scientific data

Check for updates

OPEN

DATA DESCRIPTOR

EutherianCoP. An integrated biotic and climate database for conservation paleobiology based on eutherian mammals

Alessandro Mondanaro[1]✉, Giorgia Girardi[2], Silvia Castiglione[2], Axel Timmermann[3,4], Elke Zeller[3,4], Thushara Venugopal[3,4], Carmela Serio[2], Marina Melchionna[2], Antonella Esposito[2], Mirko Di Febbraro[5] & Pasquale Raia[2]✉

We present a new database, EutherianCoP, of fossil mammals which lived globally from the Late Pleistocene to the Holocene. The database includes 13,972 fossil occurrences of 786 extant or recently extinct placental mammal species, plus 155,198 current occurrences for those of them which survived to the present. The occurrences are correlated with radiometric age information. For all species, we provide 32 different traits, inclusive of taxonomic, phenotypic, life history, biogeographic and phylogenetic information. Differently from any other compilation, the occurrences are complemented with estimates of past climatic conditions, including site-interpolated monthly and annual precipitation and temperature, leaf area index, megabiome type and net primary productivity, which are derived from transient paleo model simulations conducted with the Community Earth System Model 1.2 and the BIOME4 vegetation model. All data are further downloadable for further investigation.

Background & Summary

Conservation paleobiology is sought to provide insights about the potential fate of the living biota in the near future, by learning from the dynamics that affected past ecosystems and species[1,2]. Albeit using the fossil record to address conservation issues linked to climate change is generally well established[3], the potential of the discipline to help addressing other fundamental aspects in conservation, such as the effects of biome and habitat change, the impact of invasive species, and that of human overexploitation of species and space, is currently met with skeptism[4,5]. Consequently, most conservation paleobiology looks better on paper than in reality[6–8], and according to recent studies, students and early career scientists are much more confident about its utility than senior researchers[5]. And yet, near-time paleobiology (i.e., the discipline where conservation paleobiology roots) is as productive and rich as to inform about ecosystem resilience and reorganization after biomass fluctuation[9], major disturbances[10,11], and other impacts generated by humans or climate change[12–14]. It even contributes to the study of biodiversity change effects on carbon storage, on the amount of plant biomass eaten by primary consumers, and on biogeochemical cycles[15–17]. With such array of far-reaching, wide ranging study results[2], the limited faith in the potential of near time paleobiology to help conservation research, that is to become effective conservation paleobiology, seems perplexing to say the least, especially for terrestrial ecosystem[5].

Climate change is a powerful predictor of extinction risk across multiple temporal and spatial scales[18], and extensive, fine-grained climate data are now available for the near past[19–22]. Hence, learning from climate prehistory for studying conservation issues like extinction risk and climate-induced species dispersal is straightforward. With different, more complex questions, such as those related to ecosystem management, resilience, minimum protected area size and interspecies connectivity[23], things are much more complicated. One reason is that conservation scientists and practitioners are used to work with a plethora of data, including life-history

[1]Department of Earth Science, University of Florence, via G. La Pira 4, 50121, Florence, Italy. [2]DiSTAR, University of Naples "Federico II", 80126, via Vicinale Cupa Cintia 26, Naples, Italy. [3]IBS Center for Climate Physics, 46241, Busan, South Korea. [4]Pusan National University, 46241, Busan, South Korea. [5]EnviXLab, Department of Biosciences and Territory, University of Molise, 86090, Pesche (Isernia), Italy. ✉e-mail: alessandro.mondanaro@unifi.it; pasquale.raia@unina.it

traits, population density, genetic diversity, and biogeographical data that are hardly available for extinct species, given the punctuated, mostly erratic nature of the terrestrial fossil record. Even for simple metrics, producing usable variable estimates for extinct species can be challenging. For instance, the geographic range size is key to provide effective extinction scenarios[24,25]. Yet, given the scattered and idiosyncratic nature of the record[26], reconstructing the geographic ranges of fossil species is far from easy. The problem can be circumvented by delineating convex hulls around the fossil species occurrences or studying its co-occurrence patterns with coeval taxa[24,27,28]. More sharply, the potential range can be inferred using species distribution modeling[29] (SDM), which by mapping the species' habitat suitability across space can as also allow to estimate the effect of connectivity on extinction[30]. Phenotypic and phylogenetic data for fossil species are equally hard to come by, although they can be confidently estimated via allometric equations[31,32] and phylogenetic imputation[33]. Phylogenetic trees are themselves a fundamental tool in conservation science[34] as they allow to evaluate individual species assemblages in terms of evolutionary diversity[35], and to estimate the potential effect of competition (for limited resources) and climatic variability on species coexistence, a fruitful research field known as community phylogenetics[36]. Still, under some circumstances, phylogenetic diversity is a viable surrogate for functional diversity (which is the key feature conferring ecosystems resilience to disturbance)[37,38]. Assembling phylogenetic trees is a discipline on its own. Yet, grafting reasonably correct trees is now easier, thanks to easily manageable tools for producing formal and informal supertrees[39–41] and the availability of large phylogenetic trees repositories[42]. Phylogenetic data and trait data imputation[43], coupled with climatic and fossil occurrence records, may now provide viable tools to address the dynamics that affected past ecosystems, which is crucial to learn about the breadth of their resilience, and to understand the factors (and the intensity thereof) that favoured their formation, forced their compositional change, brought about their demise, and how all of that can be used to mitigate the future effects of climate change.

With so many different themes being relevant to conservation, and so much diverse data and facets of the species phenotype and ecology to consider, it is not surprising that large, comprehensive compilations of species traits for extant and recently extinct species are becoming common[44–49]. At least three large collections, PanTHERIA[44], and more recently and thoroughly COMBINE[50] and PHYLACINE[27], effectively include phylogenies, biogeographic and phenotypic data for nearly 6000 mammal species that lived during the last 130 ka. However complete, these databases lack climatic data and details associated to the fossil occurrences of extinct species (and extant species with fossil representative remains) and provide a limited coverage of mammal diversity in the recent past. Conversely, large compilations of the mammal fossil record, like the Paleobiology (https://paleobiodb.org) and The New and Old Worlds (NOW) databases (https://nowdatabase.org/), have very limited coverage of species traits, lack any phylogenetic information, and climate data. Here, we present a new extensive database, named EutherianCoP, meant to bridge the gap between these two kinds of compilations. EutherianCoP provides high-quality biotic (including brain and body size, life history traits, potential population density, trophic level and diet partitioned between different food types) and abiotic data (including coordinates and ages of fossil localities). All abiotic data are accompanied by fine-scale monthly and annual temperature and precipitation means, Net Primary Productivity (NPP), Leaf Area Index (LAI) and megabiome[51] type at the mean estimated age and place of the locality. The data cover 169,170 occurrences relating to nearly 800 large mammal species which lived during the last 130 ka. The goal is to deliver a ready to use, multifaceted database condensed in a single compilation, which may aid conservation paleobiology perspectives and studies by providing a comprehensive and informative snapshot of mammalian ecosystems in the past, at the global scale.

EutherianCoP comes with an application made to explore and download the occurrence (fossil sites with dating, location and climate) and the species traits data. The application was written in R using the package Shiny[52], which allows to create lightweight web applications for data visualization, exploration, and download, also suitable to construct large datasets.

Methods

The flowchart illustrating how EutherianCoP is assembled is shown in Fig. 1.

Occurrence records and dating. The earliest phase of the database building consisted in the collection of fossil records for 786 terrestrial mammal species lived during the Late Pleistocene epoch, from 130 ka to the late Holocene. To gather the fossil occurrences, we worked at global spatial scale including all known eutherian data (Fig. 2) expanding upon earlier compilations we published[53–55]. To our knowledge, there is no published fossil mammal occurrence dataset at this level of detail and coverage for the Late Pleistocene. The full list of fossil species spans from the megafauna to the small-size mammals including the following orders: Artiodactyla, Carnivora, Chiroptera, Cingulata, Didelphimorphia, Diprotodontia, Eulipotyphla, Hyracoidea, Lagomorpha, Litopterna, Macroscelidea, Perissodactyla, Pholidota, Pilosa, Primates, Proboscidea, Rodentia, Scandentia, and Soricomorpha. Overall, we collected 13,972 fossil occurrences.

We further enriched the fossil record by adding the modern occurrence of the species included in the dataset. To this aim, we downloaded the data from the "Global Biodiversity Information Facility" (GBIF; www.gbif.org/) online database by selecting "Human observation", "Observation" and "Occurrences" categories as "Basis of record" and by removing duplicates and records with unrealistic coordinates, as well as occurrences from natural history collections. After filtering procedures, the final version of dataset counts 169,187 occurrences overall (see Zenodo https://doi.org/10.5281/zenodo.14009105 to access the raw data).

EutherianCoP reports the most reliable age estimate for each fossil site (or layer within the site), by further adding the laboratory code (where available) for the dated material and the dating method to allow users estimate the data quality on herself. The dataset includes a total of 7936 radiometric data: Conv. ^{14}C (1690 items), AMS ^{14}C (5702 items), TL (80 items), OSL/IRSL (241 items), Electron Spin Resonance (122 items),

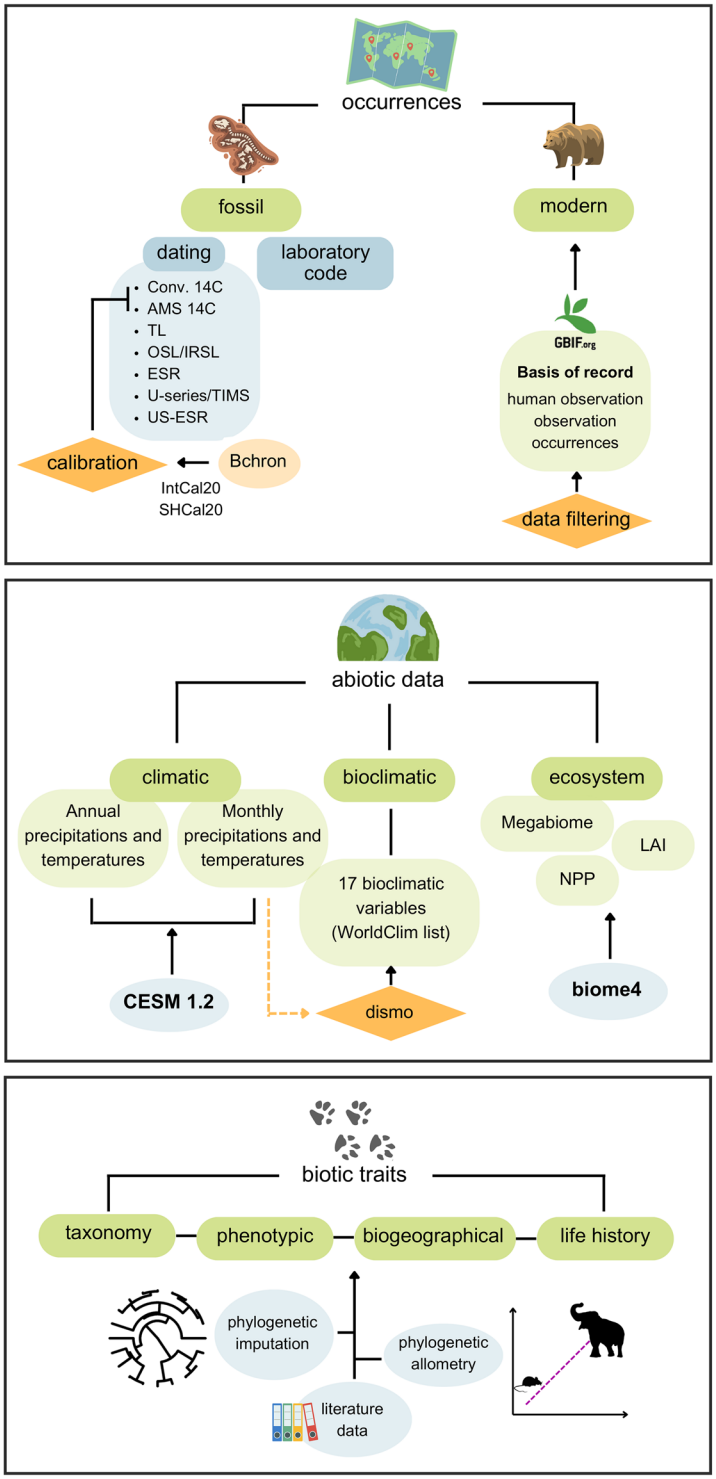

Fig. 1 Flowchart illustrating the assembly of EutherianCoP.

U-series/TIMS (115 items), combined ESR and U-series dating method (US-ESR, 42 items). A small number of date estimates (0.71%) derive from multiple dating methods estimates (e.g. TL and OSL estimates of the same stratigraphic layer). A very limited fraction of sites (0.6%) was dated through confident biochronological and/or magnetostratigraphic estimates. They were included despite the reference to non-radiometric age estimates since the age constraint on the fossil layers are especially tight in these cases.

To improve the dating quality, we removed the ages associated with reworked or intrusive materials, dates considered outliers by authors according to their chronostratigraphic models, and dates considered unrealistic because of high contamination of the dated material. We are aware that our approach is different from other published online databases. Since almost all fossil sites dated before 50 ka attested the coexistence of human and

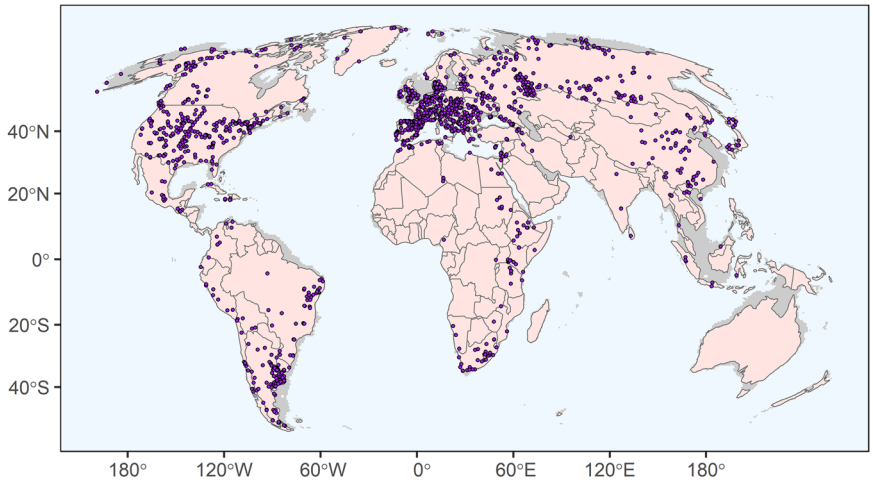

Fig. 2 Fossil eutherian mammals spatial distribution during the Late Pleistocene (130 ka to the late Holocene). Occurrences are indicated with purple points. Light pink area represents the terrestrial landscape over the last 130 ka whereas the grey area indicates the maximum extension of the landscape coastline.

faunal species within the same layers, and possibly the evidence of human-modified faunal bones, we checked some of the most common used human Palaeolithic databases as a control.

Recently, the last version of the "Radiocarbon Palaeolithic Europe Database" was published (https://ees.kuleuven.be/en/geography/projects/14c-palaeolithic/download). This database collapses a large amount of available radiometric data from literature incorporating newly published chronometric dates into a single coherent compilation. A similar approach was followed by the Neotoma database for the North America's faunal occurrences[56]. Contrary to both, we preferred not to incorporate all the available ages, focussing on the most accurate dates only. Accordingly, we accounted for the most rigorous pretreatment method in ^{14}C dating. It was demonstrated that the acid-base-oxidation stepped combustion (ABOx-SC[57,58]) provides more reliable age estimates than the traditional acid-base-acid (ABA) protocol when old charcoal material is dated. As an example, Higham *et al.*[59] showed ABA pre-treatment was not sufficient to remove contamination from charcoal materials occurred in Grotta di Fumane cave (Italy). In addition, the authors stated that "*older results obtained using ABOx–SC pre-treatment methods suggest that these ages are more reliable than the ABA-treated samples and probably comprise less contaminating carbon*". In contrast to Radiocarbon Palaeolithic Europe Database, we therefore excluded ABA dates (OxA-11360, OxA-11347 from level A2) maintaining only the ABOx–SC dates to better constraint the age of faunal assemblages recovered from the site[60].

Biological material such as bone, antler, teeth, and ivory yield the protein collagen that can be purified using multiple chemical steps that culminate in isolation and dating of XAD-purified total amino acids[61] or the specific amino acid hydroxyproline (HYP[62,63]). Recent advances in the chemical pretreatment and physical measurement increased the accuracy and precision of radiocarbon measurements and they were deeply applied to refine the chronostratigraphic relationship between the megafauna bones and the Clovis assemblages in North America[64]. Devièse *et al.*[65] analyzed the vertebrate fossils associated with some North American butchering sites stating that "*the use of XAD resins are currently the only efficient methods for removing environmental and museum-derived contaminants*". Accordingly, we excluded the ages derived from bone apatite or from incomplete ultrafiltration process if more accurate dates were available. And even in this case, we followed a different approach compared to Neotoma and the "Canadian Archaeological Radiocarbon Database (CARD)" (https://www.canadianarchaeology.ca/). For instance, XYP age derived from the mammoth bone in the Prele site (USA) was more strictly correlated with the Clovis age attribution, while the older ages generated with the standard ultrafiltration process gave significantly younger estimates[65]. In this case, CARD reported the full list of available dates for Prele while we preferred to maintain only the XYP age. Lastly, we excluded the infinite radiocarbon dates closed to the limit of the ^{14}C technique because they are often unrealistic or should be considered as minimum ages[66–68].

Since it was demonstrated that calibration against ^{14}C measurements is critical for providing a correction for changes in ^{14}C concentration within atmospheric and marine carbon reservoirs[69], all conventional dates BP (before present) were calibrated by adopting the "IntCal20" and "SHCal20" curves for Northern and Southern hemispheres, respectively[69], relying on the functionalities provided in the "*Bchron*" R package[70].

Climatic data. The second part of the database assembly consists in merging the occurrences (both fossil and extant) with climatic and ecosystem data[71]. To this aim, we retrieved the climatic and vegetational cover data originates from a 3 million year quasi-transient climate[20] and subsequential vegetation simulation[19]. The climate evolution of the past 3 million years was simulated using the Community Earth System Model (CESM) 1.2 with transient time-varying forcing of greenhouse gases (CO_2, N_2O, and CH_4)[72,73], Northern Hemispheric ice-sheets[72], and insolation conditions[74]. The model features an ocean and atmosphere resolution of approximately 3.75×3.75 degrees and utilizes the bathymetry of the Last Glacial Maximum (LGM). Due to the weaker climate sensitivity of CESM1.2 compared to paleoclimate estimates[75,76] and other earth system models[77], the CO_2 forcing was scaled by a factor of 1.5, consistent with other similar simulations[75,78,79]. In addition, the model was divided into 42

chunks and ran in parallel to reduce runtime. Each chunk starts from the same preindustrial condition, spans one interglacial/glacial cycle, and overlaps with the next to address initial conditions and spin-up time issues. And an orbital acceleration factor of 5 was used, compressing 3 million years of historical orbital data into 600,000 CESM model years, consistent with techniques in other paleoclimate model simulations. Subsequently, vegetation data was simulated by using BIOME4[80] relying on temperature, precipitation, and cloud coverage from the CESM simulation. To address model biases and issues arising from downscaling data, adjustments were made to the original pre-industrial biome data. These adjustments incorporated the difference between pre-industrial conditions and specific time slices of linearly interpolated CESM data for temperature and minimum temperature, and the percentage difference for precipitation and sunshine. The soil's percolation rate and water-holding capacity were kept constant at current levels, obtained from the original BIOME4 data. These data points were linearly interpolated where missing and extended by ten grid points in order to capture all land points.

Annual and monthly temperature and precipitation are available from the original 3 Ma CESM simulation at the closest location on a 3.75×3.75 grid and subsequently downscaled to 0.5×0.5 cell grid. As ecosystem variables, we selected Megabiome, Net Primary Productivity (NPP), and Leave Area Index (LAI) from the 3 Ma BIOME4 simulation at 0.5×0.5 resolution (https://climatedata.ibs.re.kr/data/3ma-transient-climate-simulation). Megabiome variable consists of eleven major biome types aggregated on the basis of their structure and functioning[51].

One potential use of EutherianCoP is to produce fossil data informed SDMs for mammal species. Since SDMs commonly use bioclimatic predictors[81] for most macroecological applications[82], we decided to extend the suite of climatic variables converting monthly 1000-year mean climatic data (as they come in CESM) into 17 (out of 19 overall) bioclimatic variables available from the WorldClim version 2.1 database (https://www.worldclim.org/data/bioclim.html[81]) using the "dismo" R package[83]. We excluded annual mean diurnal range (BIO2) and isothermality (BIO3) because for the CESM1.2 paleoclimate simulations sub-daily data were not saved. Overall, the full list of abiotic variables used for the analysis is available in the Table 1.

Very often, multiple dates are available for individual fossil sites or layers within a site, producing age uncertainty regardless of the rigour in accepting age estimates. To account for this uncertainty, we perform a post-hoc manipulation of the data by averaging all dates available for an individual site/layer or by using them singularly, either calibrated or uncalibrated. According to this approach, we generated four different worksheets (see below for further details).

Once the single (or averaged) calibrated (or uncalibrated) ages for site/layer were obtained, we rounded them to the 1 kiloyears (kyr) temporal resolution, that is the same of the reference climatic and ecosystem data in EutherianCoP. This step allowed us to extract the spatially and temporally climatic and ecosystemic values associated to each occurrence.

All R codes useful to generate and manipulate the data are provided as supplementary file (https://doi.org/10.5281/zenodo.13169679).

Taxonomy and phylogeny. For the species taxonomy, we used a different approach depending on the status of the species. We followed the IUCN taxonomy for the living species (www.iucnredlist.org). For the extinct species, we went through multiple sources for selecting the more updated taxonomic and species attribution. For some species, the taxonomy is still unresolved, and a number of taxonomic affiliations as they appear in the species list of individual fossil layer collections are either outdated or otherwise wrong. We provide a short list of examples to illustrate how we dealt with these issues.

The first example regards cave bears. Cave bears is an iconic group of species occurring in Late Pleistocene faunal assemblages from Europe and northern Asia. Students of cave bear remains identified several morphologically distinct species[84] generally included in either one of two large species groups, the 'large cave bears' including *Ursus spelaeus, U. eremus*, *U. kanivetz*, and *U. ingressus* found in Central and Eastern Europe and in the Urals, and the 'small cave bears' including *U. savini* and *U. rossicus* from East Europe, Urals, and Northern Siberia. Unfortunately, the phylogenetic relationships between these species and their identity are controversial, and the advent of genomic data possibly did confound the issue further. Large cave bear from Medvezhiya Cave (Urals) were attributed to *U. kanivetz* or *U. spelaeus kanivetz*, whereas the study of mitochondrial DNA showed that the large cave bear from the North Urals belongs to the *U. ingressus* haplogroup[85,86]. Since Medvezhiya Cave was the type locality of *U. spelaeus kanivetz*, the name *U. ingressus* was replaced by *U. kanivetz* according to the International Code of Zoological Nomenclature. In keeping with this, mitochondrial data recognized a nested *kanivetz-ingressus* group which is genetically distinct from the *U. spelaeus* group, that in turn includes the *U. spelaeus*, *U. ladinicus*, and *U. eremus*. Subsequent nuclear DNA analysis confirmed the genetic uniqueness and species status of the Ural large cave bear and corroborated the idea that all large cave bears of the Urals should be attributed to *U. kanivetz*[87]. However, nuclear DNA gave a completely different pattern of the cave bear phylogeny compared to mitochondrial DNA considering *U. kanivetz* as valid species sister to the European *U. spelaeus* and *U. ingressus* lineages[87]. Similarly, the small cave bear from Kizel Cave in the Middle Urals was originally assigned to *U. rossicus* but later the species was synonymized with *U. savini*[88]. Subsequent papers referred to *U. savini* for small cave bear from Urals although other researchers continued to use the name *U. rossicus*.

Therefore, we assigned to *U. kanivetz* the large cave bear materials from Urals (e.g. Medvezhiya cave, Prokoshev Cave, Secrets Cave, and Zapovednaya cave) which implies a separation from the European *spelaeus* and *ingressus* lineages, in keeping with nuclear DNA data. Regarding the small cave bear, we adopted the name *U. savini* in referring to the Urals remains, and synonymised *U. savini* and *U. rossicus*.

In North America, the mammoth lineage represents a chronospecies. Agenbroad[89] suggested the chronological progression of *Mammuthus meridionalis* into *M. columbi* and then the latter into *M. primigenius* and its insular form *M. exilis*. In contrast, some authors supported the validity of several mammoth morphotypes, *M. jeffersoni* and *M. imperator*[90]. Graham[91] rejected *M. jeffersonii* as a species distinct from *M. columbi*

variable acronym	full name	unit
Ecosystem variables		
LAI	Leaf Area Index	gC/m^2
Megabiome	Megabiome	
NPP	Net Primary Productivity	gC/m^2
Climatic variables		
PP	Annual mean Precipitation	mm/year
PPJAN	January Precipitation	mm
PPFEB	February Precipitation	mm
PPMAR	March Precipitation	mm
PPAPR	April Precipitation	mm
PPMAY	May Precipitation	mm
PPJUN	June Precipitation	mm
PPJUL	July Precipitation	mm
PPAUG	August Precipitation	mm
PPSEP	September Precipitation	mm
PPOCT	October Precipitation	mm
PPNOV	November Precipitation	mm
PPDEC	December Precipitation	mm
TS	Annual mean Temperature	°C
TSJAN	January Temperature	°C
TSFEB	February Temperature	°C
TSMAR	March Temperature	°C
TSAPR	April Temperature	°C
TSMAY	May Temperature	°C
TSJUN	June Temperature	°C
TSJUL	July Temperature	°C
TSAUG	August Temperature	°C
TSSEP	September Temperature	°C
TSOCT	October Temperature	°C
TSNOV	November Temperature	°C
TSDEC	December Temperature	°C
Bioclimatic variables		
bio1	Annual Mean Temperature	°C
bio4	Temperature Seasonality (standard deviation)	°C
bio5	Max Temperature of Warmest Month	°C
bio6	Min Temperature of Coldest Month	°C
bio7	Temperature Annual Range (bio5-bio6)	°C
bio8	Temperature Annual Range (bio5-bio6)	°C
bio9	Mean Temperature of Driest Quarter	°C
bio10	Mean Temperature of Driest Quarter	°C
bio11	Mean Temperature of Coldest Quarter	°C
bio12	Annual Precipitation	mm
bio13	Precipitation of Wettest Month	mm
bio14	Precipitation of Driest Month	mm
bio15	Precipitation Seasonality (Coefficient of Variation)	mm
bio16	Precipitation of Wettest Quarter	mm
bio17	Precipitation of Driest Quarter	mm
bio18	Precipitation of Warmest Quarter	mm
bio19	Precipitation of Coldest Quarter	mm

Table 1. Full list of climatic, ecosystem, and bioclimatic variables.

considering the former as a 'progressive' form with increased lamellar number and frequency, and thinner enamel. McDaniel & Jefferson[92] corroborated the model proposed by Agenbroad[89] considering *M. imperator* and *M. columbi* conspecific based on their dental variation in molars. In addition, Enk *et al.*[93] demonstrated that mitochondrial DNA is not compatible with existing systematic interpretations of mammoth paleontological record showing an extensive interbreeding between *M. primigenius*, *M. jeffersonii* and *M. exilis*. Until the availability of new morphological and genetic data which can better defined the mammoth taxonomic boundaries, we

preferred to consider a single *M. columbi* group, merging the materials the fossil material ascribed to columbian mammoth and its variants onto a single species name[94].

In Africa, the fossil species name *Orycteropus crassidens* was used to ascribe the material from Rusinga Island and Kanjera (Kenya) albeit the relationships with the living *O. afer* are still unclear. Unfortunately, the systematics of fossil Tubulidentata is poorly known, and phylogenetic information is extremely limited. Pickford[95] suggested merging both fossil and living specimens in the *O. afer* and Lehmann *et al.*[96] supported this view but also noticed some distinctive characters between the two morphotypes that should be investigated further. To fill this gap, Lehmann[97] performed a new cladistic analysis by using the *O. crassidens* holotype from Rusinga. He identified a unique set of *O. crassidens* apomorphies and supported the elevation of *O. crassidens* to the species level. Accordingly, we ascribed the lower Pleistocene material from Kenya to *O. crassidens*.

Some of the major nomenclature issues are related to the taxonomical relationships between *Equus burchellii* and *E. quagga*, as well as *E. caballus* versus *E. ferus*, *E. przewalskii* or *E. ferus przewalski*. Genetic analysis reclassified *E. burchelli* as *E. quagga burchelli*, considering Burchell's zebra a subspecies of the historically extinct *E. quagga quagga*[98]. In addition, more recent DNA results suggested limited reproductive isolation amongst Cape quaggas and plains zebras[99]. Overall, *E. quagga* has priority over *E. burchellii* according to the International Committee on Bionomenclature (ICB)[100] although the taxonomical rules are often hard to apply with fossil record. Indeed, many authors noted that "quagga" is not known in East Africa where "burchelli" was accepted for long time in the past[100].

The *E. caballus*-*E. ferus* dichotomy is even more controversial because of the eighteenth century habit of giving different Latin binominals to domesticated forms of a species and their wild ("ferus" in this case) ancestors. In the absence of formal decision provided by the Nomenclature Commission[101] and given the unavailability of a recent revision of the whole genus *Equus*, it is hard to establish whether the Przewalski's horse should be seen as a subspecies of *Equus ferus* or as a separate species *E. przewalskii*[99].

To avoid further conflicts about the *Equus* taxonomy, we preferred to follow the current ICB recommendations grouping *E. quagga* and *E. burchelli* into a single, *E. quagga* name. Similarly, we merged the *E. caballus*, *E. ferus*, and *E. przewalskii* fossil occurrences under a single clade *E. ferus* separating only the occurrences belonging to the *E. f. caballus*, the domestics form. The same approaches illustrated in these examples was followed thorughtout the data, giving priority to recognized authorities and to more recent papers in resolving taxonomic conflicts.

In addition to scientific species name, we provided information about the Order and Family for both extinct and extant species. The full list of taxonomic criteria adopted is available as Supporting Information.

We assembled a phylogenetic tree of eutherian mammals considering all the species included in the dataset. The tree is largely based on refs. [102–104] and the recent species-level phylogeny of mammals[105] completed with numerous additional sources for individual species as detailed in the supplementary information.

Trait data. The third part of EutherianCoP build-up regards the trait data and includes 32 variables (Table 2), partitioned into 'taxonomic', 'phenotypic' (e.g. brain and body sizes, diet, trophic level) and 'life history' (e.g. longevity, interbirth interval, age at sexual maturity, litter size) and biogeographical (e.g. range size, dispersal distance, population density). Although a large number of traits are estimated, they are based on two fundamental sources of information only: body size and phylogenetic position. Body size estimates were collected from the scientific literature and our own previous compilations. We retrieved the size of the largest individual or the averaged size among several individuals when multiple estimates were available. For fossil species, body size is usually estimated from linear regression equations correlating the body size of the individuals with individual bone or teeth measurements[31,106,107]. This is the case with our compilation except for a minority of data for which body size was estimated by applying phylogenetic imputation[33]. Phylogenetic imputation was further used to estimate missing data for diet and digestive physiology (based on refs. [48,50,108]).

Body size data are invaluable since several phenotypic traits can be calculated starting from body size estimation via allometric equations[109,110]. We used phylogenetic allometric equations to impute missing data for life history traits (lifespan, age at maturity), biogeographical traits such as home range, geographic range size and minimum and maximum population density (based on refs. [45,111]), and ecological traits such as basal and field metabolic rate (based on ref. [17]), and minimum and maximum prey size for carnivores (based on ref. [112]).

Missing trait data values were estimated under the Brownian Motion model of evolution, by using the function *phylopars* in the "Rphylopars" R package[113] and *pglmm* in the package "phyr"[114]. We compared the two approaches (phylogenetic imputation and phylogenetic allometry) in terms of Akaike Information Criterion (AIC), selecting the model that performed best (Supporting information Table 1).

In particular, we produced dietary category data including the percentage of vegetable matter, invertebrate food, mammal food and non-mammal food (further collapsed in the 'vertebrate food' category), and 'animal food' (referring to animal sources as food that cannot be addressed to any other category such as blood or unidentifiable carrion). These source data for the mammal diet and categories comes from a large collection of direct observation on mammal feeding habits published in ref. [108]. To impute missing diets, we applied phylogenetic imputation under Brownian motion using Rphylopars[113] and adding body size as covariate in the reconstruction. We further provided a second dietary category, still using ref. [108]. as a source to classify species into general dietary categories (e.g. Browser, Grazer, Frugivore, Granivore, Carnivore etc.) by classifying missing data using the Canonical Variate Analysis (CVA). CVA was implemented using body size, the percentage of each food type consumed and phylogenetic eigenvector decomposition[115] axes as predictors of the dietary category.

To impute missing data for all other variables except prey size (see below) we devised a direct comparison between phylogenetic imputation and phylogenetic allometry, adjusted for taxonomic classification. The routine, condensed in the function 'phyloallometry' (available as supplementary information), computes missing data estimates twice by using Rphylopars[113] and phylogenetic generalized linear mixed model (*pglmm*) in "phyr".

In Rphylopars, after computing the missing data we regressed the imputed data vector (including both observed and estimated data) against body size and the taxonomic categories to which the species belong (once using families and a second time using orders) plus a dummy indicating whether the single datapoint is observed or imputed either. In *pglmm*, we first calculated allometric estimates using phylogeny and taxonomic level as random factors and body size as fixed factor, using observed data only. This produces estimates for observed data which depend on taxonomy and phylogeny, besides allometry (i.e. body size). The estimates were used to produce a new linear regression to apply to the missing data (*pglmm* does not allow to predict missing data).

Eventually for both *Rphylopars* and *pglmm* the procedure provides an allometric equation models (1):

$$y \sim body\ size * imputation\ dummy + taxonomy \quad (1)$$

which were compared to each other by means of AIC. Overall, four different models are produced, either using *Rphylopars* or *pglmm* and including families or orders as taxonomic level.

For prey size only, restricted to carnivores, we used field observation of carnivorous mammal feeding habits focusing on prey size and gut capacity reported in ref. [112]. We retrieved the minimum, median and maximum prey size per predator from the data and produced linear regression models against the predator size for any of them. The linear regression models were then used to estimate prey size (minimum, median and maximum) for all the Carnivora species in the dataset. The raw species trait data and code to replicate the missing data estimation procedures are available at https://doi.org/10.5281/zenodo.13169678[116].

Data Records

The dataset is available for download at Zenodo https://zenodo.org/records/14009105[116].

The data are articulated in six different files:

1. Source occurrence record include:
    - Raw occurrence data.xlsx: the file includes two different worksheets. The first reports the faunal list of species for each site or layer within the site, the 'doi' address for reference and the reference itself, the status (whether the species is extinct or extant). Further data included are geographical coordinates and country. Sheet 2 includes dating estimates along with the dating method, and the material dated (along with lab code where available).
2. Source phenotypic data include:
    - traits.xlsx: the file includes four different datasets: 'taxonomic', 'phenotypic', 'biogeographical', and 'life history' species traits.
    - eutherian tree.txt: phylogenetic tree of the species included in the database in Newick format.
    - sources for trait data.xlsx: a reference list for trait data.
3. A 'locality data.rar' file includes:
    - calibrated combined.xlsx: the file reports a single averaged and calibrated date for each site or layer within the site.
    - calibrated multiple.xlsx: the file reports all the available calibrated dates for each site or layer within the site.
    - uncalibrated combined.xlsx: the file reports a single averaged and uncalibrated date for each site or layer within the site.
    - uncalibrated multiple.xlsx: the file reports all the available uncalibrated dates for each site or layer within the site.

These four files are completed with climatic, biome and geographic data. In particular, the geographic occurrence data are associated with climatic (annual and monthly precipitation and temperature), ecosystem (megabiome, LAI, and NPP), and bioclimatic (the 17 WordClim variables) values, in keeping with their age are associated to the occurrence spreadsheet downloads or fully navigable within the app (5).

4. Phenotypic, life history and biogeographic traits and the tree including all the species appearing in EutherianCop inclusive of imputed species data are illustrated in:
    - traits imputed.xlsx: the file includes all the taxonomic, phenotypic, life history and biogeographic data for each species, both original and imputed.
5. The shiny app to illustrate data distribution and download. This 'EutherianCop' app runs in the R environment.

In addition to these datafiles, instructions and scripts for data imputations are provided at zenodo https://doi.org/10.5281/zenodo.14009105[116].

Variable	Trait category
Order	taxonomic
Family	taxonomic
Genus	taxonomic
Species	taxonomic
Body mass (grams)	phenotypic
Basal metabolic rate (Kj day^{-1})	phenotypic
Field metabolic rate (Kj day^{-1})	phenotypic
Dietary category	phenotypic
% animal food in the diet	phenotypic
% invertebrate food in the diet	phenotypic
% mammalian food in the diet	phenotypic
% minerals in the diet	phenotypic
% non animal food in the diet	phenotypic
% vegetable matter in the diet	phenotypic
% vertebrate food in the diet	phenotypic
maximum prey size (grams)	phenotypic
mean prey size (grams)	phenotypic
minimum prey size (grams)	phenotypic
age at maturity (days)	life history
age at first reproduction (days)	life history
generation length (days)	life history
gestation length (days)	life history
interbirth interval (days)	life history
litter size	life history
litter(s) per year	life history
maximum longevity (days)	life history
density (n individuals km^2)	biogeographical
maximum density (n individuals km^2)	biogeographical
minimum density (n individuals km^2)	biogeographical
dispersal distance (km)	biogeographical
home range (km^2)	biogeographical

Table 2. Full list of taxonomic, phenotypic, biogeographical, and life history trait variables.

Technical Validation

We carefully examined the data for consistency, including taxonomy. Quality control included plotting the data first to visually detect outliers. For phylogenetic allometric equations, we tested whether subsampling the data originates allometric slopes different from those originally used to estimate the missing data by using the functions embedded in the package 'emmeans'[117]. Since our focus is on late Pleistocene terrestrial mammals, we further ensured the validity biogeographic of data by verifying that all datapoints did not fall in the marine realm, which may be the case under incorrect reporting of geographical coordinates for the fossil sites or typos. This lack of accuracy could be fatal for sites placed near the coastline where a small coordinates imprecision can result in a spatial miss-matching between the datapoints and the abiotic values. In the worst-case scenario, where a datapoint erroneously falls in the sea rather than on lands, the extraction of abiotic values is incorrect because our procedure associated missing value to the fossil site. To overcome this issue, we applied an upgraded version of the "*fix.coastal.points*" function embedded in the 'EcoPast' R package (https://github.com/francesco-carotenuto/EcoPast). This version of *fix.coastal.points* includes a series of improvements mainly addressed to work with SpatRaster (a spatially referenced surface divided into three dimensional cells) and simple feature (a geographic features made of mostly two-dimensional geometries like points or polygons) objects. We adopted *fix.coastal.points* to move datapoints to the nearest non-NA cell within a maximum distance of 5 cells by considering both climatic and ecosystem data as raster grids. We emphasize that our gridded abiotic reconstructions accounted for the past dynamic evolution of the global coastline which slightly differs among the BIOME4 and CESM simulations. However, we did not use the current Earth digital elevation models (DEM) to run *fix.coastal.points* rather we selected the SpatRaster object chronologically associated to each fossil occurrences before applying the correction.

To assess the quality of data imputation, we started by inspecting the data collected from literature. Since all variables correlate to some extent to body size, we first ran a regression of the trait value against body size, and checked one by one all the fitted trait values which are 0.30 units apart from the regression line on the log10 scale (that is the observed value is either twice or one half of the fitted value). For these data, the original sources were consulted in order to confirm the suspicious datapoint for inclusion. After discarding the data which we had

no confirmation for or evidence of dubious data quality, we started the imputation process. We estimated the missing trait values both using phylogenetic imputation and phylogenetic allometry.

Code availability

The EutherianCop applications, along with the codes implemented to manipulate the raw data for download generating the spreadsheets described in the Data Record section are available at https://doi.org/10.5281/zenodo.14009105[116].

Received: 5 August 2024; Accepted: 28 November 2024;
Published online: 13 January 2025

References

1. Dietl, G. P. & Flessa, K. W. Conservation paleobiology: putting the dead to work. *Trends Ecol Evol* **26**, 30–37 (2011).
2. Dietl, G. P. *et al*. Conservation Paleobiology: Leveraging Knowledge of the Past to Inform Conservation and Restoration. *Annu Rev Earth Planet Sci* **43**, 79–103 (2015).
3. Kiessling, W., Smith, J. A. & Raja, N. B. Improving the relevance of paleontology to climate change policy. *Proc Natl Acad Sci USA* **120**, e2201926119 (2023).
4. Dietl, G. P., Smith, J. A. & Durham, S. R. Discounting the Past: The Undervaluing of Paleontological Data in Conservation Science. *Front Ecol Evol* **7**, 108 (2019).
5. Kiessling, W., Raja, N. B., Roden, V. J., Turvey, S. T. & Saupe, E. E. Addressing priority questions of conservation science with palaeontological data. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20190222 (2019).
6. Walker, S. E. Conservation biology and conservation paleobiology meet the Anthropocene together: history matters. *Front Earth Sci (Lausanne)* **11**, 1166243 (2023).
7. Dillon, E. M. *et al*. What is conservation paleobiology? Tracking 20 years of research and development. *Front Ecol Evol* **10**, 1031483 (2022).
8. Wingard, G. L., Schneider, C., Dietl, G. P. & Fordham, D. Turning Setbacks into Stepping-stones for Growth in Conservation Paleobiology. *Front Ecol Evol* https://doi.org/10.3389/fevo.2024.1384291 (2024).
9. Bibi, F. & Cantalapiedra, J. L. Plio-Pleistocene African megaherbivore losses associated with community biomass restructuring. *Science* **380**, 1076–1080 (2023).
10. Tóth, A. B. *et al*. Reorganization of surviving mammal communities after the end-Pleistocene megafaunal extinction. *Science* **365**, 1305–1308 (2019).
11. Smith, F. A. *et al*. Late Pleistocene megafauna extinction leads to missing pieces of ecological space in a North American mammal community. *Proceedings of the National Academy of Sciences* **119**, e2115015119 (2022).
12. Pineda-Munoz, S., Wang, Y., Kathleen Lyons, S., Tóth, A. B. & McGuire, J. L. Mammal species occupy different climates following the expansion of human impacts. *Proc Natl Acad Sci USA* **118** (2021).
13. Lyons, S. K. *et al*. Holocene shifts in the assembly of plant and animal communities implicate human impacts. *Nature* **529**, 80–83 (2016).
14. Davoli, M. *et al*. Megafauna diversity and functional declines in Europe from the Last Interglacial to the present. *Global Ecology and Biogeography* **33**, 34–47 (2024).
15. Smith, F. A., Elliott, S. M. & Lyons, S. K. Methane emissions from extinct megafauna. *Nat Geosci* **3**, 374–375 (2010).
16. Doughty, C. E. *et al*. Megafauna extinction, tree species range reduction, and carbon storage in Amazonian forests. *Ecography* **39**, 194–203 (2016).
17. Pedersen, R. Ø., Faurby, S. & Svenning, J. Late-Quaternary megafauna extinctions have strongly reduced mammalian vegetation consumption. *Global Ecology and Biogeography* **32**, 1814–1826 (2023).
18. Malanoski, C., Farnsworth, A., Lunt, D. J., Valdes, P. & Saupe, E. Climate change is an important predictor of extinction risk on macroevolutionary timescales. *Science* **383**, 1130-1134 (2024).
19. Zeller, E. *et al*. Human adaptation to diverse biomes over the past 3 million years. *Science* **380**, 604–608 (2023).
20. Timmermann, A. *et al*. Climate effects on archaic human habitats and species successions. *Nature* **604**, 495–501 (2022).
21. Krapp, M., Beyer, R. M., Edmundson, S. L., Valdes, P. J. & Manica, A. A statistics-based reconstruction of high-resolution global terrestrial climate for the last 800,000 years. *Scientific Data* **8**, 1–18 (2021).
22. Barreto, E., Holden, P. B., Edwards, N. R. & Rangel, T. F. PALEO-PGEM-Series: A spatial time series of the global climate over the last 5 million years (Plio-Pleistocene). *Global Ecology and Biogeography* **7**, 1034-1045 https://doi.org/10.1111/geb.13683 (2023).
23. Sutherland, W. J. *et al*. One Hundred Questions of Importance to the Conservation of Global Biological Diversity. *Conservation Biology* **23**, 557–567 (2009).
24. Darroch, S. A. F., Saupe, E. E., Casey, M. M. & Jorge, M. L. S. P. Integrating geographic ranges across temporal scales. *Trends Ecol Evol* **37**, 851–860 (2022).
25. Silvestro, D. *et al*. A 450 million years long latitudinal gradient in age-dependent extinction. *Ecol Lett* **23**, 439–446 (2020).
26. Raja, N. B. *et al*. Colonial history and global economics distort our understanding of deep-time biodiversity. *Nat Ecol Evol* **6**, 145–154 (2021).
27. Faurby, S. *et al*. PHYLACINE 1.2: The Phylogenetic Atlas of Mammal Macroecology. *Ecology* **99**, 2626 (2018).
28. Carotenuto, F. *et al*. MInOSSE: A new method to reconstruct geographic ranges of fossil species. *Methods Ecol Evol* **11**, 1121–1132 (2020).
29. Svenning, J. C., Fløjgaard, C., Marske, K. A., Nógues-Bravo, D. & Normand, S. Applications of species distribution modeling to paleobiology. *Quat Sci Rev* **30**, 2930–2947 (2011).
30. Mondanaro, A. *et al*. The role of habitat fragmentation in Pleistocene megafauna extinction in Eurasia. *Ecography* **44**, 1619–1630 (2021).
31. Damuth, J. D. & MacFadden, B. J. *Body Size in Mammalian Paleobiology: Estimation and Biological Implications*. (Cambridge University Press, 1990).
32. Hatton, I. A., Dobson, A. P., Storch, D., Galbraith, E. D. & Loreau, M. Linking scaling laws across eukaryotes. *Proceedings of the National Academy of Sciences* **116**, 21616–21622 (2019).
33. Garland, T. Jr., Ives, A. R., Garland, T. & Ives, A. R. Using the Past to Predict the Present: Confidence Intervals for Regression Equations in Phylogenetic Comparative Methods. *American Naturalist* **155**, 346–364 (2000).
34. Diniz-Filho, J. A. F., Loyola, R. D., Raia, P., Mooers, A. O. & Bini, L. M. Darwinian shortfalls in biodiversity conservation. *Trends Ecol Evol* **28**, 689–695 (2013).
35. Redding, D. W., Mooers & A, O. Incorporating evolutionary measures into conservation prioritization. *Conservation Biology* **20**, 1670–1678 (2006).
36. Tucker, C. M. *et al*. A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biol Rev Camb Philos Soc* **92**, 698–715 (2017).
37. Mazel, F. *et al*. Prioritizing phylogenetic diversity captures functional diversity unreliably. *Nat Commun* **9**, 2888 (2018).

38. Hughes, L. J. *et al*. Global hotspots of traded phylogenetic and functional diversity. *Nature* **620**, 351–357 (2023).
39. Castiglione, S., Serio, C., Mondanaro, A., Melchionna, M. & Raia, P. Fast production of large, time-calibrated, informal supertrees with tree.merger. *Palaeontology* **65**, e12588 (2022).
40. Uyeda, J. C., Pennell, M. W., Miller, E. T., Maia, R. & McClain, C. R. The Evolution of Energetic Scaling across the Vertebrate Tree of Life. *Am Nat* **190**, 185–199 (2017).
41. Bininda-Emonds, O. R. P. The evolution of supertrees. *Trends Ecol Evol* **19**, 315–322 (2004).
42. Arnold, C., Matthews, L. J. & Nunn, C. L. The 10kTrees website: a new online resource for primate phylogeny. *Evolutionary Anthropology: Issues, News, and Reviews* **19**, 114–118 (2010).
43. Moura, M. R. *et al*. A phylogeny-informed characterisation of global tetrapod traits addresses data gaps and biases. *PLoS Biol* **22**, e3002658 (2024).
44. Jones, K. E. *et al*. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology* **90**, 2648–2648 (2009).
45. Santini, L., Isaac, N. J. B. & Ficetola, G. F. TetraDENSITY: A database of population density estimates in terrestrial vertebrates. *Global Ecology and Biogeography* **27**, 787–791 (2018).
46. Kissling, W. D. *et al*. Establishing macroecological trait datasets: digitalization, extrapolation, and validation of diet preferences in terrestrial mammals worldwide. *Ecol Evol* **4**, 2913–2930 (2014).
47. Wilman, H. *et al*. EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology* **95**, 2027–2027 (2014).
48. Lundgren, E. J. *et al*. Functional traits of the world's late Quaternary large-bodied avian and mammalian herbivores. *Sci Data* **8**, 17 (2021).
49. Middleton, O., Svensson, H., Scharlemann, J. P. W., Faurby, S. & Sandom, C. CarniDIET 1.0: A database of terrestrial carnivorous mammal diets. *Global Ecology and Biogeography* **30**, 1175–1182 (2021).
50. Soria, C. D., Pacifici, M., Di Marco, M., Stephen, S. M. & Rondinini, C. COMBINE: a coalesced mammal database of intrinsic and extrinsic traits. *Ecology* **102**, e03344 (2021).
51. Harrison, S. P. & Prentice, C. I. Climate and CO_2 controls on global vegetation distribution at the last glacial maximum: analysis based on palaeovegetation data, biome modelling and palaeoclimate simulations. *Glob Chang Biol* **9**, 983–1004 (2003).
52. Chang, W. *et al*. shiny: Web Application Framework for R. https://shiny.posit.co/ (2023).
53. Melchionna, M. *et al*. Fragmentation of Neanderthals' pre-extinction distribution by climate change. *Palaeogeogr Palaeoclimatol Palaeoecol* **496**, 146–154 (2018).
54. Belfiore, A. M. *et al*. Too much of a good thing? Supplementing current species observations with fossil data to assess climate change vulnerability via ecological niche models. *Biol Conserv* **291**, 110495 (2024).
55. Carotenuto, F. *et al*. The influence of climate on species distribution over time and space during the late Quaternary. *Quat Sci Rev* **149**, 188–199 (2016).
56. Williams, J. W. *et al*. The Neotoma Paleoecology Database, a multiproxy, international, community-curated data resource. *Quat Res* **89**, 156–177 (2018).
57. Bird, M. I. *et al*. Radiocarbon Dating of "Old" Charcoal Using a Wet Oxidation, Stepped-Combustion Procedure. *Radiocarbon* **41**, 127–140 (1999).
58. Higham, T. European Middle and Upper Palaeolithic radiocarbon dates are often older than they look: problems with previous dates and some remedies. *Antiquity* **85**, 235–249 (2011).
59. Higham, T. *et al*. Problems with radiocarbon dating the Middle to Upper Palaeolithic transition in Italy. *Quat Sci Rev* **28**, 1257–1267 (2009).
60. Tagliacozzo, A., Romandini, M., Fiore, I., Gala, M. & Peresani, M. Animal Exploitation Strategies during the Uluzzian at Grotta di Fumane (Verona, Italy). in *Vertebrate Paleobiology and Paleoanthropology* 129–150 https://doi.org/10.1007/978-94-007-6766-98 (Springer, 2013).
61. Stafford, T. W., Brendel, K. & Duhamel, R. C. Radiocarbon, 13C and 15N analysis of fossil bone: Removal of humates with XAD-2 resin. *Geochim Cosmochim Acta* **52**, 2257–2267 (1988).
62. Abelson, P. H. & Hoering, T. C. Carbon isotope fractionation in formation of amino acids by photosynthetic organisms. *Proc Natl Acad Sci USA* **47**, 623–32 (1961).
63. Marom, A., McCullagh, J. S. O., Higham, T. F. G. & Hedges, R. E. M. Hydroxyproline Dating: Experiments on the ^{14}C Analysis of Contaminated and Low-Collagen Bones. *Radiocarbon* **55**, 698–708 (2013).
64. Waters, M. R., Stafford, T. W. & Carlson, D. L. The age of Clovis—13,050 to 12,750 cal yr B.P. *Sci Adv* **6**, eaaz0455 (2020).
65. Devièse, T. *et al*. Increasing accuracy for the radiocarbon dating of sites occupied by the first Americans. *Quat Sci Rev* **198**, 171–180 (2018).
66. Zilhão, J. *et al*. Last Interglacial Iberian Neandertals as fisher-hunter-gatherers. *Science* **367**, eaaz7943 (2020).
67. Arsuaga, J. L. *et al*. Evidence of paleoecological changes and Mousterian occupations at the Galería de las Estatuas site, Sierra de Atapuerca, northern Iberian plateau, Spain. *Quat Res* **88**, 345–367 (2017).
68. Zhang, J.-F. *et al*. Radiocarbon and luminescence dating of the Wulanmulun site in Ordos, and its implication for the chronology of Paleolithic sites in China. *Quat Geochronol* **72**, 101371 (2022).
69. Reimer, P. J. *et al*. The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP). *Radiocarbon* **62**, 725–757 (2020).
70. Haslett, J., & Parnell, A.C. A simple monotone process with application to radiocarbon-dated depth chronologies. Journal of the Royal Statistical Society: Series C (Applied Statistics) **57**, 399-418 (2008).
71. Prentice, I. C. *et al*. A global biome model based on plant physiology and dominance, soil properties and climate. *J Biogeogr* **19**, 117–134 (1992).
72. Willeit, M., Ganopolski, A., Calov, R. & Brovkin, V. Mid-Pleistocene transition in glacial cycles explained by declining CO_2 and regolith removal. *Sci Adv* **5**, eaav7337 (2019).
73. Lüthi, D. *et al*. High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature* **453**, 379–382 (2008).
74. Berger, A. L. Long-Term Variations of Caloric Insolation Resulting from the Earth's Orbital Elements. *Quat Res* **9**, 139–167 (1978).
75. Friedrich, T. & Timmermann, A. Using Late Pleistocene sea surface temperature reconstructions to constrain future greenhouse warming. *Earth Planet Sci Lett* **530**, 115911 (2020).
76. Friedrich, T., Timmermann, A., Tigchelaar, M., Elison Timm, O. & Ganopolski, A. Nonlinear climate sensitivity and its implications for future greenhouse warming. *Sci Adv* **2**, e1501923 (2016).
77. Meehl, G. A. *et al*. Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. *Sci Adv* **6**, eaba1981 (2020).
78. Timmermann, A. & Friedrich, T. Late Pleistocene climate drivers of early human migration. *Nature* **538**, 92–95 (2016).
79. Timmermann, A. *et al*. Modeling Obliquity and CO_2 Effects on Southern Hemisphere Climate during the Past 408 ka*. *J Clim* **27**, 1863–1875 (2014).
80. Kaplan, J. O. *et al*. Climate change and Arctic ecosystems: 2. Modeling, paleodata-model comparisons, and future projections. *Journal of Geophysical Research: Atmospheres* **108**, 8171 (2003).

81. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**, 4302–4315 (2017).
82. Guisan, A. & Zimmermann, N. E. Predictive habitat distribution models in ecology. *Ecol Modell* **135**, 147–186 (2000).
83. Hijmans, R. J., Phillips, S., Leathwick, J. & Elith, J. dismo: Species Distribution Modeling. R package version 1.3-14, https://CRAN.R-project.org/package=dismo (2023).
84. Hofreiter, M. *et al*. Evidence for Reproductive Isolation between Cave Bear Populations. *Current Biology* **14**, 40–43 (2004).
85. Knapp, M. From a molecules' perspective – contributions of ancient DNA research to understanding cave bear biology. *Hist Biol* **31**, 442–447 (2019).
86. Stiller, M., Knapp, M., Stenzel, U., Hofreiter, M. & Meyer, M. Direct multiplex sequencing (DMPS)—a novel method for targeted high-throughput sequencing of ancient and highly degraded DNA. *Genome Res* **19**, 1843–1848 (2009).
87. Barlow, A. *et al*. Middle Pleistocene genome calibrates a revised evolutionary history of extinct cave bears. *Current Biology* **31**, 1771–1779 (2021).
88. Baryshnikov, G. F. & Puzachenko, A. Y. Morphometry of lower cheek teeth of cave bears (Carnivora, Ursidae) and general remarks on the dentition variability. *Boreas* **49**, 562–593 (2020).
89. Agenbroad, L. D. North American Proboscideans: Mammoths: The state of knowledge, 2003. *Quaternary International* **126–128**, 73–92 (2005).
90. Saunders, J. J. *et al*. Paradigms and proboscideans in the southern Great Lakes region, USA. *Quaternary International* **217**, 175–187 (2010).
91. Graham, R. W. Taxonomy of North American mammoths. *The Colby Mammoth Site: taphonomy and archaeology of a Clovis kill site in northern Wyoming* (Frison, G. C. and Todd, L. C. *editors)*. University of New Mexico Press, Albuquerque 165–229 (1986).
92. McDaniel, G. E. & Jefferson, G. T. Mammoths in our midst: The proboscideans of Anza-Borrego Desert State Park®, Southern California, USA. *Quaternary International* **142–143**, 124–129 (2006).
93. Enk, J. *et al*. Mammuthus Population Dynamics in Late Pleistocene North America: Divergence, Phylogeography, and Introgression. *Front Ecol Evol* **4**, 42 (2016).
94. Lister, A. M. On the type material and evolution of North American mammoths. *Quaternary International* **443**, 14–31 (2017).
95. Pickford, M. Orycteropus (Tubulidentata, Mammalia) from Langebaanweg and Baard's Quarry, Early Pliocene of South Africa. *C R Palevol* **4**, 715–726 (2005).
96. Lehmann, T., Vignaud, P., Likius, A. & Brunet, M. A new species of Orycteropodidae (Mammalia, Tubulidentata) in the Mio-Pliocene of northern Chad. *Zool J Linn Soc* **143**, 109–131 (2005).
97. LEHMANN, T. Phylogeny and systematics of the Orycteropodidae (Mammalia, Tubulidentata). *Zool J Linn Soc* **155**, 649–702 (2009).
98. Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A. & Wilson, A. C. DNA sequences from the quagga, an extinct member of the horse family. *Nature* **312**, 282–284 (1984).
99. Orlando, L. *et al*. Revising the recent evolutionary history of equids using ancient DNA. *Proceedings of the National Academy of Sciences* **106**, 21754–21759 (2009).
100. Prins, H. H. T. & Gordon, I. J. Are Equids Evolutionary Dead Ends? in *The Equids: a Suite of Splendid Species.* 1–25 https://doi.org/10.1007/978-3-031-27144-11 (Springer, Cham, 2023).
101. Grubb, P. Order perissodactyla. *Mammal species of the world: a taxonomic and geographic reference* (Wilson, D. E. and Reeder, D. M. *eds.). 3rd ed*. Johns Hopkins University Press, Baltimore, Maryland 629–636 (2005).
102. Dembitzer, J., Castiglione, S., Raia, P. & Meiri, S. Small brains predisposed Late Quaternary mammals to extinction. *Sci Rep* **12**, 3453 (2022).
103. Raia, P., Carotenuto, F. & Meiri, S. One size does not fit all: No evidence for an optimal body size on islands. *Global Ecology and Biogeography* **19**, 475–484 (2010).
104. Raia, P. *et al*. Rapid action in the Palaeogene, the relationship between phenotypic and taxonomic diversification in Coenozoic mammals. *Proceedings of the Royal Society B: Biological Sciences* **280**, 20122244 (2013).
105. Álvarez-Carretero, S. *et al*. A Species-Level Timeline of Mammal Evolution Integrating Phylogenomic Data. *Nature* **602**, 263-267 https://doi.org/10.1038/s41586-021-04341-1 (2021).
106. Larramendi, A. Estimating tusk masses in proboscideans: a comprehensive analysis and predictive model. *Hist Biol* 1–14 https://doi.org/10.1080/08912963.2023.2286272 (2023).
107. Dantas, M. A. T. Estimating the body mass of the Late Pleistocene megafauna from the South America Intertropical Region and a new regression to estimate the body mass of extinct xenarthrans. *J South Am Earth Sci* **119**, 103900 (2022).
108. Lintulaakso, K., Tatti, N. & Žliobaitė, I. Quantifying mammalian diets. *Mammalian Biology* **103**, 53–67 (2023).
109. Calder, W. A. *Size, Function, and Life History*. (Courier Corporation, 1996).
110. Brown, J. H. & West, G. B. *Scaling in Biology*. (Oxford University Press, USA, 2000).
111. Broekman, M. J. E. *et al*. *HomeRange*: A global database of mammalian home ranges. *Global Ecology and Biogeography* **32**, 198–205 (2023).
112. De Cuyper, A. *et al*. Predator size and prey size–gut capacity ratios determine kill frequency and carcass production in terrestrial carnivorous mammals. *Oikos* **128**, 13–22 (2019).
113. Goolsby, E. W., Bruggeman, J. & Ané, C. Rphylopars: fast multivariate phylogenetic comparative methods for missing data and within-species variation. *Methods Ecol Evol* **8**, 22–27 (2017).
114. Ives, A., Dinnage, R., Nell, L. A., Helmus, M. & Li, D. phyr: Model Based Phylogenetic Analysis. https://CRAN.R-project.org/package=phyr (2020).
115. Diniz-Filho, J. A. F., de Sant'Ana, C. E. R. & Bini, L. M. An Eigenvector method for estimating phylogenetic inertia. *Evolution* **52**, 1247–1262 (1998).
116. Mondanaro, A., Raia, P., Melchionna, M. & Castiglione, S. EutherianCoP. An integrated biotic and climate database for conservation paleobiology based on eutherian mammals. *Zenodo.* https://doi.org/10.5281/zenodo.14009105 (2024).
117. Lenth, R. V. emmeans: Estimated Marginal Means, aka Least-Squares Means. https://github.com/rvlenth/emmeans (2023).

Acknowledgements

PR was supported by project OMEGA (P2022RZ4PL, CUP:E53D23014440001), financed by the European Union NextGenerationEU (Mission 4, Component 1) fund. AT, EZ and TV were supported by the Institute for Basic Science grant no. IBS-R028-D1.

Author contributions

A.M. conceived the study. A.M. and P.R. lead the writing of the draft. A.M. cured the database with significant input by Md.F. and G.G. S.C., G.G., A.E. and C.S. assembled the phylogenetic tree and provided missing data estimation. M.M. produced the EutherianCoP application with significant support by G.G., C.S., S.C. A.T., E.Z. and T.V. provided the climate data and helped with the integration of the data within the database. Each author contributed significantly to writing the manuscript, preparing the figures and files for submission.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-024-04181-4.

Correspondence and requests for materials should be addressed to A.M. or P.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.