unique genera, locations, and paleo-climatic records across all the stratigraphic stages (counts within stages do not sum up to these counts across stages because genera, locations, and paleo-climatic records can extend over multiple stages).

