Dance of the Cave Bear: Honouring the Scientific Legacy of Björn Kurtén

An assessment of body size and dietary biases in fossil mammal assemblages of the Pleistocene of Eurasia

Abigail K. Parker^{1,*}, Diana Pushkina^{1,2} & Liping Liu^{2,3,4}

- ¹⁾ Department of Computer Science, P.O. Box 68, FI-00014 University of Helsinki, Finland (*corresponding author's e-mail: abigail.parker@helsinki.fi)
- ²⁾ Department of Geoscience and Geography, P.O. Box 64, FI-00014 University of Helsinki, Finland
- ³⁾ The Swedish Museum of Natural History, P.O. Box 50007, SE-114 18 Stockholm, Sweden
- ⁴⁾ Bolin Center for Climate Research, Stockholm University, P.O. Box 50741, SE-10691 Stockholm, Sweden

Received 2 Feb. 2024, final version received 22 Apr. 2024, accepted 25 Apr. 2024

Parker, A. K., Pushkina, D. & Liu, L. 2024: An assessment of body size and dietary biases in fossil mammal assemblages of the Pleistocene of Eurasia. — *Ann. Zool. Fennici* 61: 253–280.

Paleoecological and ecometric studies of animal communities' functional composition require quantification of whether fossil faunas reliably record trait distributions of living communities. We analyzed whether body size and dietary biases exist within Pleistocene Eurasian sites included in the NOW database, compared with modern communities. Based on mass distributions, we discriminated small mammal sites, large mammal sites, and mixed sites in this record. Large mammals made up 50% of occurring genera and, on average, 74% of genera occurring within sites. Mixed sites with more generic occurrences (17% of sites) fell within modern communities' variation in their proportions of body size and dietary categories. Both large and small mammal sites fell outside the range of modern variation. Whereas most modern communities included 50%–60% herbivorous genera, small and large mammal fossil sites record more herbivorous genera (average 70%). Overall, large mammal sites predominated in this record, and herbivores were over-represented relative to non-herbivores.

Introduction

Björn Kurtén (1924–1988) was internationally known for advancing quantitative analysis in studies of ecology and evolution, including innovative studies on paleoenvironments, evolution rates, and population dynamics of fossil mammal communities (Kurtén 1952, 1953, 1958, 1968). His influence stimulated many later researchers to explore data-driven analyses in broad topics, many of which have focused on comparing spatial and temporal communities, or faunas derived

Edited by Oscar E. Wilson, University of Helsinki

from a single fossil locality (Fortelius *et al.* 2002, Jernvall & Fortelius 2004, Eronen *et al.* 2010, Liu *et al.* 2012, Žliobaitė *et al.* 2018, Lintulaakso & Kovarovic 2023). While Kurtén pursued applying quantitative approaches, he was also aware of biases introduced by limited data (Kurtén 1953).

Today, the fossil record appears less limited, as information about localities and occurrences is increasingly available in databases (Allmon *et al.* 2018, Smith *et al.* 2023, Žliobaitė *et al.* 2023). For example, the NOW Database (NOW — Database of fossil mammals, https:// doi.org/10.5281/zenodo.4268068, Žliobaite et al. 2023) enables ongoing Kurténian quantitative investigation of mammalian faunas (Fortelius et al. 2023). At the time of writing this paper, the NOW database contained 7838 fossil localities, and 16 573 mammal taxa. However, it is a long-acknowledged fact that a given fossil locality preserves only a small portion of the whole community that lived there (Darwin 1859, Andrews et al. 1979). The challenges of these preservational biases are ever-present in paleontology and must be considered especially in comparisons between sites in databases (Nanglu & Cullen 2023). Here, we took the example of the NOW database and interrogated bias in the proportions of fossils that have been recovered belonging to different trait-based ecological groupings, or guilds. We considered mammals from Pleistocene sites in Europe and Asia to quantify such biases within the record upon which Kurtén focused most of his work (Kurtén 1968).

Biases in the fossil record are introduced at many levels, beginning from differential weathering and decomposition of different organisms pre-burial, followed by differential burial and preservation of their remains in different depositional environments (Behrensmeyer 1984, Kidwell & Behrensmeyer 1988, Behrensmeyer et al. 2000). Then there are physical biases in the age of exposed rocks and proportion of geological settings, and finally human-induced biases in sampling, including unequal collecting effort across geography, biased collecting of taxa of interest, including through choice of collection methods, and unequal effort in taxonomic description of collected fossils. Combined, these biases in preservation and sampling lead to spatial and temporal heterogeneity in the makeup of fossil assemblages. Fossil faunas from different sites may have different compositions due to either variation in taphonomic processes or actual differences in taxonomic occurrences.

This poses a problem for paleontologists seeking to draw reliable conclusions about faunal associations or paleoenvironmental inferences based on incomplete fossils. Many studies indicate that biased fossil assemblages distort the understanding of ancient ecosystems and diversification (Raup 1979, Smith 2001, Butler et al. 2012, Benson et al. 2022b). However, biases in the composition of fossil assemblages are particularly problematic when those faunas are treated as communities for analyses of ecological function. Ecometric studies examine the relationships between functional traits within communities and aspects of the environment where they live (Eronen et al. 2010, Polly & Head 2015). Therefore, such studies rely on community sampling that is not biased in terms of the functional traits of interest. While Polly and Sarwar (2014) found that ecometric relationships between modern mammal communities and environmental variables are robust to downsampling of up to 50% of species, Faith et al. (2019a) found a strong effect of uneven sampling across fossil sites on their ecometric means, cautioning against the use of mean trait values from raw fossil occurrence data for paleoenvironmental inference. Although general ecological trends can be estimated using common species present in $\geq 25\%$ of localities (Jernvall & Fortelius 2002), studies of biases are useful to uncover more detailed ecological relationships in fossil communities.

We investigated the extent to which a large sample of Quaternary localities was biased in terms of the preservation of fossil mammals based on the fundamental ecological traits of diet and body size. There are various ecometric models based on skeletal correlates of diet in mammals (Fortelius et al. 2002, Eronen et al. 2010, Liu et al. 2012, Galbrun et al. 2018, Oksanen et al. 2019, Meloro and Sansalone 2022), and distributions of body sizes have been investigated as indicators of paleoenvironment (Legendre 1986, Morgan et al. 1995, Polly & Head 2015, Parker et al. 2023). Consideration of body size distributions in fossil faunas is complicated by the phenomenon of "large mammal sites" and "small mammal sites." Paleontologists have long recognized that some depositional environments preserve mostly small vertebrate bones and others mostly large vertebrates (Andrews et al. 1979). These depositional environments differ in characteristics including sedimentation rate, grain size, and flow energy, with lower-energy environments with fine-grained sediments being more likely to preserve small animal remains (Boggs 2006). Large mammal

localities likely have a higher discovery rate as bones are more easily observed weathering out of rock surfaces, while small mammal fossil sampling typically results not from surface collecting but from sieving of sediments, followed by labor-intensive identification of the bones from sieves (McKenna et al. 2005). The different sampling techniques chosen by researchers with different interests can bias the composition of recovered faunas, while geography and political factors bias which areas are prospected and sampled and how intensely. Fossils of larger-bodied taxa are over-represented in many records, even in well-sampled formations, despite the high diversity and abundance of smaller-bodied taxa (Damuth 1981, Alroy 1998, Brown et al. 2013a, Benson 2018, Behrensmeyer et al. 2019). Smallsized fossils suffer more overall from preservational and sampling biases (Behrensmeyer et al. 1979, Cooper et al. 2006). Peláez-Campomanes and Meulen (2009) surveyed the NOW database of fossil mammals and found that European sites preserve differing patterns of genus counts for large and small mammals. We used an up-todate and supplemented version of that database for all of Eurasia, also at the genus level, and developed a method to distinguish and define large mammal and small mammal sites in this record. As part of this method, we defined a body mass cutoff between large and small mammals tailored to this data set, which is higher than the 500 g or 1 kg cutoffs typically used in previous works (Morgan et al. 1995, Lovegrove & Haines 2004, Lintulaakso & Kovarovic 2016, Wilson et al. 2022).

We quantified community composition of faunas occurring in Quaternary fossil sites and compared their trait distributions at a coarse level to modern Eurasian communities sampled from range maps. Recent advances in geospatial data availability for modern species, including in the Phylacine database from which we drew these range maps (Faurby *et al.* 2018), enabled us to build taxonomic community samples that we could compare directly with fossil faunas. We can interrogate the modern record to understand how traits are distributed in modern-day communities and how those distributions vary across space and climate (Allen *et al.* 2006). With an understanding of this modern variation, in this case within cate-

gorical traits of body size and diet, we could evaluate whether fossil faunas match modern communities in their community compositions. At long timescales, trait or diversity gradients observed today may not hold (Fenton et al. 2023), and analyses of Quaternary mammal faunas have shown that taxa which occur in different environments today occurred together in the past, indicating that those faunas inhabited environments that were "non-analog", or not equivalent to any environments observed in the modern world (Graham 2005, Polly & Eronen 2011). While such non-analog assemblages exist, we expect that modern patterns in community composition in terms of coarse categorical ecological traits should be followed across Quaternary faunas. For example, we could deem a fossil fauna comprising 99% carnivores to be badly biased against other trophic levels based on ecological theory, and we can quantify this bias using statistical comparison to modern communities.

The modern day is the time point for which we have the most complete record of species distributions and, hence, the best sampling of co-occurring communities. However, we acknowledge that these species distributions are severely impacted by human activity and climate change (Borges et al. 2019, Pacifici et al. 2020, Faurby and Araújo 2018), and human impacts have fundamentally altered guild structures in extant communities (Malhi et al. 2016, Cooke et al. 2022). Our comparison of body size and dietary composition between fossil faunas and modern communities, therefore, also prompts consideration of whether community structures have changed over time. Over long timescales, ecological structures of communities may have evolved, so the relative diversities of species within body size and dietary categories in the past could differ from the range of variation observed in modern communities; however, such evolution in community structure can also only be discovered through measurement of fossil faunas' guild compositions (Andrews et al. 1979).

The primary purpose of this work was to show how trait-based biases manifest in the fossil record of Quaternary Eurasian mammals. Quantifying these biases through comparison to modern sites contributes to answering the question posed by Žliobaitė and Fortelius (2022) of whether the fossil record is "looking under a streetlamp" shining light on only a few of the animals that lived in the mostly unilluminated past. In this case, the question is whether there are two types of streetlamps, which filter for visibility of only small or only large mammals. We defined body mass groups and these site types based on breaks in observed data, allowing us to measure the strength of size- and diet-based filters due to the combined action of taphonomic and sampling biases.

Material and methods

Occurrence data

Modern community samples were drawn from the Phylacine database (Faurby et al. 2018), which provides trait data (body size and diet) and range maps for 5831 mammal species. We used the Phylacine range maps to sample modern communities at 21 002 sample points across the Eurasian landmass (geography mapped in Fig. 1B). These sample points were defined in the Eco-ISEA3H database, which uses a discrete global grid system to divide the globe into hexagonal cells with equal areas (Mechenich & Žliobaitė 2023). Eco-ISEA3H resolution 9, which we used here, has 50.3 km spacing between hexagon centroid points. We considered those species whose range maps overlap with the centroid point of a hexagon to be present in the present community for that cell, and determined communities by genus by recording which genera have any species occurring in each cell. This range-based approach to defining communities is more suitable for comparison to fossil communities because it includes spatial averaging across the hexagonal grid cells that parallels the temporal averaging across the depositional period of fossil sites (Saarinen et al. 2010, Du & Behrensmeyer 2018).

The fossil occurrences used in our analyses were drawn from the NOW Database (NOW — Database of fossil mammals, https://doi. org/10.5281/zenodo.4268068, Žliobaite *et al.* 2023) on 21 September 2023. This database is constantly updated and curated by a team of experts (Žliobaitė *et al.* 2023; geography of all sites mapped in Fig. 1A). We supplemented this NOW download with an additional 128 localities from Russia and Kazakhstan, which are currently private due to use in other ongoing projects and which increase sampling of higher-latitude Asian sites. We drew the occurrences for these sites (as documented in the GitHub repository associated with this article at https:// github.com/abbaparker/Size-Diet-Bias) from the literature, and they will be publicly available in the future in the NOW database. This database lists occurrences by site, defining each locality as a geographically-constrained location where fossils have been found in one sedimentary layer or a sedimentary sequence. Therefore, the age ranges of localities can vary greatly, with some dated by absolute methods to within some thousands of years, while others, especially localities surveyed earlier in history, have the ages of their deposits constrained only by mammalian biostratigraphy within some million years or more. The database follows the locality designations of original publications, with updated dates and occurrences from recent papers where available. Sites with large age ranges, therefore, can represent either assemblages of fossils deposited quickly, but in strata that have been imprecisely dated, or assemblages that have been deposited over a protracted period, in which case the resulting fauna includes time-averaging of the taxa that ranged across the area at several different times. To account for possible issues with the most imprecisely dated sites, we removed data from sites whose age ranges were greater than 2 Myr. For remaining sites, genus count does not increase at sites with wider ranges between their bounding dates (Appendix 1). Therefore, the first case, where localities' assemblages typically date to a more constrained period, but lack of precise dating causes them to be assigned only to a broad time bin, seems to be more common. The relative consistency in genus count across site age ranges (Appendix 1) supports the assumption that each of these sites can be considered a comparable community sample, though time-averaging on scales from seasons to many millennia certainly exists in many assemblages. While faunal changes that occurred over these periods of deposition act to inflate occurrence counts for sites, few assemblages include more



Fig. 1. The number of mammal genera observed at (**A**) 1018 Pleistocene and Holocene fossil sites in Eurasia, and (**B**) 21 002 modern sampling locations across Eurasia (occurrences from the Phylacine database (Faurby *et al.* 2018) at Eco-ISEA3H hexagonal samples). The vertical (latitude) and horizontal (longitude) histograms accompanying each map show the number of sites in every 5° bin of latitude or longitude, with the bar colors indicating the average genus count per sample within that bin.

genera than are typically observed at the closest modern sample point (hexagon centroids from the Eco-ISEA3H data set resolution 9). Additional occurrence data filtering included excluding any sites with minimum ages over 2.58 Ma, to limit to the Pleistocene. We did not set a minimum age boundary because a high proportion of sites (n = 134) have age ranges crossing the Holocene/Pleistocene boundary. Sixteen sites had ages constrained fully to under 11.7 kya; we retained this minority of Holocene sites in the data set but refer to the collective fossil sites as Pleistocene throughout the text for brevity. Additionally, we removed all sites outside the geographic boundaries of the Eurasian continent and all occurrences of marine mammal taxa (in families Phocidae, Otariidae, Odobenidae, Delphinidae, Balaenidae, Balaenopteridae, Dugongidae, Eschrichtiidae, Iniidae, Monodontidae, Neobalaenidae, Phocoenidae, Physeteridae, Trichechidae, Ziphiidae, Platanistidae, and genus Enhydra).

Of the 11 875 occurrences in this subset of the NOW database, only 346 (2.9%) were not identified to the genus level, whereas 2817 (23.7%) were unidentified to the species level. Therefore, we chose to carry out assemblage analyses at the genus level, so those occurrences could be included. Analyzing communities at the genus rather than species level has minimal impact; among the 10 397 genus occurrences at these sites, 8.7% had more than one species from the same genus present together (7.3% with two species, 1.0% with three species, 0.4% with four or more species per genus). In the modern world, 75.4% of genus occurrences at global Eco-IS-EA3H sampling points have one species occurring in the genus (in Eurasia, this figure is 76.5% of occurrences, with 14.2% of genus occurrences having two co-occurring species and 9.3% of genus occurrences having three or more co-occurring species within the genus).

Definition of small and large mammal groups

To create a data-defined mass-based cutoff between small and large mammals, we fit bimodal normal distributions to data sets of terrestrial mammal body mass. Throughout, we used \log_{10} values of body mass in grams due to the geometric normality observed in biological variation, especially for body mass distributions (Gingerich 2000, Benson *et al.* 2022a). For the fossil genera occurring in NOW (n = 452), we

used body mass estimates from the following sources: Phylacine database (n = 229) (Faurby et al. 2018), Cooke et al. (2022) (n = 60), the MOM database (Smith *et al.* 2003) (n = 10), the Pantheria database (n = 4) (Jones *et al.* 2009), and masses saved in the NOW database (n = 3). We took the average for each genus of its species occurring in each database, and where data was available from multiple data sets, we used the average mass value from the data set including the most species in the genus. For 146 occurring genera, there were no mass estimates available in the literature, so we assigned to those genera the average body mass of their family (averaged for all species in the family in the Phylacine database, which has the highest overall species sampling). We did not estimate body masses for three genera from families with no living representatives.

Modern species and genus body masses were drawn from the Phylacine database (Faurby *et al.* 2018). As for the fossil occurrences, we removed fully marine taxa, as coded in the database, as well as freshwater taxa from the families Delphinidae, Iniidae, Platanistidae, Phocoenidae, and Trichechidae, and the genera *Enhydra* and *Neovison*. For tests of body mass distributions, we removed species whose body masses in Phylacine are phylogenetically imputed (rather than reported from measurement), then found genus averages across the species in each genus. For tests of body mass distributions in Eurasia, we considered only taxa with occurrences at any of the Eco-ISEA3H Eurasian centroid points.

For the distributions of mammal body masses depicted in Fig. 2, we fit bimodal normal distributions using the *normalmixEM* function implemented in the R package mixtools. For a number k distributions, this function determines the mean (μ) and standard deviation (σ) of normal distributions best fitting the data, as well as the mixing proportion λ describing what proportion of observations fall within each of the distributions. We set the σ values for each distribution to be equal, ensuring that the standard deviations of each distribution spanned an equal range of the mass values. We used the initial condition $\sigma = 1$ and $\mu = c(1,5)$, which began tests for means between \log_{10} mass values from 1 to 5. We fit bimodal distributions (k = 2) for the body mass data sets of



Fig. 2. Distribution of body masses for data sets of mammal genera. **– A**: Estimated masses for genera occurring in the Pleistocene and Holocene of Eurasia in the NOW database fossil sites. **– B**: All modern mammal genera in the Phylacine database. **– C**: Modern mammal genera occurring in Eurasia in the Phylacine database. The red and blue curves are the results of fitting two normal distributions to each data set; their relative heights correspond to the proportion of genera that fall in each group; for their mean and SD values *see* Table 1. Dashed lines indicate the body mass value at which the probability of the genus falling in either group is equal. For equivalents of **B** and **C** at species-level resolutions *see* Appendix 4.

all genera occurring in the Pleistocene of Eurasia (Fig. 2A), all modern genera (Fig. 2B), modern genera occurring on the Eurasian landmass (Fig. 2C), all modern species in the Phylacine database, and those modern species occurring in Eurasia (Appendix 2). For each of these bimodal distributions, we defined the cutoff point between large and small mammals within the data set at the x-value of the point where the two modelled distributions intersect. This point is the mass at which a taxon has an equal probability of belonging to either of the two groups' distributions, and the cutoff value is dependent on λ , the relative proportion of observations within each group. We compared those λ values, as indicators of relative occurrence rates between large and small mammals, between data sets.

We assigned body masses to all genera occurring at each of the 719 Pleistocene Eurasian sites with over four genus occurrences. Then, we calculated distribution statistics for body mass at each site. We fit both bimodal and trimodal normal distributions to the mean body masses in the data set to determine divisions between sites with small/large or small/medium/large mean body masses (Fig. 3, Appendix 3 including sites with 3 or fewer genus sample counts). The mean values for these distributions represent the mean body size preserved at these two/three types of sites. Where the curves cross (marked by dashed lines on Fig. 3), the mean body mass for a site has an equal probability of belonging to either site type. Our results used the trimodal normal distribution division to separate small, mixed, and large mammal sites; based on the divisions in Fig. 3B, small mammal sites are those with mean body mass by genus under log₁₀ value 2.8 (628 g), large mammal sites are those with mean body mass over log₁₀ value 4.1 (12.7 kg), and mixed sites have mean masses between those two cutoffs. We counted large and small herbivores at each site, using the cutoff from the bimodal distribution in Fig. 2C, then calculated the percentage of genera at each site that were



Fig. 3. Distribution of mean body masses at Pleistocene/Holocene Eurasian fossil sites in the NOW database. -A: Overlaid with two fit normal distributions dividing "small mammal" and "large mammal" sites. -B: Overlaid with three fit normal distributions dividing "small mammal,", "mixed," and "large mammal" sites. Dashed lines indicate mean body mass values for which the probability of falling into two of these groups is equal.

small mammals. We then summarized these data for small, mixed, and large sites.

Dietary categorization

We used a simple binary classification of genus diets, with herbivores consuming a diet of 50% or more plant material, and all other taxa being classified as non-herbivores. This simplified categorization masks variation in the proportions of carnivores, omnivores, and insectivores within the non-herbivore category. Variation in the dietary composition of omnivores in particular complicates dietary categorization, and previous definitions of omnivory vary (Pineda-Munoz & Alroy 2014, Wilman et al. 2014, Lintulaakso et al. 2023, Reuter et al. 2023). We opted for this binary classification because it can be applied in a straightforward manner to categorizing the diets of extinct taxa. For modern genera, diet categories were determined based on diet data from Phylacine (Faurby et al. 2018), which reports the percentage of each species' diet that is plant, invertebrate, and vertebrate material. We took the mean of the percentage plant material

in the diets of species within each genus; if it was \geq 50%, the genus was categorized as herbivorous, and, if not, non-herbivorous. We used the same dietary category as modern for fossil genera within the Phylacine data set (n = 378). For extinct fossil genera, we assigned the family average diet in cases where there was no variation in herbivore/non-herbivore categorization within modern genera in the family (n = 28). For extinct fossil genera, we assigned the family average diet in cases where there was no variation in herbivore/non-herbivore categorization within modern genera in the family (n = 32). For 27 remaining genera, we determined their likely diets from the literature (Willemsen 1992, Parfitt & Harrison 2011, Domingo et al. 2016, Bartolini Lucenti & Rook 2021). We were unable to determine diet for 15 extinct genera that belong to modern families with variation in their diets.

Within modern community samples and fossil assemblages, we considered total counts of herbivore and non-herbivore genera. We calculated the percentage of genera categorized as herbivores, and compared these percentages between fossil assemblages and modern communities, as well as between small mammal and large mammal fossil assemblages across geographic space and time. We tested for body mass and diet community composition only for fossil sites with four or more genera occurring. We chose this threshold because it matches the minimum number of mammal genera sampled at any of the modern Eurasian sample points geographically closest to these fossil sites.

Ordination of variation in mammalian community structures

To quantify structural differences between fossil assemblages and modern communities in Eurasia, we employed a generalized linear latent variable model (GLLVM; Hui et al. 2015). This model is based on four variables for each sample point: the number of genera observed that are small herbivores, small non-herbivores, large herbivores, and large non-herbivores. GLLVMs summarize variation in ecological communities across latent variables (LVs) using occurrence or abundance data from those communities; these LVs can be used to relate guild structure to environmental gradients (Niku et al. 2019, Wilson & Parker 2023). Since we sampled only four guild categories $(2 \times$ 2 body size and diet classes), we fit a single LV to a negative binomial distribution of the genus counts per guild. The GLLVM was run on data including all fossil sites and a random sample of 8000 modern communities, to reduce matrix row count due to computational limitations. This model was implemented using the gllvm function in the R package gllvm (https://cran.r-project.org/ web/packages/gllvm/index.html), with 1000 as the maximum number of interactions and incorporating a random row effect to account for non-independence of samples close in space (Niku et al. 2019). We then used a t-test to test whether the mean LV scores for the fossil assemblages, divided into small, mixed, and large mammal sites, differ from the mean LV score for modern Eurasian communities. We also used a Kolmogorov-Smirnov test to test whether the LV scores for the site groups differ in their distributions from the LV distribution for the modern sites. Differences were considered significant at p < 0.05.

To visualize variation in community structure between fossil assemblages and modern Eur-

asian samples, we used the proportion of occurring taxa at each site that are small mammals and the proportion that are herbivores as plot axes. These proportions indicate trait-based community structure at a broad scale; they also allow for comparison in size and trait composition between fossil assemblages and modern samples that differ greatly in their absolute occurrence counts. We calculated a convex hull around the modern sample points in this small proportion vs. herbivore proportion space, using the function chull in R. We scored whether each Pleistocene site in the large, mixed, and small mammal site groups fell within that convex hull, and reported the percentage that do, for which the community composition in this space falls within modern observed variation.

Results

Division between small and large mammals

Large mammals are overrepresented in the fossil record of Pleistocene mammals in Eurasia. Whereas large mammals make up 20%-27% of named mammals in the modern data sets considered, they contribute 50.6% of genera described in this fossil record (Table 1 and Fig. 2). The bimodal distributions of modern mammal masses, summarized in Table 1 and Fig. 2, are generally consistent with previous analyses of body size distributions, which found that mammals exhibit a right-skewed distribution of body masses (Brown et al. 1993, Gaston & Blackburn 2000, Lovegrove & Haines 2004, Allen et al. 2006, Lyons & Smith 2013). The small/large mammal cutoff values varied somewhat between samples, spanning from 1-3 kg in body mass (Table 1). The cutoff of 1.09 kg $(\log_{10} \text{ value})$ 3.04) recovered for mammal species in Eurasia agreed with the division of small and large mammals at 1 kg body mass that Lovegrove and Haines (2006) described as the interface between mammalian locomotory modes. However, because our community analyses were all at the genus level, we chose to use the equivalent cutoff calculated for modern genera in Eurasia. This cutoff is at 1.76 kg (\log_{10} value 3.25), and



Fig. 4. The number of mammal genera observed at 1018 Pleistocene/Holocene Eurasian fossil localities plotted against the number of mammal genera occurring today at the closest sample in the Phylacine Eco-ISEA3H data set to each locality; *see* also Appendix 4 for equivalent counts within size and diet categories.

we used it throughout our analyses. Because 1.76 kg was the lowest large/small division recovered within the genus data sets (Table 1), it included slightly more genera within the better-sampled large mammal sites than the other cutoffs (Fig. 2).

Generic diversity

We compared the diversity of genera at the Pleistocene Eurasian fossil sites and modern community samples (Figs. 1, 4 and 5). The genus count at the nearest modern Eco-ISEA3H sample was, in almost all cases, higher than that observed for each fossil site (Fig. 2). There was low sampling of fossil sites at high and low latitudes, with most sites falling between 35° and 55°N. This latitudinal band also contains the most land area in Eurasia (see Fig. 1B latitude histogram, counting equal-area hexagons), but the fossil record outside this band is still depauperate relative to available land area, especially in the Indomalayan tropics. In Fig. 4, the increased variation away from the line (indicating 1:1 ratio of fossil to modern genera) at low latitudes (darker-colored points) emphasizes the paucity of the fossil record in diverse tropical regions. Today, Eurasian generic diversity peaks from 10°S to 25°N latitude (Fig. 1B), but in the Pleistocene, the highest genus counts are observed from 25° to 50°N.

Longitudinal patterns of genus count are obscured in the fossil record due to the low site count in the Middle East and central Asia (Fig. 1A). In modern samples, genus richness reaches over 100 at low-latitude sites around 100°E, but does not exceed around 50 in Europe

Table 1. Summary of the bimodal normal distribution fit (using R function *normalmixEM*) to five data sets of taxonomic body mass data for mammals. Group 1 in each case is the "small mammal" group, and Group 2 is the "large mammal" group. The mean mass for each normal distribution and its standard deviation are shown, along with the percentage of taxa in each group. This percentage is taken into account when calculating the cutoff value between groups (*see* dashed lines in Fig. 2 and Appendix 2).

Data set	Groups	Mean mass (log ₁₀ mass) (g)	SD	Percentage in group	Cutoff (log ₁₀ mass (g) with equal probability for both groups)
NOW Pleistocene Eurasia occurring	Group 1	1.98	0.81	49.4	3.42
genera	Group 2	4.89	0.81	50.6	= 2.61 kg
Phylacine modern all terrestrial	Group 1	1.96	0.81	72.9	3.56
genera	Group 2	4.59	0.81	27.1	= 3.61 kg
Phylacine genera in Eurasia	Group 1	1.89	0.72	74.1	3.25
	Group 2	4.19	0.72	25.9	= 1.76 kg
Phylacine modern all terrestrial	Group 1	1.74	0.76	80.0	3.31
species	Group 2	4.24	0.76	20.0	= 2.06 kg
Phylacine species in Eurasia	Group 1	1.65	0.69	77.7	3.04
	Group 2	3.93	0.69	22.3	= 1.09 kg



Fig. 5. The number of mammal genera observed at (A) 1018 Pleistocene/Holocene fossil sites in Eurasia, and (B) 21 002 modern sampling points across Eurasia, plotted against the longitude of their geographic locations. Site latitudes are indicated by the color of each point; *see* also Appendix 5.

(Fig. 5B). In contrast, fossil sites in Europe yielded high genus counts, almost matching those of East Asian sites (Fig. 5A). This relative enrichment of the low-longitude sites relative to higher longitude sites indicates either more thorough fossil sampling and description in Europe relative to Asia or less contrast in environmental conditions across this longitudinal gradient at some times in the Pleistocene.

Fossil site types

There is a clear distinction between fossil sites that preserve large and small mammal fossils. The distribution of mean body masses for Pleistocene Eurasian sites can be characterized as a bimodal distribution of sites with low/high mean body mass of occurrences (Fig. 3A), but a trimodal distribution fits the data more closely (Fig. 3B). This trimodal distribution suggests that some sites have taphonomic conditions that preserve mostly small mammals (red curve), others have taphonomic conditions that preserve mostly large mammals (blue curve), and others have exceptional preservation where both large and small mammals are preserved (purple curve). Using this trimodal distribution (Fig. 3B), 65.9% of sites fall into the "large mammal only" group, 16.5% of sites fall in the "small mammal only" group, and 17.6% of sites are mixed sites. Small

mammal sites are, on average, slightly older than the other sites (with a midpoint age of 1 Ma) and have wider age ranges reported (0.75 Ma, in contrast to 0.44 Ma on average for the other sites).

We broke down the occurrences in each site type in terms of body size and diet categories (Tables 2 and 3) for sites with four or more genera occurring. The four-genus threshold was set based on the minimum genus count at the sample of modern sites closest to the fossil sites (Fig. 4, excluding the single sample from a mammoth site on the Arctic island Severnaya Zemlya, where only one modern mammal genus occurs). We then compared sites with four or more genera with modern communities (Table 4) in terms of their community composition. Across all sites, there is a mean of approximately 10 large mammal genera and 4 small mammal genera occurring (Table 2). By diet category, all sites have a mean of approximately 9 herbivores and 5 non-herbivore genera (69% herbivores). In contrast, modern Eurasian sites have a mean of 25 small mammal genera and 15 large mammal genera, with 55% of the total genera being herbivores, on average (Table 4).

As expected, the small mammal sites have many more small than large mammals, and *vice versa*. Most of the mixed sites have between 34%–58% small mammal genera. Modern mammal communities have an average of 61.7% small mammal genera (data plotted in Fig. 6A). Therefore, small mammals are generally underrepresented in this record, even at mixed sites (*see* also Appendix 4). The mixed sites have, on average, more than twice as many occurrences (mean = 24) as the small- or large-only sites (respective means 11, 12) and therefore represent those sites with a higher preservation rate (Table 3), though they still do not match modern sites in generic diversity (Fig. 4). The more diverse mammal assemblages from the mixed

sites have, on average, 59% herbivore genera, whereas the small and large mammal sites both average over 70% herbivores. Modern communities average 55% herbivore genera; this figure is 54% when only large mammals are counted, and 57% for only small mammals. Therefore, the mixed sites preserve assemblages with closer to the proportion of dietary guilds expected from modern communities.

Table 2. The average genus counts of small and large mammal genera at different site types (small/large cutoff = 1.76 kg), and the average percentages of small mammal genera within these site types as defined using the cutoffs in Fig. 3B. The percentages of large mammal genera, under this two-category classification, can be found by subtracting the percentage of small genera from 100%.

Sites	Mean midpoint site age (Ma)	Mean site age range (Ma)		Total genus count	Small genus count	Large genus count	Percentage small genera
All	0.75	0.49	mean	13.83	3.92	9.9	26.3
			max–min	4–52	0–31	0–33	0-100
			SD	9.02	5.5	7.01	0.33
Small	1.02	0.75	mean	10.76	9.63	1.12	91.0
			max–min	4–27	3–25	0–8	60-100
			SD	6.13	5.37	1.62	11.7
Mixed	0.87	0.44	mean	23.59	10.46	13.1	45.9
			max-min	5–52	2–31	2–32	25–75
		SD	12.0	5.75	7.42	12.5	
Large	0.65	0.44	mean	12.16	0.94	11.22	6.0
-			max-min	4–46	0–13	3–33	0-37.5
		SD	6.95	1.63	6.01	8.5	

Table 3. The average genus counts of herbivore and non-herbivore genera at different site types (non-herbivores are any genus where modern species consume < 50% plant foods), and the average percentages of herbivorous mammal genera within these site types as defined using the cutoffs in Fig. 3B. The percentages of non-herbivores, under this two-category classification, can be found by subtracting the percentage of small genera from 100%.

Sites	Total genus Herbivore genus count count		Non-herbivore genus count	Percentage herbivores	
All					
mean	13.83	8.84	4.7	68.9	
max-min	4–52	0–29	0–26	0-100	
SD	9.02	5.10	4.73	0.2	
Small					
mean	10.76	7.66	3.04	70.4	
max-min	4–27	0–23	0–15	0-100	
SD	6.13	5.09	3.04	25.9	
Mixed					
mean	23.59	13.34	9.85	59.3	
max-min	5–52	1–29	0–26	20-100	
SD	12.02	6.65	6.5	17.1	
Large					
mean	12.16	8.02	3.83	71.0	
max-min	4-46	2–29	0–18	18.2-100	
SD	6.95	3.96	3.58	17.9	

Community composition

Having defined the small, mixed, and large mammal fossil assemblages from the Pleistocene of Eurasia, we next quantified how the faunas they preserve compare in their trait composition to modern Eurasian faunas. We used the percentage of small mammal genera and the percentage of herbivore genera as two axes defining a community composition space (Fig. 6). Using these proportions rather than total counts enabled us to make comparisons with modern sites where total genus counts are much higher (see Appendix 4 for how genus count in each trait category differs between fossil faunas and the nearest modern Eco-ISEA3H sample point). Within modern Eurasian communities, variation in this size/diet space relates to temperature (Fig. 6B) and latitude (Fig. 6C). The fossil assemblages tend to have a higher proportion of herbivorous mammals than the modern communities, and small and large mammal sites are separated along the body size proportion axis as expected (Fig. 6A).

We used a convex hull to define the space on these axes occupied by modern communities; the counts of fossil sites falling inside and outside this hull are listed in Table 5. Under 20% of large mammal sites fall within the range of modern Eurasian variation, while just over a quarter of small mammal sites do. In contrast, the majority of mixed sites have size and diet proportions that are observed in modern communities — 80% of these sites with higher generic sampling fall within the modern convex hull.

Our GLLVM analysis also revealed comparable community structure between mixed fossil sites' faunas and modern communities. The single latent variable we used to summarize variation in counts of genera within size/diet guilds had positive loadings for large herbivores (1.0) and small non-herbivores (2.77), and negative loadings for large non-herbivores (-1.35) and small herbivores (-1.13). Small and mixed sites had lower mean LV scores and greater standard deviations, in comparison to large sites or modern samples (Table 6 and Fig. 7). Fossil assemblages overall,

Table 4. Genus counts by diet type for modern Eurasian communities containing all mammals, only large mammals, and only small mammals, and the same for communities with four or more genera present. The 4+ genus subsamples are directly comparable to the fossil sites, where we consider only assemblages with four or more genera; sample counts here show how many modern Eurasian communities are above this threshold.

Sample	Sample count (number of Eco-ISEA3H sampled communities)		Total genus count	Herbivore genus count	Non-herbivore genus count	Percentage herbivores
Total communities	21002	mean	39.18	21.25	17.87	55.5
		max-min	1–111	1–56	0–59	26.1-100
		SD	17.46	8.51	9.83	8.5
Total communities	20983	mean	39.21	21.27	17.89	55.5
with 4+ genera		max–min	4–111	2–56	1–59	26.1-87.5
		SD	17.43	8.49	9.82	8.4
Large mammal communities	21002	mean	14.56	7.85	6.71	54.4
		max–min	1–45	0–26	0–21	0–100
		SD	7.31	3.96	3.69	10
Large mammal communities	20831	mean	14.66	7.9	6.76	54.1
with 4+ large genera		max-min	4–45	2–26	0–21	22.2-100
		SD	7.26	3.93	3.66	9.4
Small mammal communities	20965	mean	24.61	13.34	11.18	57.1
		max-min	1–75	0–38	0-40	0-100
		SD	11.8	5.78	7.08	12.2
Small mammal communities	20866	mean	24.72	13.58	11.24	57.0
with 4+ small genera		max-min	4–75	2–38	1-40	18.2-85.7
-		SD	11.77	5.74	7.06	12



Fig. 6. – A: Proportion of small (vs. large) mammal genera vs. the proportion of herbivore genera present at 718 Eurasian fossil sites and 20 983 modern Eurasian sample points, each with four or more mammal genera present. The convex hull indicates the area in this space occupied by modern Eurasian communities, excluding those with proportion small under 0.1 and proportion herbivore over 0.8, which are outliers. – B: The modern Eurasian sample points plotted on the same axes, with samples colored according to the mean annual temperature (MAT, °C). – C: The same sample points colored according to latitude. The convex hulls in B and C (gray) are the same as shown in A.

as well as the small and large site subsamples, had significantly different mean LV scores (*t*-test) and significantly different distributions of LV scores (Kolmogorov-Smirnov test) vs. modern communities (Table 6). The LV scores of mixed sites, however, had a mean and distribution comparable to those of modern mammal communities in Eurasia.

Discussion

Geographic biases

The Pleistocene fossil record of Eurasia is relatively well-sampled compared to most regions and ages; this is visible from inspection of site densities mapped from the NOW Database or the Paleobiology Database, which shows much denser sampling of Pleistocene terrestrial fossil sites in North America and Europe (paleobiodb. org). Large-scale comparisons of mammal assemblages are often only possible for these regions, which limits analysis of evolutionary trends and climatic relationships to these continents (e.g. Fritz et al. 2016, Huang et al. 2017, Janis 2023). The same bias in observation density exists in the modern world, where species records and ecological studies are highly concentrated in Europe and North America (Martin et al. 2012), frequently leading studies of modern ecological traits to focus on patterns on these continents (van Buskirk 2003, Olalla-Tárraga et al. 2006, Reif 2013). Our sample included both the densely-sampled record of Europe and many areas with sparse sampling, for example the Indomalayan tropics, where fossil genus counts are well below the diversity expected based on the modern latitudinal biodiversity gradient (Fig. 1; Rolland et al. 2014). In this Pleistocene sample, genus richness is highest on average at temperate-latitude sites. This is likely due to the poor preservation and exposure of fossils in dense

Table 5. Count of fossil faunas for which the proportions of small mammals and herbivorous mammals plot (in Fig. 6) within the convex hull containing modern Eurasian mammal communities. Note that the convex hull used for these counts surrounds all modern sample points, while that plotted in Fig. 6 excludes outliers.

Sites	Number of sites	Sites inside modern hull	Sites outside modern hull	Percentage of sites inside hull
Large	483	389	94	19.5
Mixed	119	96	23	80.7
Small	116	31	85	26.7

vegetation-covered tropical regions, as well as sampling intensity centered on sites in continental Europe and China/southern Russia, in particular at sites with archaeological finds (*see* peaks in Fig. 5A). However, particularly for large mammals, the higher relative generic diversity at high latitudes in the fossil record (*see* Appendix 5 for Fig. 5 with the axis exchanged for latitude) may reflect true patterns during interglacial periods and prior to megafaunal extinctions (Smith *et al.* 2018, Cooke *et al.* 2022).

Here, we took steps to address the non-uniform distribution of fossil sites across Eurasia in the NOW database by supplementing it with more records from China, Russia and central Asia. In our data set, small, mixed, and large mammal fossil assemblages were found across Eurasia, with no clear relationships between genus counts in guild categories and geography. However, the remaining areas with low site counts (Figs. 1A and 5A), notably in the Middle

Table 6. Differences between means (*t*-test) and distributions (KS test) of the LV scores for the listed sets of fossil sites and the LV scores for 8000 randomly sampled Eurasian mammal communities. The LV scores summarize community composition in terms of genus count within the guild categories of large herbivores, large non-herbivores, small herbivores, and small non-herbivores. The mean scaled LV score for the modern test points is 0.0027 ± 0.0965 SD (distributions plotted in Fig. 7). Differences at p < 0.05 (set in boldface) were considered significant.

Sites Nu of	Number		t-test	Mean ± SD scaled LV1 score	KS test	
	01 51(65	t	p		D	р
Large	483	-6.202	9.927e-10	-0.0153 ± 0.0595	0.2108	< 2.2e-16
Mixed	119	1.38	0.1704	-0.0102 ± 0.1016	0.1069	0.1372
Small All	116 718	-6.291 -7.862	5.62e-09 1.065e-14	-0.0651 ± 0.1155 -0.0225 ± 0.0811	0.3427 0.1696	4.342e-12 < 2.2e-16



Fig. 7. Densities of latent variable scores summarizing guild composition at the 718 Eurasian Pleistocene/Holocene fossil sites, divided into small mammal, mixed, and large mammal sites, and 8000 randomly-sampled modern Eurasian mammal communities. The t-test and Kolmogorov-Smirnov test revealed that the small and large mammal sites are significantly different in community composition along this latent variable from modern samples, while mixed sites' mammal assemblages are not significantly different (Table 6).

East and south and southeastern Asia, where modern tropical diversity leads us to expect richer fossil faunas (Fig. 4), remain a priority for exploration to enable synthesis of patterns in taxonomic richness and community structure across space and climate. The NOW database is continually updated, so future iterations of these analyses featuring more sites or increased sampling across mammal groups at existing sites may reveal stronger geographic patterns, including for narrower time bins.

Body size distributions and biases

The bimodal distribution of mammal body masses in the fossil record (with means at body masses 95 g and 75 kg, Table 1) prompts consideration of the components making up modern mammal mass distributions (Fig. 2). The long right tail of modern species (Appendix 2) or genus (Fig. 2B) mass distributions can be described as being composed of multiple normal distributions, with lower occupation of the larger mass categories (Brown *et al.* 1993, Lovegrove & Haines 2004). The theory behind the existence of normal distributions of body mass is that an optimum body mass exists for which reproduc-

tive output is maximized (Brown et al. 1993, Maurer 1998). Such optima exist because of metabolic scaling, wherein the energy acquired from food scales allometrically with body mass, so larger organisms acquire less energy per unit body mass (Brown & Sibly 2006). For mammals, this optimum has been suggested to be 100 g (Brown et al. 1993), which agrees in some cases with our results for the mean body size of the small mammal group in the bimodal distributions (Table 1), with the closest fit being the \log_{10} value 1.98 for small mammals in the Pleistocene fossil genera, equivalent to 95 g. Log-normal variation away from the optimum or mean is explained as being due to competition between species, as they partition resources in their environments (Allen et al. 2006). The higher counts of smaller species (skewness in the distribution) can be explained by the smaller-scale niche partitioning that occurs between small-bodied taxa (Hutchinson & MacArthur 1959). Alternatively, this skewness could be accounted for by higher speciation rates for small taxa (Gould 1988, Maurer et al. 1992, but see Liow et al. 2008), which is expected since rates of all kinds also scale allometrically with body size (Brown et al. 2004), or by elevated extinction risk for larger taxa (Janis & Carrano 1990, Cardillo 2003).

However, our results showing a bimodal distribution in body masses among fossil mammals support that, rather than modern mammal masses being distributed unimodally with skewness, mammalian body mass distributions comprise multiple size optima, resulting in overlapping distributions. Such overlapping multi-modal distributions were identified in Lovegrove and Haines (2004), where each distribution is defined by a locomotory category. Plantigrade mammals are the smallest and more diverse, with digitigrade and unguligrade mammals being successively larger, with a different size optimum observed for each. The non-plantigrade large mammals are less diverse today than in the past due to Pleistocene extinctions, reducing the skewness of the total distribution (Lovegrove & Haines 2004). Lovegrove and Haines's (2004) tests for bimodal cutoffs between large and small mammals vary across regions and spatial scales, but for the Nearctic realm, which overlaps with most of the Eurasian landmass we consider here, they find a cutoff at log_{10} mass value 3.14, similar to those in Table 1. In addition to differences in size optima based on locomotion, Liow et al. (2008, 2009) propose that mammals with small body size are better able to adopt hibernation/ torpor or burrowing behaviors to buffer against environmental change, which introduces selective pressures on size and can account for the relatively higher count of small mammal taxa.

The cutoff of 1.76 kg that we calculated between large and small Eurasian mammal genera differs from the round-number cutoffs frequently used by other researchers, such as 500 g (Morgan et al. 1995, Croft 2001, Tougard & Montuire 2006, Travouillon & Legendre 2009, Lintulaakso & Kovarovic 2016) or 1 kg (Behrensmeyer et al. 1979, Du & Behrensmeyer 2018, Fraser et al. 2022, Hardy & Badgley 2023). Future studies distinguishing between body size guilds in the fossil record could utilize our methodology to tailor the cutoffs between those size guilds to the distribution of masses within their group of interest. Table 1 and Fig. 2 show how these cutoffs can vary with geographic scope or taxonomic resolution. While in some cases, round numbers such as a 100 g mean for small mammal mass (Brown et al. 1993) or a 1 kg division between locomotory

guilds (Lovegrove & Haines 2004) are supported by theory, in most cases cutoffs for including taxa in size guilds should be informed by the statistical properties of the observed variation around multiple size modes, such as the two peaks in body mass frequency for fossil mammal genera (*see* Fig. 2A).

The roughly 50/50 division between small and large mammal genera we recover in the fossil record of Pleistocene Eurasia is unexpected based on these expectations from ecological theory. It suggests that the record has a strong bias towards preservation and/or description of large mammals. Behrensmeyer et al. (1979) found that small animals have a lower likelihood of their bones fossilizing due to pre-burial processes such as destruction of bones by carnivores and higher rates of surface weathering. Large mammals are disproportionately preserved both in the Holocene of Europe and in the Miocene globally (Crees et al. 2019, Žliobaitė & Fortelius 2022), and other fossil records also exhibit bias towards preserving large body size (Cooper et al. 2006, Brown et al. 2013b, 2022, Behrensmeyer et al. 2016). With regards to description, large mammals such as perissodactyls and proboscideans are frequent topics of paleontological study (Žliobaitė & Fortelius 2022), and Crees et al. (2019) found that large Holocene mammal taxa were described earlier in history than smaller mammals. Even in the modern day, and especially in the historical record, there is bias in reporting levels of sightings of modern mammals, with large mammals being more frequently noted as "charismatic megafauna" (Monsarrat & Kerley 2018). Alroy (2003) notes that, for modern mammals, taxonomic errors in the naming of small mammal taxa occur at a higher rate than for larger-bodied mammals, acting to inflate the population of species with masses plotting in the "small mammal" group.

However, the increased occupation of the large mammal size category in the Pleistocene of Eurasia relative to today also reflects the effects of Late Quaternary extinctions (Lovegrove & Haines 2004, Smith *et al.* 2018). Widespread extinction of mammals in the Late Quaternary is well-documented (Martin & Klein 1989, Flannery & Roberts 1999, Barnosky *et al.* 2004, Faith 2014). Large-bodied mammals were dispropor-

tionately affected by these extinctions (Smith et al. 2018: fig. 1). This size selectivity was unprecedented relative to earlier vertebrate extinction events (Smith et al. 2018), and extinction mechanisms triggered by human impacts have been proposed to explain this selectivity (Barnosky et al. 2004, Sandom et al. 2014, Smith et al. 2018). Human hunting features selection for larger-sized prey, potentially uniquely among pressures reducing populations, so even with low rates of hunting humans could have driven mammal species to extinction (Alroy 2001, Brook & Johnson 2006, Ben-Dor & Barkai 2021). Multiple recent studies have concluded that human migration and growing population density are the most likely cause for megafaunal extinction (Koch & Barnosky 2006, Prescott et al. 2012, Sandom et al. 2014, Saltré et al. 2016, Araujo et al. 2017, Andermann et al. 2020). These migrations occurred concurrent with climatic changes, which also contributed to extinction risk, particularly during the last deglaciation, when rapid temperature changes were linked to mammalian extinctions at least in Europe (Cooper et al. 2015, Wan & Zhang 2017). Recovery from population crashes, whether triggered by rapid environmental change or hunting, is particularly challenging for large mammals because of their low reproductive rates (Janis & Carrano 1991, Zuo et al. 2012). The unique impact of humans over recent millennia may have led modern body mass distributions in mammals (but also in other terrestrial vertebrates) to be depauperate in larger-bodied taxa, relative to what would have been the norm at other times in earth history.

In addition to bias in the genus count of small *vs*. large mammals in this Eurasian record, there is bias in the proportion of sites preserving small and large mammals. This result fits with paleontologists' intuitive understanding that there are separate "small mammal" and "large mammal" sites. However, our fitting of bimodal and trimodal distributions to the mean body size of faunas found across a database of sites is an early attempt at quantifying this small site/large site divide: 67% of localities with four or more generic occurrences were classified as large mammal sites when using the cutoffs defined by the trimodal distribution shown in Fig. 3C. We suggest that future studies investigating ques-

tions specific to large mammal or small mammal communities use similar cutoffs based on the mean body size of taxa observed at sites. For example, selecting only the large mammal sites and removing small mammal occurrences there will result in a subset of sites with generally similar preservation, allowing comparison between those large mammal faunas.

The higher occurrence of large mammal sites than small or mixed sites is likely driven by size-based recognition bias, as fossils of large bones could be recognized weathering out of sediments. Relative ease of surface collecting of large mammal bones, without need for the sieving and microscopy necessary for small mammal specimen collection and identification, may largely account for the approximately fivefold higher count of large mammal than small mammal sites in this record. However, taphonomic factors across depositional environments also favor the preservation of large mammals in the rock record over small ones (Behrensmeyer et al. 1979, 2000). Further research categorizing the depositional environments of these mammal sites could quantify the relative influence of such taphonomic biases vs. biases introduced in the human processes of selective collection and taxonomic identification.

Dietary biases

The majority of the sites included in this study have proportionally fewer non-herbivores than the modern communities do (Fig. 6). This bias against non-herbivore taxa that we recovered here, while noted previously (Wolff 1975, Andrews et al. 1979, Jernvall & Fortelius 2004), has received less attention in the literature than body size biases. In cases of archaeological sites, which are numerous among the sites studied here, herbivores may be over-represented because humans preferentially hunted herbivores in the Holocene (Crees et al. 2019). There is no analogous separation of herbivore-bearing and carnivore-bearing sites to the large/small site division discussed above (Fig. 6). Because carnivores have lower population densities, due to their higher trophic level, their fossils are less likely to be sampled at localities with lower sampling intensity (Damuth 1993, Turvey & Blackburn 2011, Žliobaitė & Fortelius 2022). Because we lumped omnivores, insectivores, and carnivores together in our non-herbivore category, we could not distinguish preservation patterns between these categories. However, since most omnivores and insectivores are small mammals, their fossil records are likely poor. Our latent variable analysis resulted in the genus count of small non-herbivores, with double biases against them, having the largest impact on LV scores, with communities plotting further to the right in Fig. 7 being those with more representatives of this guild. Future evaluation of how complete community sampling within fossil faunas is could use the presence of small non-herbivores to indicate good sampling.

The offset in proportion of herbivores and carnivores between fossil sites as a whole and modern communities may not be entirely a sampling artifact. Populations of livestock have displaced many native herbivore species (Veblen et al. 2016, Hempson et al. 2017, Filazzola et al. 2020). When most herbivory is done by a handful of domesticated species, the diversity of herbivores in communities drops, and the total biomass of wild mammals is today a small fraction of the global biomass of livestock (Greenspoon et al. 2023). Estimates of what species would co-occur in communities without human impacts, such as the "present natural" range map data set available for the Phylacine database (Faurby et al. 2018) can be used to test whether modern community herbivore proportions have been significantly shifted (down the y-axis of Fig. 6A) from their pre-anthropogenic state.

Additionally, non-analog community structures have been identified in the fossil record, for example in the Miocene of South America, where carnivores had low diversity (Croft *et al.* 2018), or the Plio-Pleistocene of Africa, where carnivores had higher diversity than today (Faith *et al.* 2019b). These are cases where differences in the composition of the carnivore guild are not due to biased sampling but reflect real differences in the structure of communities across continents or ages. We expect faunas of increasing age to fall farther outside modern ecospace, and propose that future GLLVM analyses, including fitting additional latent variables, can be used to quantify how relative guild occupation across communities converged over time towards community structures observed today.

Comparability of fossil faunas and modern communities

As discussed above, the strong signal of anthropogenic impacts on modern mammals prevents interpretation of trait differences between modern communities and fossil faunas purely as biases in the fossil record. However, ordination of community trait variation across a large geographic range of modern sites, as in Fig. 6, reveals that the range of variation in trait composition of modern communities clusters relative to the more scattered trait proportions of fossil faunas. Furthermore, the temperature and latitude color gradients in Fig. 6B and C reveal that even these coarse binary size and diet categories carry environmental signals, with high-latitude/ cooler regions having high proportions of herbivorous genera, and tropical/warmer regions having high proportions of small mammal genera. Taxon counts within guilds defined by size and diet, therefore, can act as ecometrics with utility for reconstructing paleoenvironments (Wilson & Parker 2023). This utility is limited, however, to fossil sites falling within the range of variation of modern training data (the convex hull in Fig. 6), which is a small proportion of the sites sampled here. We can deem sites plotting far from the modern cluster, for example those with 100% small or large mammals, as subject to strong biases in taphonomy and/or sampling, or else as representing faunas from truly non-analog past environments. However, the fact that the sites with the best sampling, the mixed sites (Table 3), fall within modern variation suggests that the other Pleistocene Eurasian sites would have modern analogs as well if the complete communities what lived at those sites had been preserved.

The mixed sites, while making up only 17% of total sites, stand out as being comparable to modern communities. With their increased generic diversity and lower mean proportion of herbivores, they exemplify exceptional preservation, at least for their mammal communities,

because they preserve a distribution of traits that can be reasonably considered to represent a functioning community (Nanglu & Cullen 2023). Having geospatial range data for a highly comprehensive set of modern taxa, which are produced by the IUCN (https://www.iucnredlist. org) and collected in the Phylacine database (Faurby et al. 2018), allows us to create community samples that, while abstracted relative to on-the-ground sampling, are standardized globally to allow for unbiased comparison. We drew genus counts in trait categories from these communities to quantify expectations based on modern communities. These figures (Table 4) can be compared to other assessments of fidelity in fossil faunas. As one example, Wolff (1975) conducted a study of Late Pleistocene sites in California and concluded that bulk sediment sampling yielded faunas with community composition comparable to modern mammal communities; however, these samples yielded 94% herbivores and 10% large mammals, which would place them among the small mammal sites in Fig. 6A, and not within modern community variation.

While the mixed sites are comparable to modern communities, large mammal sites are by far the most numerous. With this being the state of the record, one approach could be to focus research on sites falling in this mixed category, or intensifying field collection (e.g. sampling for small mammal material at large mammal sites) with the aim of shifting more sites into the mixed category, which generally exhibits higher genus counts (Nanglu & Cullen 2023). The other approach, focusing on data synthesis rather than collection work, is to limit analyses to the richer set of large mammal sites. Since the best record is that of large herbivores, then existing ecometric methods based on dental traits of these taxa remain a good target for future studies (Eronen et al. 2010, Liu et al. 2012, Evans 2013, Žliobaitė et al. 2016, 2018, Oksanen et al. 2019). Based on our quantification of the Eurasian record from the Pleistocene in comparison to today, we echo the advice of Damuth (1982) that analysis of fossil faunas should focus on a single trophic level, with primary consumers having the highest abundance in both living populations and the fossil record. Particularly, primary consumers

with body sizes above 1.78 kg are the most abundant in the Eurasian Pleistocene record, so consideration of trait distributions within this group, rather than between groups, can minimize biases between faunas.

Conclusions

We introduced a method to define large and small mammal groups based on bimodal distributions, which calculates cutoffs based on statistical properties of body mass distributions instead of arbitrary cutoff values. This method, which revealed fundamental divisions between body size classes between 1-4 kg for modern and fossil mammal datasets, can also be used to calculate trait divisions for other groups. Using the cutoff of 1.76 kg defined for modern Eurasian genera, large mammals make up 50% of Eurasian genera occurring in the Pleistocene, while they comprise 20%-30% of modern mammal taxa, indicating clear bias towards preservation and/or description of larger mammals. However, Pleistocene megafaunal extinctions also reduced the proportion of large taxa in communities today, so quantification of how fossil communities were non-analog in terms of body size remains a goal for future work.

We also fit distributions to categorize fossil mammal sites as small mammal, mixed, or large mammal sites based on their mean body masses of preserved genera. The utility of this site categorization was demonstrated by the differences in community composition in terms of both size and diet that we quantified using convex hulls and latent variable ordination. Both large and small mammal sites differ significantly from modern communities in having a higher proportion of herbivorous genera, approximately 70% on average. In contrast, the minority of mixed sites preserve dietary and size proportions more similar to modern faunas. Because over 60% of sites fell in the large mammal site type, we conclude that large-scale analyses of fossil mammal community composition may be limited to consideration of these large mammal sites, but caution that biases against non-herbivorous taxa prevent those large mammal communities from being functionally complete samples.

Acknowledgements

This work was supported by the Research Council of Finland projects 341620 and 341623. We thank the entire Helsinki paleontology community, and particularly the organizing committee of the Björn Kurtén Centenary Symposium for their work bringing this volume together.

References

- Allen, C. R., Garmestani, A. S., Havlicek, T. D., Marquet, P. A., Peterson, G. D., Restrepo, C., Stow, C. A. & Weeks, B. E. 2006: Patterns in body mass distributions: sifting among alternative hypotheses. — *Ecology Letters* 9: 630–643.
- Allmon, W. D., Dietl, G. P., Hendricks, J. R. & Ross, R. M. 2018: Bridging the two fossil records: paleontology's "big data" future resides in museum collections. — In: Rosenberg, G. D. & Clary, R. M. (eds.), *Museums at the forefront of the history and philosophy of geology: history made, history in the making*: Special paper 535. Geological Society of America, Boulder, CO, https://doi. org/10.1130/2018.2535(03).
- Alroy, J. 1998: Cope's rule and the dynamics of body mass evolution in North American fossil mammals. — Science 280: 731–734.
- Alroy, J. 2001: A multispecies overkill simulation of the end-Pleistocene megafaunal mass extinction. — *Science* 292: 1893–1896.
- Alroy, J. 2003: Taxonomic inflation and body mass distributions in North American fossil mammals. — Journal of Mammalogy 84: 431–443.
- Andermann, T., Faurby, S., Turvey, S. T., Antonelli, A. & Silvestro, D. 2020: The past and future human impact on mammalian diversity. — *Science Advances* 6, https:// doi.org/10.1126/sciadv.abb2313.
- Andrews, P., Lord, J. M. & Nesbit Evans, E. M. 1979: Patterns of ecological diversity in fossil and modern mammalian faunas. — *Biological Journal of the Linnaean Society* 11: 177–205.
- Araujo, B. B., Oliveira-Santos, L. G. R., Lima-Ribeiro, M. S., Diniz-Filho, J. A. F. & Fernandez, F. A. 2017: Bigger kill than chill: the uneven roles of humans and climate on late Quaternary megafaunal extinctions. *Quaternary International* 431: 216–222.
- Barnosky, A. D., Koch, P. L., Feranec, R. S., Wing, S. L. & Shabel, A. B. 2004: Assessing the causes of Late Pleistocene extinctions on the continents. — *Science* 306: 70–75.
- Bartolini Lucenti, S. & Rook, L. 2021: "Canis" ferox revisited: diet ecomorphology of some long gone (Late Miocene and Pliocene) fossil dogs. — Journal of Mammalian Evolution 28: 285–306.
- Behrensmeyer, A. K. 1984: Taphonomy and the fossil record: the complex processes that preserve organic remains in rocks also leave their own traces, adding another dimension of information to fossil samples. — American

Scientist 72: 558-566.

- Behrensmeyer, A. K., Western, D. & Boaz, D. E. D. 1979: New perspectives in vertebrate paleoecology from a recent bone assemblage. — *Paleobiology* 5: 12–21.
- Behrensmeyer, A. K., Kidwell, S. M. & Gastaldo, R. A. 2000: Taphonomy and paleobiology. — *Paleobiology* 26: 103–147.
- Ben-Dor, M. & Barkai, R. 2021: Prey size decline as a unifying ecological selecting agent in Pleistocene human evolution. — *Quaternary* 4, 7, https://doi.org/10.3390/ quat4010007.
- Benson, R. B. J. 2018: Dinosaur macroevolution and macroecology. — Annual Review of Ecology, Evolution, and Systematics 49: 379–408.
- Benson, R. B., Godoy, P., Bronzati, M., Butler, R. J. & Gearty, W. 2022a: Reconstructed evolutionary patterns for crocodile-line archosaurs demonstrate impact of failure to log-transform body size data. — *Communications Biology* 5, 171, https://doi.org/10.1038/s42003-022-03071-y.
- Benson, R. B. J., Brown, C. M., Campione, N. E., Cullen, T. M., Evans, D. C. & Zanno, L. E. 2022b: Comment on "The influence of juvenile dinosaurs on community structure and diversity". — *Science* 375, https://doi. org/10.1126/science.abj5976.
- Boggs, S. 2006: Principles of sedimentology and stratigraphy. — Pearson Prentice Hall, Upper Saddle River, NJ.
- Borges, C. M., Terribile, L. C., Oliveira, G. d., Lima-Ribeiro, M. d. S. & Dobrovolski, R. 2019: Historical range contractions can predict extinction risk in extant mammals. — *PLoS ONE* 14(9), e0221439, https://doi.org/10.1371/ journal.pone.0221439.
- Brook, B. W. & Johnson, C. N. 2006: Selective hunting of juveniles as a cause of the imperceptible overkill of the Australian Pleistocene megafauna. — Alcheringa: An Australasian Journal of Palaeontology 30: 39–48.
- Brown, J. H. & Sibly, R. M. 2006: Life-history evolution under a production constraint. — *Proceedings of the National Academy of Sciences* 103: 17595–17599.
- Brown, J. H., Marquet, P. A. & Taper, M. L. 1993: Evolution of body size: consequences of an energetic definition of fitness. — *The American Naturalist* 142: 573–584.
- Brown, C. M., Campione, N. E., Mantilla, G. P. W. & Evans, D. C. 2022: Size-driven preservational and macroecological biases in the latest Maastrichtian terrestrial vertebrate assemblages of North America. — *Paleobiology* 48: 210–238.
- Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. & West, G. B. 2004: Toward a metabolic theory of ecology. — *Ecology* 85: 1771–1789.
- Brown, C. M., Evans, D. C., Campione, N. E., O'Brien, L. J. & Eberth, D. A. 2013a: Evidence for taphonomic size bias in the Dinosaur Park Formation (Campanian, Alberta), a model Mesozoic terrestrial alluvial–paralic system. — *Palaeogeography, Palaeoclimatology, Palaeoecology* 372: 108–122.
- Brown, C. M., Campione, N. E., Giacomini, H. C., O'Brien, L. J., Vavrek, M. J. & Evans, D. C. 2013b: Ecological modelling, size distributions and taphonomic size bias in dinosaur faunas: a comment on Codron et al. (2012).

- Butler, R. J., Brusatte, S. L., Andres, B. & Benson, R. B. 2012: How do geological sampling biases affect studies of morphological evolution in deep time? A case study of pterosaur (Reptilia: Archosauria) disparity. — *Evolution* 66: 147–162.
- Cardillo, M. 2003: Biological determinants of extinction risk: why are smaller species less vulnerable? — Animal Conservation 6: 63–69.
- Cooke, R., Gearty, W., Chapman, A. S. A., Dunic, J., Edgar, G. J., Lefcheck, J. S., Rilov, G., McClain, C. R., Stuart-Smith, R. D., Kathleen Lyons, S. & Bates, A. E. 2022: Anthropogenic disruptions to longstanding patterns of trophic-size structure in vertebrates. — *Nature Ecology & Evolution* 6: 684–692.
- Cooper, A., Turney, C., Hughen, K. A., Brook, B. W., McDonald, H. G. & Bradshaw, C. J. A. 2015: Abrupt warming events drove Late Pleistocene Holarctic megafaunal turnover. — *Science* 349: 602–606.
- Cooper, R. A., Maxwell, P. A., Crampton, J. S., Beu, A. G., Jones, C. M. & Marshall, B. A. 2006: Completeness of the fossil record: estimating losses due to small body size. — *Geology* 34: 241–244.
- Crees, J. J., Collen, B. & Turvey, S. T. 2019: Bias, incompleteness and the 'known unknowns' in the Holocene faunal record. — *Philosophical Transactions of the Royal Society B* 374, 20190216, https://doi.org/10.1098/ rstb.2019.0216.
- Croft, D. 2001: Cenozoic environmental change in South America as indicated by mammalian body size distributions (cenograms). — *Diversity and Distributions* 7: 271–287.
- Croft, D. A., Engelman, R. K., Dolgushina, T. & Wesley, G. 2018: Diversity and disparity of sparassodonts (Metatheria) reveal non-analogue nature of ancient South American mammalian carnivore guilds. — *Proceedings of the Royal Society B* 285, 20172012, http://doi.org/10.1098/ rspb.2017.2012.
- Damuth, J. 1981: Population density and body size in mammals. — *Nature* 290: 699–700.
- Damuth, J. 1982: Analysis of the preservation of community structure in assemblages of fossil mammals. — *Paleobiology* 8: 434–446.
- Damuth, J. 1993: Cope's rule, the island rule and the scaling of mammalian population density. — *Nature* 365: 748–750.
- Darwin, C. 1859: On the origin of species. John Murray, London.
- Domingo, M. S., Domingo, L., Abella, J., Valenciano, A., Badgley, C. & Morales, J. 2016: Feeding ecology and habitat preferences of top predators from two Miocene carnivore-rich assemblages. — *Paleobiology* 42: 489–507.
- Du, A. & Behrensmeyer, A. K. 2018: Spatial, temporal and taxonomic scaling of richness in an eastern African large mammal community. — *Global Ecology and Biogeography* 27: 1031–1042.
- Eronen, J. T., Polly, P. D., Fred, M., Damuth, J., Frank, D. C., Mosbrugger, V., Scheidegger, C., Stenseth, N. C. & For-

telius, M. 2010: Ecometrics: The traits that bind the past and present together. — *Integrative Zoology* 5: 88–101.

- Evans, A. R. 2013: Shape descriptors as ecometrics in dental ecology. — Virtual Morphology and Evolutionary Morphometrics in the new millenium 24: 133–140.
- Faith, J. T. 2014: Late Pleistocene and Holocene mammal extinctions on continental Africa. — *Earth-Science Reviews* 128: 105–121.
- Faith, J. T., Du, A. & Rowan, J. 2019a: Addressing the effects of sampling on ecometric-based paleoenvironmental reconstructions. — *Palaeogeography, Palaeoclimatol*ogy, *Palaeoecology* 528: 175–185.
- Faith, J. T., Rowan, J. & Du, A. 2019b: Early hominins evolved within non-analog ecosystems. — *Proceedings* of the National Academy of Sciences 43: 21478–21483.
- Faurby, S. & Araújo, M. B. 2018: Anthropogenic range contractions bias species climate change forecasts. — *Nature Climate Change* 8: 252–256.
- Faurby, S., Davis, M., Pedersen, R. Ø., Schowanek, S. D., Antonelli, A. & Svenning, J.-C. 2018: PHYLACINE 1.2: the phylogenetic atlas of mammal macroecology. — *Ecology* 99, 2626, https://doi.org/10.1002/ecy.2443.
- Fenton, I. S., Aze, T., Farnsworth, A., Valdes, P. & Saupe, E. E. 2023: Origination of the modern-style diversity gradient 15 million years ago. — *Nature* 614: 708–712.
- Filazzola, A., Brown, C., Dettlaff, M. A., Batbaatar, A., Grenke, J., Bao, T., Peetoom Heida, I. & Cahill, J. F. Jr. 2020: The effects of livestock grazing on biodiversity are multi-trophic: a meta-analysis. — *Ecology Letters* 23: 1298–1309.
- Flannery, T. F. & Roberts, R. G. 1999: Late Quaternary extinctions in Australasia. — In: MacPhee, R. D. E. (ed.) *Extinctions in near time: causes, contexts, and consequences*: 239–255. Springer, Boston, MA.
- Fortelius, M., Eronen, J., Jernvall, J., Liu, L., Pushkina, D., Rinne, J., Tesakov, A., Vislobokova, I., Zhang, Z. & Zhou, L. 2002: Fossil mammals resolve regional patterns of Eurasian climate change over 20 million years. — Evolutionary Ecology Research 4: 1005–1016.
- Fortelius, M., Agustí, J., Bernor, R. L., De Bruijn, H., van Dam, J. A., Damuth, J., Eronen, J. T., Evans, G., van den Hoek Ostende, L. W., Janis, C. M., Jernvall, J., Kaakinen, A., von Koenigswald, W., Lintulaakso, K., Liu, L., Ataabadi, M. M., Mittmann, H.-W., Pushkina, D., Saarinen, J., Sen, S., Sova, S., Säilä, L. K., Teskov, A., Vepsäläinen, J., Viranta, S., Vislobokova, I., Werdelin, L., Zhang, Z. & Žliobaitė, I. 2023: The origin and early history of NOW as it happened. — In: Casanovas-Vilar, I., van den Hoek Ostende, L. W., Janis, C. M. & Saarinen, J. (eds.), *Evolution of Cenozoic land mammal faunas and ecosystems*: 7–32. Springer International Publishing, Cham.
- Fraser, D., Villaseñor, A., Tóth, A. B., Balk, M. A., Eronen, J. T., Barr, W. A., Behrensmeyer, A. K., Davis, M., Du, A., Faith, J. T., Graves, G. R., Gotelli, N. J., Jukar, A. M., Looy, C. V., McGill, B. J., Miller, J. H., Pineda-Munoz, S., Potts, R., Shupinski, A. B., Soul, L. C. & Lyons, K. S. 2022: Late quaternary biotic homogenization of North American mammalian faunas — *Nature Communications* 13, 3940, https://doi.org/10.1038/s41467-022-

31595-8.

- Fritz, S. A., Eronen, J. T., Schnitzler, J., Hof, C., Janis, C. M., Mulch, A., Böhning-Gaese, K. & Graham, C. H. 2016: Twenty-million-year relationship between mammalian diversity and primary productivity. — *Proceedings of the National Academy of Sciences* 113: 10908–10913.
- Galbrun, E., Tang, H., Fortelius, M. & Žliobaitė, I. 2018: Computational biomes: the ecometrics of large mammal teeth. — *Palaeontologia Electronica* 21: 1–31.
- Gaston, K. J. & Blackburn, T. M. 2000: Body size. In: Pattern and process in macroecology: 201–271. John Wiley & Sons, Hoboken, NJ.
- Gingerich, P. D. 2000: Arithmetic or geometric normality of biological variation: an empirical test of theory. — Journal of Theoretical Biology 204: 201–221.
- Gould, S. J. 1988: Trends as changes in variance: a new slant on progress and directionality in evolution. — *Journal of Paleontology* 62: 319–329.
- Graham, R. W. 2005: Quaternary mammal communities: Relevance of the individualistic response and non-analogue faunas. — *Paleontological Society Papers* 11: 141–158.
- Greenspoon, L., Krieger, E., Sender, R., Rosenberg, Y., Bar-On, Y. M., Moran, U., Antman, T., Meiri, S., Roll, U., Noor, E. & Milo, R. 2023: The global biomass of wild mammals. — *Proceedings of the National Academy* of Sciences 120, e2204892120, https://doi.org/10.1073/ pnas.2204892120.
- Hardy, F. C. & Badgley, C. 2023: Mammalian faunal change of the Miocene Dove Spring Formation, Mojave region, southern California, USA, in relation to tectonic history. — GSA Bulletin, https://doi.org/10.1130/B37082.1.
- Hempson, G. P., Archibald, S. & Bond, W. J. 2017: The consequences of replacing wildlife with livestock in Africa. — Scientific Reports 7, 17196, https://doi.org/10.1038/ s41598-017-17348-4.
- Huang, S., Eronen, J. T., Janis, C. M., Saarinen, J. J., Silvestro, D. & Fritz, S. A. 2017: Mammal body size evolution in North America and Europe over 20 Myr: similar trends generated by different processes. — *Proceedings of the Royal Society B*, 20162361, https://doi. org/10.1098/rspb.2016.2361.
- Hui, F. K. C., Taskinen, S., Pledger, S., Foster, S. D. & Warton, D. I. 2015: Model-based approaches to unconstrained ordination. — *Methods in Ecology and Evolution* 6: 399–411.
- Hutchinson, G. E. & MacArthur, R. H. 1959: A theoretical ecological model of size distributions among species of animals. — *The American Naturalist* 93: 117–125.
- Janis, C. M. 2023: Asymmetry of evolutionary patterns between new world and old world equids and among new world equine tribes. — In: Casanovas-Vilar, I., van den Hoek Ostende, L. W., Janis, C. M. & Saarinen, J. (eds.), Evolution of Cenozoic land mammal faunas and ecosystems: 25 Years of the NOW Database of Fossil Mammals: 143–164. Springer International Publishing, Cham.
- Janis, C. M. & Carrano, M. 1991: Scaling of reproductive turnover in archosaurs and mammals: why are large terrestrial mammals so rare? — Annales Zoologici Fennici

28: 201-216.

- Jernvall, J. & Fortelius, M. 2002: Common mammals drive the evolutionary increase of hypsodonty in the Neogene. *Nature* 417: 538–540.
- Jernvall, J. & Fortelius, M. 2004: Maintenance of trophic structure in fossil mammal communities: site occupancy and taxon resilience. — *The American Naturalist* 164: 614–624.
- Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., Safi, K., Sechrest, W., Boakes, E. H., Carbone, C., Connolly, C., Cutts, M. J., Foster, J. K., Grenyer, R., Habib, M., Plaster, C. A., Price, S. A., Rigby, E. A., Rist, J., Teacher, A., Bininda-Emonds, O. R. P., Gittleman, J. L., Mace, G. M. & Purvis, A. 2009: PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. — *Ecology* 90: 2648–2648.
- Kidwell, S. M. & Behrensmeyer, A. K. 1988: Overview: ecological and evolutionary implications of taphonomic processes. — Palaeogeography, Palaeoclimatology, Palaeoecology 63: 1–13.
- Koch, P. L. & Barnosky, A. D. 2006: Late Quaternary extinctions: state of the debate. — *Annual Review of Ecology*, *Evolution, and Systematics* 37: 215–250.
- Kurtén, B. 1952: The Chinese Hipparion fauna. Commentationes Biologicae, Societas Scientiarum Fennica 13: 1–82.
- Kurtén, B. 1953: On the variation and population dynamics of fossil and recent mammal populations. — Acta Zoologica Fennica 76: 1–118.
- Kurten, B. 1958: Life and death of the Pleistocene cave bear, a study in paleoecology. — Acta Zoologica Fennica 90: 1–59.
- Kurtén, B. 1968: Pleistocene mammals of Europe. Weidenfeld and Nicolson, London.
- Lee, H. L., Fortelius, M., Lintulaakso, K., Mannila, H. & Stenseth, N. C. 2009: Lower extinction risk in sleepor-hide mammals. — *The American Naturalist* 173: 264–272.
- Legendre, S. 1986: Analysis of mammalian communities from the late Eocene and Oligocene of southern France. — Palaeovertebrata 16: 191–212.
- Lintualaakso, K. & Kovarociv, K. 2016: Diet and locomotion, but not body size, differentiate mammal communities in worldwide tropical ecosystems. — *Palaeogeography, Palaeoclimatology, Palaeoecology* 454: 20–29.
- Lintulaakso, K., Tatti, N. & Žliobaitė, I. 2023: Quantifying mammalian diets. — Mammalian Biology 103: 53–67.
- Lintulaakso, K. & Kovarovic, K. 2023: Continent-wide patterns in mammal community structure: diet, locomotion, and body mass. — In: Casanovas-Vilar, I., van den Hoek Ostende, L. W., Janis, C. M. & Saarinen, J. (eds.), Evolution of Cenozoic land mammal faunas and ecosystems: 25 Years of the NOW Database of Fossil Mammals: 95–110. Springer International Publishing, Cham.
- Liow, L. H., Fortelius, M., Bingham, E., Lintulaakso, K., Mannila, H., Flynn, L. & Stenseth, N. C. 2008: Higher origination and extinction rates in larger mammals. — *Proceedings of the National Academy of Sciences* 105: 6097–6102.

- Liu, L., Puolamäki, K., Eronen, J. T., Ataabadi, M. M., Hernesniemi, E. & Fortelius, M. 2012: Dental functional traits of mammals resolve productivity in terrestrial ecosystems past and present. — *Proceedings of the Royal Society B* 279: 2793–2799.
- Lovegrove, B. G. & Haines, L. 2004: The evolution of placental mammal body sizes: evolutionary history, form, and function. — *Oecologia* 138: 13–27.
- Lyons, S. K. & Smith, F. A. 2013: Macroecological patterns of body size in mammals across time and space. — In: Smith, F. A. & Lyons, S. K. (eds.), *Animal body size: linking pattern and process across space, time, and taxonomic group*: 116–144. University of Chicago Press, Chicago.
- Malhi, Y., Doughty, C. E., Galetti, M., Smith, F. A., Svenning, J.-C. & Terborgh, J. W. 2016: Megafauna and ecosystem function from the Pleistocene to the Anthropocene. — *Proceedings of the National Academy of Sciences* 113: 838–846.
- Martin, P. S. & Klein, R. G. 1989: *Quaternary extinctions: a prehistoric revolution.* University of Arizona Press, Tuscon, AZ.
- Martin, L. J., Blossey, B. & Ellis, E. 2012: Mapping where ecologists work: biases in the global distribution of terrestrial ecological observations. — *Frontiers in Ecology* and the Environment 10: 195–201.
- Maurer, B. A. 1998: The evolution of body size in birds. II. The role of reproductive power. — *Evolutionary Ecology* 12: 935–944.
- Maurer, B. A., Brown, J. H. & Rusler, R. D. 1992: The micro and macro in body size evolution. — *Evolution* 46: 939–953.
- McKenna, M. C., Bleefield, A. R. & Mellett, J. S. 2005: Microvertebrate collecting: large-scale wet sieving for fossil microvertebrates in the field. — In: Leiggi, P. & May, P. (eds.), Vertebrate paleontological techniques, vol. 1: 93–112. Cambridge University Press, Cambridge.
- Mechenich, M. F. & Žliobaitė, I. 2023: Eco-ISEA3H, a machine learning ready spatial database for ecometric and species distribution modeling. — *Scientific Data* 10: 77, https://doi.org/10.1038/s41597-023-01966-x.
- Meloro, C. & Sansalone, G. 2022: Palaeoecological significance of the "wolf event" as revealed by skull ecometrics of the canid guilds. — *Quaternary Science Reviews* 281, 107419, https://doi.org/10.1016/j.quascirev.2022.107419.
- Monsarrat, S. & Kerley, G. I. 2018: Charismatic species of the past: biases in reporting of large mammals in historical written sources. — *Biological Conservation* 223: 68–75.
- Morgan, M. E., Badgley, C., Gunnell, G. F., Gingerich, P. D., Kappelman, J. W. & Maas, M. C. 1995: Comparative paleoecology of Paleogene and Neogene mammalian faunas: body-size structure. — *Palaeogeography, Palae*oclimatology, *Palaeoecology* 115: 287–317.
- Nanglu, K. & Cullen, T. M. 2023: Across space and time: A review of sampling, preservational, analytical, and anthropogenic biases in fossil data across macroecological scales. — *Earth-Science Reviews* 244, 104537, https://doi.org/10.1016/j.earscirev.2023.104537.

- Niku, J., Hui, F. K. C., Taskinen, S. & Warton, D. I. 2019: gllvm: fast analysis of multivariate abundance data with generalized linear latent variable models in R. — *Meth*ods in Ecology and Evolution 10: 2173–2182.
- Oksanen, O., Žliobaitė, I., Saarinen, J., Lawing, A. M. & Fortelius, M. 2019: A Humboldtian approach to life and climate of the geological past: estimating palaeotemperature from dental traits of mammalian communities. *— Journal of Biogeography* 46: 1760–1776.
- Olalla-Tárraga, M. Á., Rodríguez, M. Á. & Hawkins, B. A. 2006: Broad-scale patterns of body size in squamate reptiles of Europe and North America. — *Journal of Biogeography* 33: 781–793.
- Pacifici, M., Rondinini, C., Rhodes, J. R., Burbidge, A. A., Cristiano, A., Watson, J. E. M., Woinarski, J. C. Z. & Di Marco, M. 2020: Global correlates of range contractions and expansions in terrestrial mammals. — *Nature Communications* 11, 2840, https://doi.org/10.1038/s41467-020-16684-w.
- Parfitt, S. A. & Harrison, D. L. 2011: New material of the shrew Macroneomys Fejfar, 1966 (Mammalia, Soricomorpha, Soricidae) from the British early Middle Pleistocene, with comments on its palaeobiology and European range. — Acta Zoologica Cracoviensia A 54: 31–37.
- Parker, A. K., Müller, J., Boisserie, J.-R. & Head, J. J. 2023: The utility of body size as a functional trait to link the past and present in a diverse reptile clade. — *Proceedings of the National Academy of Sciences* 120, e2201948119, https://doi.org/10.1073/pnas.2201948119.
- Peláez-Campomanes, P. & Meulen, A. J. 2009: Diversity of mammals in the Neogene of Europe: comparing data quality of large and small mammals in the NOW database. — *Hellenic Journal of Geosciences* 44: 105–115.
- Pineda-Munoz, S. & Alroy, J. 2014: Dietary characterization of terrestrial mammals. — *Proceedings of the Royal Society B* 281, 20141173, https://doi.org/10.1098/ rspb.2014.1173.
- Polly, P. D. & Sarwar, S. 2014: Extinction, extirpation, and exotics: effects on the correlation between traits and environment at the continental level. — *Annales Zoologici Fennici* 51: 209–226.
- Polly, P. D. & Eronen, J. T. 2011. Mammal associations in the Pleistocene of Britain: implications of ecological niche modelling and a method for reconstructing palaeoclimate. — In: Ashton, A., Lewis, S. G. & Stringer, C. (eds.), *The ancient human occupation of Britain*, 279–204. Elsevier, Amsterdam.
- Polly, P. D. & Head, J. J. 2015: Measuring earth-life transitions: ecometric analysis of functional traits from late Cenozoic vertebrates. — In: Polly, P. D., Head, J. J. & Fox, D. L. (eds.), *Earth-life transitions: paleobiology in the context of Earth system evolution*: 21–46. The Paleontological Society, Cambridge.
- Prescott, G. W., Williams, D. R., Balmford, A., Green, R. E. & Manica, A. 2012: Quantitative global analysis of the role of climate and people in explaining late Quaternary megafaunal extinctions. — *Proceedings of the National Academy of Sciences* 109: 4527–4531.
- Raup, D. M. 1979: Biases in the fossil record of species and

genera. — Bulletin of the Carnegie Museum of Natural History 13: 85–91.

- Reif, J. 2013: Long-term trends in bird populations: a review of patterns and potential drivers in North America and Europe. — Acta Ornithologica 48: 1–16.
- Reuter, D. M., Hopkins, S. S. B. & Price, S. A. 2023: What is a mammalian omnivore? Insights into terrestrial mammalian diet diversity, body mass and evolution. — *Proceedings of the Royal Society B* 290, 20221062, https:// doi.org/10.1098/rspb.2022.1062.
- Rolland, J., Condamine, F. L., Jiguet, F. & Morlon, H. 2014: Faster speciation and reduced extinction in the tropics contribute to the mammalian latitudinal diversity gradient. — *PLoS Biology* 12(1), e1001775, https://doi. org/10.1371/journal.pbio.1001775.
- 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? — Evolutionary Ecology Research 12: 363–376.
- Saltré, F., Rodríguez-Rey, M., Brook, B. W., Johnson, C. N., Turney, C. S. M., Alroy, J., Cooper, A., Beeton, N., Bird, M. I., Fordham, D. A., Gillespie, R., Herrando-Pérez, S., Jacobs, Z., Miller, G. H., Nogués-Bravo, D., Prideaux, G. J., Roberts, R. G. & Bradshaw, C. J. A. 2016: Climate change not to blame for late Quaternary megafauna extinctions in Australia. — *Nature Communications* 7, 10511, https://doi.org/10.1038/ncomms10511.
- Sandom, C., Faurby, S., Sandel, B. & Svenning, J.-C. 2014: Global late Quaternary megafauna extinctions linked to humans, not climate change. — *Proceedings of the Royal Society B* 281, 20133254, https://doi.org/10.1098/ rspb.2013.3254.
- Smith, A. B. 2001: Large-scale heterogeneity of the fossil record: implications for Phanerozoic biodiversity studies. — *Philosophical Transactions of the Royal Society* of London B 356: 351–367.
- Smith, F. A., Smith, R. E. E., Lyons, S. K. & Payne, J. L. 2018: Body size downgrading of mammals over the late Quaternary. — *Science* 360: 310–313.
- Smith, F. A., Lyons, S. K., Ernest, S. K. M., Jones, K. E., Kaufman, D. M., Dayan, T., Marquet, P. A., Brown, J. H. & Haskell, J. P. 2003: Body mass of late Quaternary mammals. — *Ecology* 83: 3403–3403, https://doi. org/10.1890/02-9003.
- Smith, J., Rillo, M. C., Kocsis, A. T., Dornelas, M., Fastovich, D., Huang, H.-H. M., Jonkers, L., Kiessling, W., Li, Q., Liow, L. H., Margulis-Ohnuma, M., Meyers, S., Na, L., Penny, A. M., Pippenger, K., Renaudie, J., Saupe, E. E., Steinbauer, M. J., Sugawara, M., Tomašových, A., Williams, J. W., Yasuhara, M., Finnegan, S. & Hull, P. M. 2023: BioDeepTime: a database of biodiversity time series for modern and fossil assemblages. *Global Ecology and Biogeography* 32: 1680–1689.
- Tougard, C. & Montuire, S. 2006: Pleistocene paleoenvironmental reconstructions and mammalian evolution in South-East Asia: focus on fossil faunas from Thailand. — *Quaternary Science Reviews* 25: 126–141.
- Travouillon, K. J. & Legendre, S. 2009: Using cenograms to investigate gaps in mammalian body mass distributions in Australian mammals. — *Palaeogeography, Palaeocli-*

matology, Palaeoecology 272: 69-84.

- Turvey, S. T. & Blackburn, T. M. 2011: Determinants of species abundance in the Quaternary vertebrate fossil record. — *Paleobiology* 37: 537–546.
- van Buskirk, J. 2003: Habitat partitioning in European and North American pond-breeding frogs and toads. — Diversity and Distributions 9: 399–410.
- Veblen, K. E., Porensky, L. M., Riginos, C. & Young, T. P. 2016: Are cattle surrogate wildlife? Savanna plant community composition explained by total herbivory more than herbivore type. — *Ecological Applications* 26: 1610–1623.
- Wan, X. & Zhang, Z. 2017: Climate warming and humans played different roles in triggering Late Quaternary extinctions in east and west Eurasia. — *Proceedings* of the Royal Society B 284, 20162438, http://doi. org/10.1098/rspb.2016.2438.
- Willemsen, G. F. 1992: A revision of the Pliocene and Quaternary Lutrinae from Europe. — *Scripta Geologica* 101: 1–115.
- Wilman, H., Belmaker, J., Simpson, J., de la Rosa, C., Rivadeniera, M. M. & Jetz, W. 2014. Eltontraits 1.0: species-level foraging attributes of the world's birds and mammals. — *Ecology* 95: 2027–2027, https://doi. org/10.1890/13-1917.1.
- Wilson, O. E., Fortelius, M. & Saarinen, J. 2022: Species discovery and dental ecometrics: good news, bad news and recommendations for the future. — *Historical Biology* 35: 678–692.
- Wilson, O. E. & Parker, A. K. 2023: Low predator competition indicates occupation of macro-predatory niches by giant Miocene reptiles at La Venta, Colombia. — Palaeogeography, Palaeoclimatology, Palaeoecology 632, 111843, https://doi.org/10.1016/j.palaeo.2023.111843.
- Wolff, R. G. 1975: Sampling and sample size in ecological analyses of fossil mammals. — *Paleobiology* 1: 195–204.
- Žliobaitė, I. & Fortelius, M. 2022: On calibrating the completometer for the mammalian fossil record. — *Paleobi*ology 48: 1–11.
- Žliobaitė, I., Tang, H., Saarinen, J., Fortelius, M., Rinne, J. & Rannikko, J. 2018: Dental ecometrics of tropical Africa: linking vegetation types and communities of large plant-eating mammals. — *Evolutionary Ecology Research* 19: 127–147.
- Žliobaitė, I., Rinne, J., Toth, A. B., Mechenich, M., Liu, L., Behrensmeyer, A. K. & Fortelius, M. 2016: Herbivore teeth predict climatic limits in Kenyan ecosystems. — *Proceedings of the National Academy of Sciences* 113: 12751–12756.
- Žliobaite, I., Fortelius, M., Bernor, R. L., van den Hoek Ostende, L., Janis, C. M., Lintulaakso, K., Säilä, L. K., Werdelin, L., Casanovas-Vilar, I., Croft, D., Flynn, L., Hopkins, S. S. B., Kaakinen, A., Kordos, L., Kostopoulos, D. S., Pandolfi, L., Rowan, J., Tesakov, A., Vislobokova, I., Zhang, Z., Aiglstorfer, M., Alba, D. M., Arnal, M., Antoine, P. O., Belmaker, M., Bilgin, M., Boisserie, J.-R., Borths, M., Cooke, S. B., Dam, J. van, Delson, E., Eronen, J. T., Fox, D., Furió, A. F. M., Giaourtsakis, I. X., Holbrook, L., Hunter, J., López-Tor-

res, S., Ludtke, J., Minwer-Barakat, R., van der Made, J., Mennecart, B., Pushkina, D., Rook, L., Saarinen, J., Samuels, J. X., Sanders, W., Silcox, M. & Vepsäläinen, J. 2023: The NOW Database of Fossil Mammals. — In: Casanovas-Vilar, I., van den Hoek Ostende, L. W., Janis, C. M. & Saarinen, J. (eds.), *Evolution of Cenozoic* *land mammal faunas and ecosystems*: 33–42. Vertebrate Paleobiology and Paleoanthropology, Springer, Cham, https://doi.org/10.1007/978-3-031-17491-9 3.

Zuo, W., Smith, F. A. & Charnov, E. L. 2013: A life-history approach to the Late Pleistocene megafaunal extinction. — *The American Naturalist* 182: 524–531.



Appendix 1. The observed genus count for each of 1018 sites with minimum ages under 2.58 Ma, plotted against the site age range (maximum age – minimum age). The bases for these ages are listed in the NOW database (https://doi.org/10.5281/zenodo.4268068) and in the occurrence file available on GitHub (https://github.com/abbaparker/Size-Diet-Bias).



Appendix 2. Body size distributions of modern mammal species. $- \mathbf{A}$: All species occurring globally in the Phylacine data set, with species for which mass was imputed removed. $- \mathbf{B}$: The subset of those species occurring in Eurasia. Both histograms fit with a bimodal normal distribution where both distributions have equal standard deviations, with *y*-axis scaling for the proportion of species occurring in each distribution (*see* Table 1).



Appendix 3. Mean body size for all 1018 fossil localities, averaged across occurring genera. The same histogram subsampled to sites with four or more genera is shown in Fig. 3.



Appendix 4. Fossil genus count vs. genus count at the nearest modern sample, for (A) large mammal genera, (B) herbivore genera, (C) small mammal genera, and (D) non-herbivore genera; see also Fig. 4.



Appendix 5. The number of mammal genera observed at (A) 1018 fossil sites in Eurasia, and (B) 21 002 modern sampling points across Eurasia, plotted against the latitude of their geographic locations. Site longitudes are indicated by the color of each point; see also Fig. 5.