TITLE

The emergence of modern zoogeographic regions in Asia examined through climate–dental trait association patterns

AUTHOR LIST

Liping Liu[1,2*], Esther Galbrun[3*], Hui Tang[1,4,5], Anu Kaakinen[1], Zhongshi Zhang[6], Zijian Zhang[7], Indrė Žliobaitė[8,1]

AFFILIATIONS

[1]Department of Geosciences and Geography, University of Helsinki, P.O. Box 64, University of Helsinki, FI-00014, Finland.
[2]Department of Palaeobiology, the Swedish Museum of Natural History, P.O. Box 50007, Stockholm, SE-104 05, Sweden.
[3]School of Computing, University of Eastern Finland, Technopolis, Microkatu 1, Kuopio, FI-70210, Finland.
[4]Climate System Research Unit, Finnish Meteorological Institute, P.O. Box 503, Helsinki, FI-00101, Finland.
[5]Department of Geosciences, University of Oslo, P.O. Box 1022, Oslo NO-0315, Norway.
[6]Department of Atmospheric Science, School of Environmental Studies, China University of Geosciences, 388 Lumo Road, Wuhan, 430074, China.
[7]Key Laboratory of Cenozoic Geology and Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, 19, Beitucheng Western Road, Chaoyang District, Beijing, 100029, China.
[8]Department of Computer Science, University of Helsinki, P.O. Box 64, University of Helsinki, FI-00014, Finland.
Corresponding authors: Liping Liu (liping.liu@helsinki.fi) and Esther Galbrun (esther.galbrun@uef.fi)

ABSTRACT

The complex and contrasted distribution of terrestrial biota in Asia has been linked to active tectonics and dramatic climatic changes during the Neogene. However, the timings of the emergence of these distributional patterns and the underlying climatic and tectonic mechanisms remain disputed. Here, we apply a computational data analysis technique, called redescription mining, to track these spatiotemporal phenomena by studying the associations between the prevailing herbivore dental traits of mammalian communities and climatic conditions during the Neogene. Our results indicate that the modern latitudinal zoogeographic division emerged after the Middle Miocene climatic transition, and that the modern monsoonal zoogeographic pattern emerged during the late Late Miocene. Furthermore, the presence of a montane forest biodiversity hotspot in the Hengduan Mountains alongside Alpine fauna on the Tibetan Plateau suggests that the modern distribution patterns may have already existed since the Pliocene.

INTRODUCTION

The Asian continent is characterized by a complex and contrasted distribution of climate and biota. Not only is there a stark north–south zoogeographic division between the Indomalayan and Palearctic realms [1, 2], but also an east–west zoogeographic distributional pattern controlled by the East Asia Monsoon [3–5].

Moreover, due to the presence of the Tibetan Plateau, the Asian continent hosts montane biodiversity hotspots at low latitudes [6], of which there is no equivalent elsewhere on the planet. The diverse biotas make Asia one of the most interesting areas for studying how the modern biogeographical regions emerged following tectonic and climatic changes. The emergence of the modern biogeographical regions have been linked to the topographic rise of the Tibetan Plateau [7], monsoon circulation [8–11] and transformations of the global climate [12]. The timing of these changes and their controlling mechanisms are still debated and have been key questions in the field of paleontology. The difficulty is not only the sparsity and incompleteness of fossil records, in terms of both space and time, but also the lack of methods allowing to delineate geographical regions in the past, where the distribution of fossil communities

has no comparable extant relatives. Alternatively, there have been several studies attempting to track modern biotas back through time along phylogenies to shed light on their evolution, but they either provide only indirect phylogenetic evidence [13–15] or are incompatible with fossil evidence [5].

In this work we analyze the dental traits of large herbivore communities, in order to track changes in mammal communities and their climatic contexts in the deep past. Previous works showed that the distribution of dental traits across mammalian communities correlate well with climatic conditions, in the present [16] and in the past [8, 10], and can be used to build global and regional predictive models for precipitation and temperature [17, 18], as well as for productivity [19] or vegetation cover [20], based on such traits. A computational data analysis methodology called redescription mining was recently tailored to biogeographic studies [21, 22], aiming to identify local patterns of association between the dental traits of mammalian communities, on one hand, and the climatic conditions, on the other hand. Here, we employ this methodology to generate and explore hypotheses about the emergence of the modern zoogeographic regions. First, we extract redescriptions, i.e. local patterns of association, from high resolution data about mammalian communities and climatic conditions across Asia in the present day. Then, we evaluate the redescriptions on data of fossil mammalian communities and modeled paleoclimatic conditions during five intervals of the Neogene, that is, since 22 million years ago. Through this lens, we examine the build-up of modern zoogeographic regions in Asia.

RESULTS

Redescriptions are pairs of logical statements about the values of the data variables. In our context, they capture the interplay between dental traits of mammalian communities and climatic conditions. More specifically, we consider seven dental traits and nineteen bioclimatic variables, listed in Table 1. These dental traits, introduced and studied in previous work [17, 18, 23], are understood to have good correlations with the encountered environmental conditions, including the climate.

A logical statement, also known as a query, specifies a range of values the involved variables might take, typically by means of thresholds. Given data recording the values of these variables at the localities within our study area, each query implicitly selects a subset of localities, those localities where the values satisfy the requirements stated in the query. For example, a query over the bioclimatic variables might require the mean annual temperature (TMeanY) to be lower than 15.4°C, selecting the localities with colder climate. On the other hand, a query over the dental traits might require the fraction of species with structural fortification of cusps (SF) to be lower than 22.2% and the fraction of bunodont species (BU) to be lower than 35.7%, selecting the localities where the prevalence of structural fortification of cusps and the prevalence of bunodonty are both low.

A redescription then consists of a pair of queries, here respectively over dental traits and bioclimatic variables, that select similar subsets of localities, thereby capturing a pattern of association between the involved variables and value ranges. The similarity of the two subsets of localities is measured using the Jaccard coefficient, denoted as J, and generally referred to as the accuracy of the redescription, while their intersection is referred to as the support of the redescription, whose size as a percentage of the total number of localities studied is denoted as $\mathrm{supp}\,\%$. Note that redescriptions are not predictive models, the accuracy is a similarity measure calculated on the considered dataset and is not related to a prediction task. Intuitively, the closer the accuracy of a redescription is to one, the stronger is the association between the corresponding conditions in the dataset. Redescription mining algorithms build and evaluate many pairs of queries, looking for those that have the highest Jaccard coefficients, while fulfilling user-defined constraints (e.g. on the complexity of the queries and the number of satisfying localities). As a major difference, redescription mining is a descriptive approach capturing patterns of associations between subsets of variables that hold locally, whereas machine learning approches such as regression produce a global predictive model, assuming that the same relation between the variables holds throughout the dataset.

Characterizing modern zoogeographic regions

Having applied a redescription mining algorithm on our present-day dataset, we manually select among the obtained results nine redescriptions that have a high accuracy on that dataset and together provide a good coverage of the study area as well as of the different dental traits variables.

The selected redescriptions, denoted as rA–rI, are shown in Figure 1. For each redescription, we list the query over dental traits variables ($q_{\mathbf{D}}$), the query over bioclimatic variables ($q_{\mathbf{C}}$), the accuracy (J) as well as the size of its support as a percentage of the total number of present-day localities ($\mathrm{supp}\,\%$). The status of each redescription across the study area is visualized as a map, with a dot for each present-day locality, whose color indicate whether the queries of the considered redescription are satisfied at that locality. We use the same color code throughout to represent the four possible cases: purple where both the dental traits query and the climate query are satisfied; red where the dental traits query is satisfied but the climate query is not; blue where the climate query is satisfied but the dental traits query is not; gray where neither the dental traits query nor the climate query are satisfied. The support of each redescription, defined as the set of localities where both queries are satisfied, is hence drawn in purple. Note that the queries used as example above are from an actual obtained redescription, namely rA.

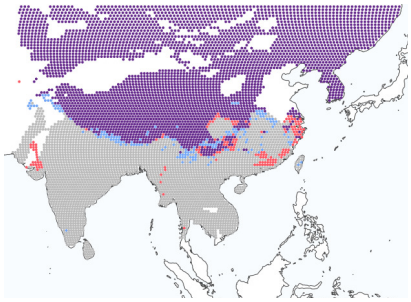
We see that the redescriptions delineate areas corresponding to prominent ecoregions and notable mammalian distribution patterns in present-day Asia [2].

Among the nine redescriptions, five (rA, rB, rD, rF and rI) outline a north–south zoogeographic pattern across the study area. Redescriptions rA, rD and rF select regions in the northern half of the study area, requiring low temperatures, whether on average throughout the year (rA: $[\mathrm{TMeanY} \leq 15.4]$ and rD: $[-5.5 \leq \mathrm{TMeanY} \leq 19.1]$) or at the lowest of the coldest month (rF: $[-29.6 \leq \mathrm{TMinColdM} \leq 6.0]$). On the other hand, redescriptions rB and rI select regions in the southern half of the study area, requiring mild to warm temperatures. Specifically, the average temperature should not drop below moderate during the driest quarter (rB: $[2.0 \leq \mathrm{TMeanDryQ}]$ and rI: $[-4.0 \leq \mathrm{TMeanDryQ}]$). Overall, the cold climate of the northern half is associated with a low prevalence of bunodonty among the mammalian communities (rA: $[\mathrm{BU} \leq 0.357]$) and structural fortification of cusps (rA: $[\mathrm{SF} \leq 0.222]$) as well as a high prevalence of acute lophs (rD: $[0.154 \leq \mathrm{AL}]$) and exclusively obtuse lophs (rF: $[0.235 \leq \mathrm{OO}]$) among the mammalian communities. Vice versa, the warm climate of the southern half is associated with a high prevalence of bunodonty (rB: $[0.286 \leq \mathrm{BU}]$) and moderate prevalence of obtuse lophs (rI: $[0.222 \leq \mathrm{OL} \leq 0.571]$) among the mammalian communities.

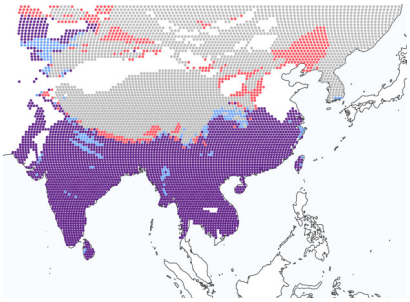
Three redescriptions (rC, rG and rH) outline a southeast–northwest zoogeographic pattern across the study area. Redescription rC selects regions in the east and south of the study area that receive high precipitation amounts during the wettest month ($[117.0 \leq \mathrm{PWetM}]$), which correspond to the regions under the influence of the Asian monsoon. Redescriptions rG and rH select more specifically East and Southeast Asian Monsoon regions, i.e. the same regions as rC but excluding inland India, requiring respectively limited diurnal temperature variations ($[\mathrm{TMeanRngD} \leq 12.3]$) and a mild warm season ($[17.0 \leq \mathrm{TMeanWarmQ} \leq 30.6]$). In terms of dental traits, redescriptions rG and rH require a low prevalence of hypsodonty (HYP) among the mammalian communities, while rC requires a low prevalence of hypsodonty or a non-negligible prevalence of structural fortification of cusps (SF) (or both).

The remaining redescription (rE) does not fall squarely in either group, as it covers the Indian subcontinent and Southeast Asia, but also extends northward over the Tibetan plateau, requiring high isothermality, on one hand, and a low prevalence of acute lophs, on the other hand.

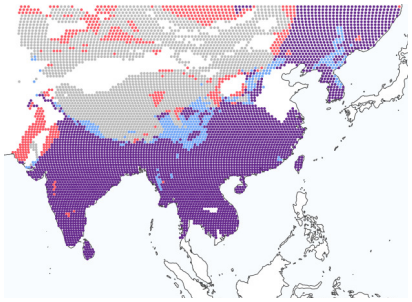
The climate queries of redescriptions rA and rC involve a couple of dental traits each, using respectively a conjunction and a disjunction. To properly understand such redescriptions, it can be useful to consider separately the subqueries involving each variable. Such further examination (see maps visualization in [24]) reveals that the conditions on SF apply almost exclusively to India. Indeed, prevalences of structural fortification of cusps above 15.4% or even 22.2% are found throughout the subcontinent and hardly anywhere else. Furthermore, lower prevalences of bunodonty ($\mathrm{BU} \leq 0.357$) and higher prevalences of hypsodonty ($1.733 < \mathrm{HYP}$) are commonly encountered today in inland Indian, unlike in East China and Southeast Asia.



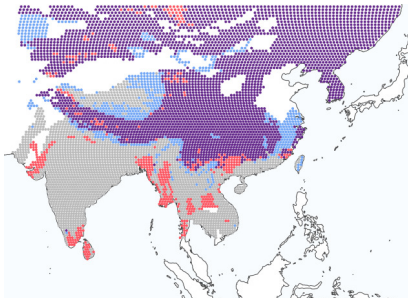
rA $\mathrm{J} = 0.914 \quad \mathrm{supp}\,\% = 54.61$
$q_{\mathbf{D}} = [\mathrm{SF} \leq 0.222]$ AND $[\mathrm{BU} \leq 0.357]$
$q_{\mathbf{C}} = [\mathrm{TMeanY} \leq 15.4]$



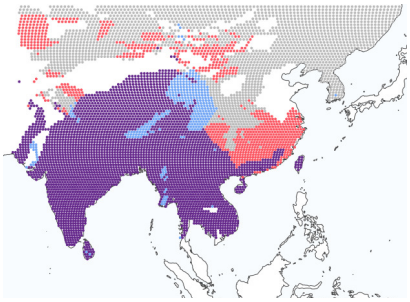
rB $\mathrm{J} = 0.810 \quad \mathrm{supp}\,\% = 47.31$
$q_{\mathbf{D}} = [0.286 \leq \mathrm{BU}]$
$q_{\mathbf{C}} = [2.0 \leq \mathrm{TMeanDryQ}]$



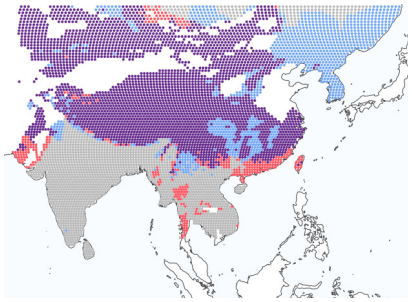
rC $\mathrm{J} = 0.787 \quad \mathrm{supp}\,\% = 54.90$
$q_{\mathbf{D}} = [\mathrm{HYP} \leq 1.733]$ OR $[0.154 \leq \mathrm{SF}]$
$q_{\mathbf{C}} = [117.0 \leq \mathrm{PWetM}]$



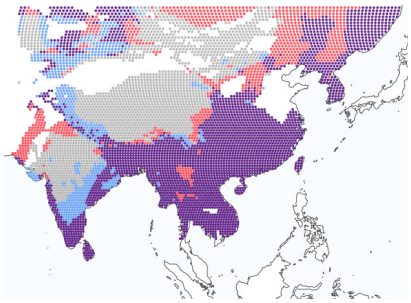
rD $\mathrm{J} = 0.747 \quad \mathrm{supp}\,\% = 51.33$
$q_{\mathbf{D}} = [0.154 \leq \mathrm{AL}]$
$q_{\mathbf{C}} = [-5.5 \leq \mathrm{TMeanY} \leq 19.1]$



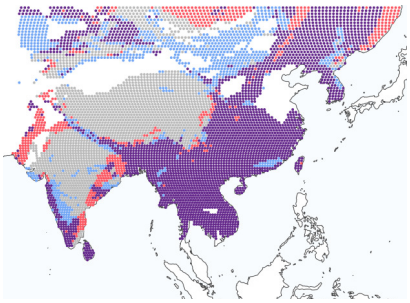
rE $\mathrm{J} = 0.729 \quad \mathrm{supp}\,\% = 49.94$
$q_{\mathbf{D}} = [\mathrm{AL} \leq 0.308]$
$q_{\mathbf{C}} = [32.0 \leq \mathrm{TIso}]$



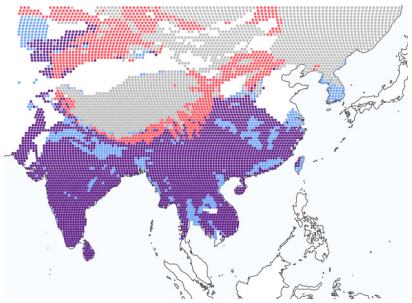
rF $\mathrm{J} = 0.676 \quad \mathrm{supp}\,\% = 47.19$
$q_{\mathbf{D}} = [0.235 \leq \mathrm{OO}]$
$q_{\mathbf{C}} = [-29.6 \leq \mathrm{TMinColdM} \leq 6.0]$



rG $\mathrm{J} = 0.658 \quad \mathrm{supp}\,\% = 45.91$
$q_{\mathbf{D}} = [\mathrm{HYP} \leq 1.857]$
$q_{\mathbf{C}} = [\mathrm{TMeanRngD} \leq 12.3]$



rH $\mathrm{J} = 0.657 \quad \mathrm{supp}\,\% = 45.91$
$q_{\mathbf{D}} = [\mathrm{HYP} \leq 1.778]$
$q_{\mathbf{C}} = [17.0 \leq \mathrm{TMeanWarmQ} \leq 30.6]$



rI $\mathrm{J} = 0.643 \quad \mathrm{supp}\,\% = 45.03$
$q_{\mathbf{D}} = [0.222 \leq \mathrm{OL} \leq 0.571]$
$q_{\mathbf{C}} = [-4.0 \leq \mathrm{TMeanDryQ}]$

Figure 1 Redescriptions rA–rI in the present-day dataset. Localities that satisfy both queries, only the dental traits query, only the climate query and neither queries, are drawn in purple, red, blue and gray, respectively. For each redescription, we list the query over dental traits variables ($q_{\mathbf{D}}$), the query over bioclimatic variables ($q_{\mathbf{C}}$), the accuracy (J) as well as the size of its support as a percentage of the total number of localities ($\mathrm{supp}\,\%$). See maps visualization in [24].

Next, we take some of these redescriptions from the present-day and evaluate them on our dataset from the past, which combines data from the fossil record with paleoclimate model simulations, aiming to track the emergence of these modern zoogeographical regions during the Neogene.

After evaluating them on the past dataset, we can visualize the status of each redescription for fossil localities from a given time interval as a map, using the same color code. For reference, we show the status of the redescription on the present-day in the background. We show the maps for redescriptions rB and rC in Figure 2. Further maps are provided in [24].

Emergence of the modern north–south zoogeographic division

Redescriptions rA and rB best capture the north–south zoogeographic division of the present day. In fact, the two redescriptions are almost logical complements of one another and tell essentially the same story, despite some differences. Thus, we focus on rB that lends itself to easier interpretation, thanks in particular to its simpler dental traits query.

In Figure 2a, we see that fossil localities at low and middle latitudes are drawn in purple, meaning that the conditions at these localities satisfy both the dental traits and climate queries of redescription rB. In other words, fossil localities from the Early Miocene (23.03–15.97 Ma) in that area match the conditions of modern tropical and subtropical areas as captured by rB. On the other hand, fossil localities at high latitudes are drawn in gray, meaning that cold winters and low fractions of bunodont species are encountered there. This indicates that a north–south division already existed then. Blue and purple dots represent localities that satisfy the climatic conditions of the redescription. Such localities appearing northward of the present-day support of rB (i.e. falling on a gray background to the north of the purple background) suggests that warm hospitable conditions in the Early Miocene extended further north than they do today.

The north–south pattern is also visible in the Middle Miocene (15.97–11.63 Ma) (Figure 2b). We note the presence of several red dots and a single blue dot among the gray dots at high latitudes. This means that the fossil communities have a fraction of bunodont species above the threshold specified in the dental query, that the redescription associates to warmer conditions, while the climate model predicts relatively cold temperatures, below the threshold specified in the climate query, for most localities at high latitudes. From the faunal perspective, the bunodont mammals seem to have achieved maximum geographic extent during the Middle Miocene.

In the early Late Miocene (11.63–7.246 Ma) (Figure 2c), the northern boundary of rB has retreated southward, with the fossil localities supporting rB (purple dots) falling exclusively within its support in the present-day data (purple background). Localities outside this area are mostly gray, with a few exceptions in blue, i.e. low prevalence of bunodonty but warmer temperatures. The northern boundary of rB in the early Late Miocene appears to have been similar to the present-day situation, suggesting that a north–south zoogeographic pattern comparable to the modern may have emerged before the Late Miocene.

The northern boundary of rB appears to be stable in the late Late Miocene (7.246–5.333 Ma) (Figure 2d), and is also fairly clear in the Pliocene (5.333–2.58 Ma) (Figure 2e), at least in terms of climate. The four blue dots in Southeast Asia correspond to high-altitude sites of the Hengduan Mountains in Yunnan (China) and Gwebin in Burma that appear to host mammalian communities with a low prevalence of bunodonty, unlike the mammalian communities found in the same or neighboring sites during the Miocene.

Emergence of the modern southeast–northwest zoogeographic pattern

Redescriptions rC, rG and rH capture the southeast–northwest zoogeographic distributional pattern of the present day. In fact, species with structural fortification of cusps never exceeds 12.5% of fossil communities in our dataset, and are entirely absent in most cases. Therefore, the condition on SF in the dental query of rC is not satisfied at any of the fossil localities, so that, when evaluated on the past dataset, redescription rC behaves as if the dental query consisted only of the condition on HYP. This means that wherever the dental query is satisfied (red or purple dots in Figure 2f–j) the condition on low prevalence of hypsodonty holds true. Thus, the dental queries of the three

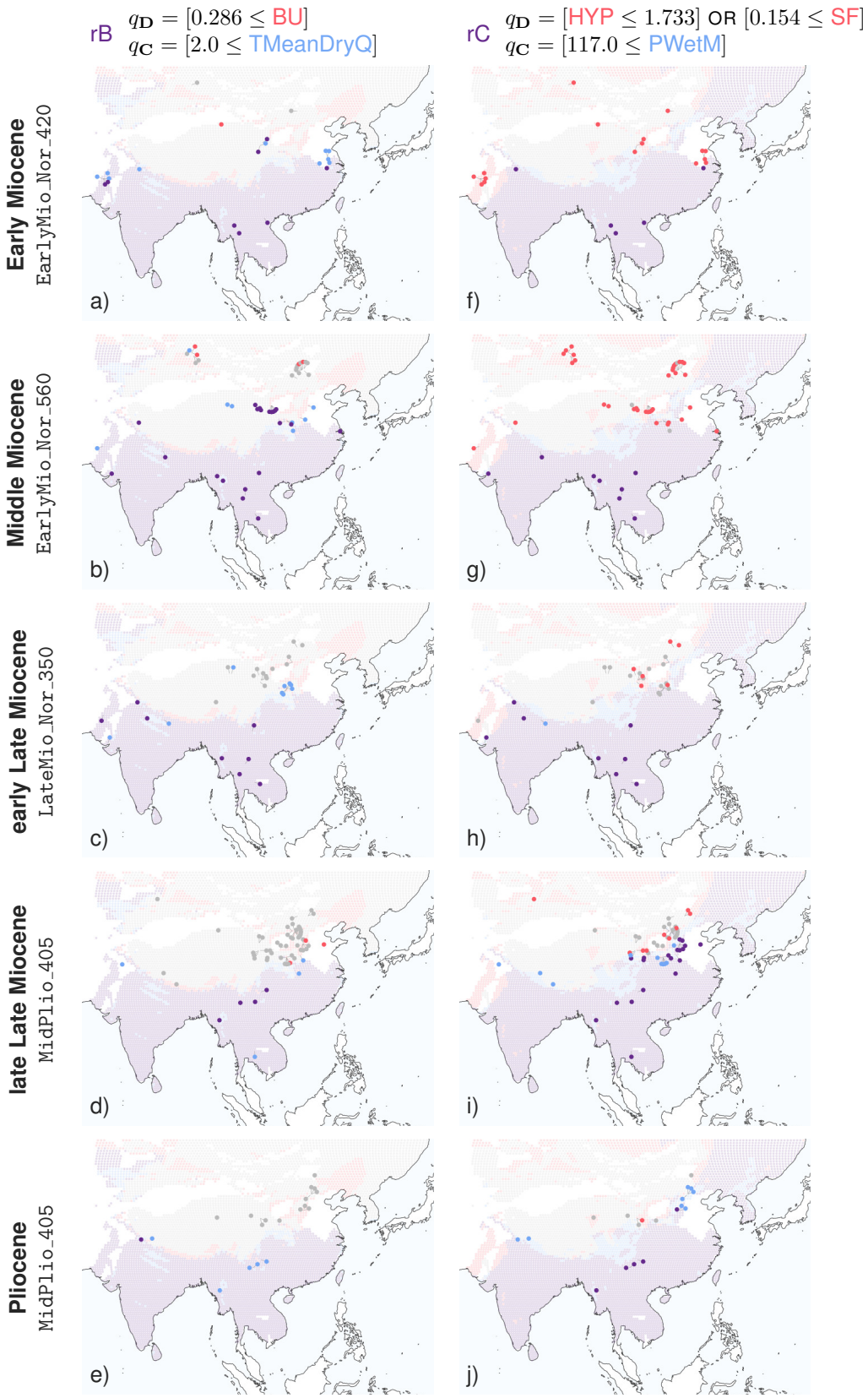


Figure 2 Focus maps of redescriptions rB and rC (columns) evaluated on fossil localities from the different time intervals, considering the corresponding paleoclimate model simulation (rows). Fossil localities that support both queries, only the dental traits query, only the climate query and neither queries, are drawn in purple, red, blue and gray, respectively. Present-day localities are drawn in the background, for reference. See maps visualization in [24].

redescriptions are very similar, all requiring low prevalences of hypsodonty, with somewhat different thresholds. We focus on redescription rC, that involves the lowest threshold, paired with a condition on precipitation (PWetM).

In Figure 2f and g, we see that all but two dots are either red or purple, meaning that during the Early and Middle Miocene, nearly all fossil communities in our dataset exhibit low prevalences of hypsodonty. This is in contrast to the modelled climate conditions indicating that fossil localities receiving high amounts of precipitation during the wettest month (dots that are either blue or purple) are resticted mainly to the south of the study area

In the early Late Miocene (Figure 2h), higher prevalences of hypsodont and mesodont taxa are found among fossil communities at middle latitudes, with overall more gray dots in northern China, while the boundary of the climate query appears to be fairly stable, spanning east–west across the study area.

In the late Late Miocene (Figure 2i), localities satisfying the climate query are found further in northeastern China, with the boundary tilting counter-clockwise. The prevalence of hypsodonty also appears more contrasted across this boundary. In particular, a majority of localities have higher (resp. lower) prevalences of hypsodonty in the northwestern (resp. southeastern) part of East China, with more gray dots than red (resp. more purple dots than blue). This suggests that the distribution of precipitation and of hypsodonty changed from a primarily north–south pattern to a southeast–northwest pattern, similar to the monsoonal pattern found in the present-day data. Fossil data from the Pliocene is rather limited, but the pattern appears to persist (Figure 2j).

The dynamics of zoogeography in the context of climate and tectonics

To provide context for our results and shed more light onto the processes that might have impacted the Asian faunal distribution during the Neogene, we analyze existing quantitative proxy data of temperature, precipitation and orographic height of the Tibetan Plateau, along with the dynamics of two main dental traits and two main bioclimatic variables within subregions of the study area, during the Neogene (Figure 3).

Figure 3a shows the global temperature trend (based on [25]) along with the average and standard deviation of BU and TMeanY calculated over localities from northern and southern Asia (delimited as specified in Supplementary Table 3) for each of the five time intervals. We see that the dynamics of bunodonty in Asia during the Neogene follow global temperature changes. The prevalence of bunodont species remains consistently higher in southern Asia than in northern Asia throughout the analysis period.

Figure 3b shows the modeled mean annual precipitation for East Asia (based on [26]) along with the average and standard deviation of HYP and PTotY calculated over localities from northwestern and southeastern China (delimited as specified in Supplementary Table 3) for each of the five time intervals.

We see that average hypsodonty and precipitation are clearly negatively correlated in southeastern China, in agreement with the expectation that higher hypsodonty is associated with more arid climates. On the other hand, in northwestern China the variables follow unsynchronized increasing trends, contradicting this expectation.

In addition, Figure 3c shows elevation estimates obtained with either fossil-based or isotope-based methods for various sites of the Tibetan Plateau.

Discussion

Among the climate constraints of the nine selected redescriptions (rA–rI), eight contain temperature variables and only one a precipitation variable, pointing to the strong relation between temperature and the present-day mammalian distribution in Asia. This is not unexpected, since temperatures present a strong latitudinal gradient on a large scale while the latitudinal diversity gradient is well known to be the strongest distributional pattern in the natural world [27]. When the modern latitudinal gradient was established is still an open question, however, especially within the terrestrial realm [28].

In this study, we are able to delineate a southern zoogeographic region through time via the association of mammalian bunodonty and a lower limit on the temperature of the driest quarter with redescription rB. The 0°C isotherm of the average temperature in January is one of the criteria separating the modern Palaearctic and Indomalayan

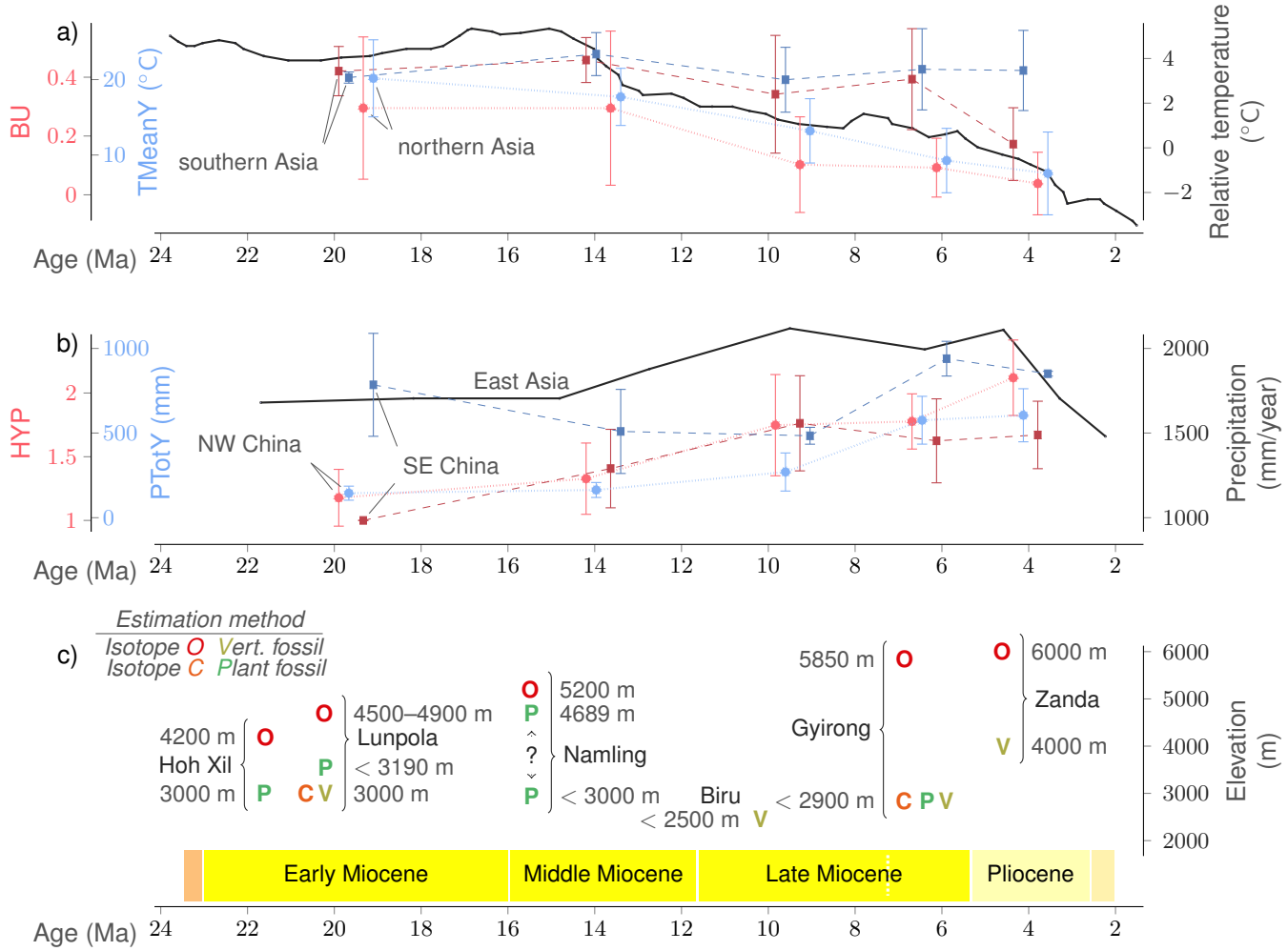


Figure 3 Temperature, precipitation, elevation, bunodonty and hypsodonty trends through the Neogene. a) Global temperature trend (based on [25]), along with bunodonty and mean annual temperature average values in northern and southern Asia. b) Modeled mean annual precipitation for East Asia (based on [26]), along with hypsodonty and annual precipitation average values in northwestern (NW) and southeastern (SE) China. Average values and standard deviations (represented as error bars) are calculated over the localities in each group, which number (n) 18, 37, 23, 51 and 15 in northern Asia, 3, 8, 7, 5 and 4 in southern Asia, 6, 29, 14, 33 and 11 NW China, 7, 8, 7, 20 and 6 in SE China, respectively for the five time intervals. c) Elevation estimates for the Tibetan Plateau (data resources in Supplementary Table 1).

realms [29]. The southern region satisfying rB, and such that $[2.0 \leq \text{TMeanDryQ}]$, hence corresponds well to the Indomalayan realm.

The prevalence of bunodonty (BU) was at its highest in both southern and northern Asia during the warm Early and Middle Miocene (Figure 3a), and the northern boundary of rB lay further north than today (Figure 2a and b). The maximum expansion of the warm and humid zone during the Middle Miocene coincides with the Middle Miocene Climate Optimum, around 17–15 Ma [25]. Following the significant cooling at the end of the Middle Miocene, the prevalence of bunodont species declined notably in both regions. Mammalian communities with high proportions of bunodonts disappeared permanently from high and middle latitudes and retreated to tropical and subtropical regions in the Late Miocene and Pliocene (Figure 2c–e). This suggests that the modern Indomalayan realm was established

most likely after the Middle Miocene Climate Transition, around 15–13 Ma [25, 30], which is consistent with results from both marine fossil evidence [31] and phylogenetic modeling [15].

The increase of mean hypsodonty in the Middle Miocene relates to the immigration of hypsodont/mesodont rhinocerotids and bovids from West Asia and Europe [32]. This coincides with the appearance of grasslands vegetation in Asia during the Middle Miocene [33]. The overall low hypsodonty in faunal communities in our data suggests grassland and dry conditions must have been very limited in East Asia during this interval [32, 33].

The notable increase of mean hypsodonty in the early Late Miocene can be linked to the prevailing aridity at middle latitudes in East Asia [12] and the prominent expansion of grassland in the Asian inland [34, 35]. Concurrent with the decline of bunodonty in Asia, the immigration and subsequent radiation of hypsodont *Hipparion* and bovids in Eurasia during the early Late Miocene [32, 36] led to a substantial increase of hypsodonty (HYP) (Figure 3b) and other traits associated with non-bunodont taxa (OL, OO, OT and SF). High precipitation areas were limited to southern Asia during this interval (Figure 2h). The occurrences of global mid-latitude aridity during the late Neogene have been interpreted as a result of the global climate cooling and are well documented in West Europe, America, East Africa, Australia and China [12, 30, 37–42].

The high point of mean ordinated hypsodonty in southeastern China during the early Late Miocene is at odds with the simulated precipitation peak [26] (Figure 3b). This apparent disagreement might be due to the difference in the considered regions. Fossil locality data mainly comes from localities lying at the middle latitudes of East China, whereas the simulations of Farnsworth *et al.* [26] are for the whole East Asia.

During the Middle Miocene and early Late Miocene the prevalence of hypsodonty is comparable and undergoes a similar increase across mammalian communities of northwestern and southeastern China. From the late Late Miocene (ca. 7 Ma), a disparity appears between northwestern and southeastern China, which strengthens during the Pliocene. Relatively higher precipitation combined with lower mean ordinated hypsodonty start to dominate in southeastern China, in contrast to northwestern China where lower precipitation and higher average hypsodonty prevail. Such changes in regional precipitation regimes have been suggested to link to the onset or intensification of summer monsoon in East Asia and aridity in central Asia [32, 43, 44] as a result of the significant uplift or lateral extension of the Tibetan Plateau around this time [43, 45].

In the Pliocene, the mammalian communities of the Hengduan Mountains and central Myanmar show a drop in the fraction of bunodont species (Figure 2d and e), indicating that warm-favoring fauna disappeared from these areas. The low temperatures to which such mammalian communities point is consistent with the appearance of an altitudinal vegetation zonation [46–48] and the absence of previously prevailing middle-sized hominids in this area during the Pliocene [49]. Today's Hengduan Mountains are identified as a montane forest biodiversity hotspot [6], our results support the emergence of this hotspot during the Pliocene [50]. In the Pliocene, the fossil communities of the Tibetan Plateau are characterized by species endemic to the central Plateau [51], on one hand, and snow-adapted Zanda fauna in the southern Plateau [52, 53], on the other hand, indicating that Alpine fauna had already established on the Tibetan Plateau. The contemporaneous presence of Alpine fauna on the Tibetan Plateau and a hotspot of montane forest on its southeast extension in the Hengduan mountains suggests that the modern faunal diversity in Asia has been fully developed since the Pliocene. Our result is consistent with the analysis of biodiversity based on phylogenetic modeling, and bridges the gap between phylogenetic and fossil evidence [5].

Our study further highlights that modern Asian distribution has been shaped by global climate change and the tectonic evolution of the Tibetan Plateau. While the impact from global climate changes is rather clear, the impact of the Tibetan Plateau on the Asian biota and climate during the Neogene is difficult to trace because the orogenic history of the Tibetan Plateau is the subject of much controversy, involving both biotic and abiotic evidence [54–57]. Most of the evidence of a high Plateau relies on oxygen stable isotope paleoaltimetry estimations, but this method comes with many uncertainties [58, 59]. The differences between elevation estimates obtained with oxygen isotope and other proxies are considerable, with the former yielding much higher estimates (Figure 3c). Looking through the lens of dental traits and climate redescriptions, the Tibetan Plateau started to differ from neighboring regions on its southeast extension (e.g. Yunnan), being distinctively colder, only from the early Late Miocene (Figure 2b and c).

This supports a relatively low Tibetan Plateau during the Early Neogene and is consistent with most of the biotic evidence [54].

Redescription mining captured patterns of association between bunodonty and warm conditions (rB), on one hand, and between low hypsodonty and high precipitation (rC), on the other hand, from the present-day dataset. While these associations hold relatively robustly in the late Neogene, they seem much weaker when evaluated on the data from the early Neogene. Many localities at middle latitudes have a low prevalence of bunodonty but mild temperatures during the dry quarter in the Early and Middle Miocene (blue dots in Figure 2a and b), and vice versa for several localities at higher latitudes in the Middle Miocene (red dots in Figure 2b). The disagreement between mean ordinated hypsodonty plotted over time and simulated precipitation seems even more prominent, with practically all of the localities from the northern half of the study area having a low prevalence of hypsodonty but also having fairly dry simulated climate, in the Early and Middle Miocene (red dots in Figure 2f and g). These disagreements might be due to a combination of reasons, including the following two.

First, the modern patterns of associations between dental traits and climatic conditions might not yet have applied in the early Neogene, not least because some of the dental traits characteristic of current large mammal communities were still in early stages of evolution. Indeed, in the Early Miocene the fossil localities in question (blue dots in Figure 2a) host communities with dominant non-bunodont perissodactyls (tapirs, chalicotheres, rhinos), small-sized deers and moschids. Considering the nearest living relatives, an evolutionary perspective that should also be taken cautiously, such fauna suggests warm conditions [60], despite consisting of a low proportion of bunodonts. Higher bunodonty is generally associated with low seasonality, but its relation to temperaure in those times might not have been analogous.

From the Middle Miocene onwards, perissodactyls presumably favoring warm and humid conditions decreased abruptly [61], while bunodont primates, suids and elephants occupied the forest habitats [32], forming high bunodont and low hypsodont mammalian communities. The same patterns of association between dental traits and climatic conditions as today are expected to apply to them. The warm and humid conditions in the high latitudes during the Miocene are not only supported by the presumably forest-adapted taxa *Pliopithecus*, *Kubanochoerus* and *Gomphotherium* [62–66], there is also ample evidence of warm and humid conditions related to the Middle Miocene Climatic Optimum at high latitudes and inland areas [34, 67–69]. Hence, this first explanation is plausible for the Early Miocene, but not so later on.

Second, the simulated climate variables might not adequately reflect the prevailing conditions not least due to large uncertainties in the prevailing physical boundary conditions (e.g. atmospheric carbon dioxide, orbital forcings, geography, topography and vegetation), especially in the early Neogene. Indeed, prominent paleoclimate models have been found to underestimate temperatures at high latitudes for the middle Miocene [70], which matches with our observation. Furthermore, when looking at the predictions of paleoclimate models, it is generally understood to be more meaningful to consider relative values and trends of changes rather than absolute values, which are more difficult to predict reliably. A bias-correction procedure is applied to the predictions of paleoclimate models, increasing the robustness of the obtained values. Yet, because climate queries contain strict thresholds on absolute values, their evaluation in the past can be sensitive to systematic deviations that might persist in the predictions. This is not such a problem with dental traits, which for the most part are fractional values.

The combination of dental traits and climate is necessary and valuable as it enables us to automatically identify the relevant variables and thresholds and to extract meaningful queries from the present-day dataset thanks to redescription mining. Here, when using the redescriptions to reason about the past, we relied more heavily on the dental traits queries, eventhough such data is of course not free of uncertainties and potential biases. This approach allowed us to shed light on the timing of the emergence of modern zoogeographic regions in Asia. Further studies are needed to resolve the discord between dental traits and climate simulations, especially in the Early and Middle Miocene, which hopefully would shed more light not only on the biogeographic patterns during the early Neogene in Asia, but also provide further methodological insights for modelling mammalian communities and climate.

METHODS

Study area and time intervals

Similarly to our previous study on the present-day ecoregions of China and Southern Asia [22], the area studied in this work is focused on China, the Indian subcontinent and Southeast Asia, that have a distinct climate system, unlike any other region in the world.

We extend our study area from 40°N to 50°N to cover most of Northern China, a region that was excluded from our previous research, but comprises rich Neogene fossils localities. We obtain the fossil data for countries in this area, namely China, India, Pakistan, Bangladesh, Sri Lanka, Burma, Laos, Buhdan, Cambodia, Thailand, and Vietnam. The fossil data span from the Early Miocene to the Pliocene (23–2.5 Ma), split into five time intervals (Table 2), namely Early Miocene (23.03–15.97 Ma, 23 localities), Middle Miocene (15.97–11.63 Ma, 42 localities), early Late Miocene (11.63–7.246 Ma, 27 localities), late Late Miocene (7.246–5.333 Ma, 56 localities) and Pliocene (5.333–2.58 Ma, 17 localities).

The datasets of modern species occurrences and of bioclimatic variables (sites $\times$ species and sites $\times$ climate, respectively) come from [18]. We use square grid cells of 50 $\times$ 50 km as units of analysis.

Species occurrence data

Present-day species occurrence data originally come from the list of the International Union for Conservation of Nature (IUCN, https://www.iucn.org/) processed by Oksanen *et al.* [18].

Fossil occurrence data are downloaded from the *New and Old Worlds* (NOW) database (https://nowdatabase.org/) For the purpose of this study we updated the Asian fossil data in NOW following published literature.

We focus on the large herbivorous mammals which we select by taxonomic orders. That is, we focus on Perissodactyla, Artiodactyla, Primates, and Proboscidea. We only consider sites (present-day grid cells and fossil localities) that report at least three species of large mammalian herbivores.

Dental traits data

The functional characteristics of the teeth of plant eating mammals, such as crown height, scale isometrically with the size of the animal [71], allowing to directly describe animal communities in terms of the distribution of their functional dental characteristics. The most common macroscopic characteristic is hypsodonty, which describes how tall a tooth is in relation to its width or length. The more hypsodont, the more durable to wear the tooth is. Mean hypsodonty of a community has been widely used as a proxy for precipitation [8]. The proxy predicts precipitation primarily, although not exclusively, due to hypsodonty being common in grass-dominated ecosystems or otherwise open habitats but rare in (temperate) woody habitats, and open habitats in turn being typically drier than woody habitats. Pointed or rounded structures on the surface of a tooth are called cusps, while ridges of enamel connecting cusps are called lophs. Bunodonty refers to a dental morphology with separate cusps that are not fused into elongated, lophlike structures. This dental morphology is typical for omnivores and frugivores, including most suids and primates. Faunal communities with a high proportion of bunodont species are typically found in relatively warm and humid forested environments that lack seasonal differences. Another characteristic commonly used to estimate climatic conditions is the presence of longitudinal lophs. Acute lophs and obtuse lophs respectively designate sharp edges and blunt edges across the chewing direction. Globally, the greater the prevalence of taxa having lophs in the community, the cooler the temperature is expected to be [19] in the harsh season. High average loph counts are typically associated with cold habitats and plant food which is harder to chew (especially during the harsh season). Structural fortifications are reinforcement of cusps making them more prominent and resistant to dental wear. Occlusal topography refers to the surface of a teeth being flat or non-flat.

Because the redescriptions from our previous work [21] could not capture the specificities of South–East Asia sufficiently well, here we consider an extended tailored set of dental traits. The core scoring scheme follows Žliobaitė *et al.* [17], but we score acute lophs (AL) for selenodonts as introduced by Oksanen *et al.* [18], and include two extra

traits, namely bunodonty (BU) and exclusively obtuse lophs (OO) as described by Saarinen *et al.* [23]. Both traits can be uniquely determined from the other traits [17]. Specifically, BU is positive if none of AL, OL, OT and SF is, that is, there are no lophs of any kind. OO, as the name suggests, is positive if OL is positive, but none of AL, OT and SF is. All in all, we use seven dental traits variables to describe morphological characteristics of molar teeth: hypsodonty (HYP), bunodonty (BU), presence acute lophs (AL), presence of obtuse or basin-like lophs (OL), structural fortification of cusps (SF), flatness of occlusal topography (OT), and exclusively obtuse lophs (goat-type) (OO). HYP is ordinal with three distinct values, whereas the other variables are binary (the characteristic is either present or absent).

We start from the scores for present-day dental traits from Galbrun *et al.* [22], which we update, including scoring AL for selenodonts and additional traits of BU and OO. We score dental traits for the fossil species specifically for this study. Dental traits are primarily scored at the species level from images in the literature and museum photos from our own archives. When no information for a particular species is available, we assign the dominant score of the genus, tribe, subfamily or family that the species belongs to, as relevant. The scores used in this study are provided in [24].

To prepare the datasets of dental traits for sites as needed for mining and evaluating redescriptions, we average individual dental traits of species occurring at each site. Hence, we obtain a dataset with seven dental traits variables (Table 1), such that each variable takes values in the unit interval and records the fraction of species in the mammalian community of the considered site in which the characteristic is present, with the exception of HYP, which is the average over the occurring species of the ordinal values.

Climate data

On the climate side, we use the standard nineteen bioclimatic variables known as *Bioclim* (Table 1).

The modern observational data (*WorldClim2*, http://www.worldclim.com/version2) are originally based on Fick and Hijmans [72]. We prepare the present-day dataset following the methodology described in Galbrun *et al.* [22]. In particular, the data are mapped to the same 50×50 km grid to match the species occurrences and dental traits data.

To prepare the past climate dataset, we used paleoclimate models to simulate climate data, specifically monthly precipitation, maximum and minimum temperature (averaged over multiple years), from which the bioclimatic variables are derived and interpolated for each fossil locality.

For each time interval, we select a paleoclimate model and associated parameters, as indicated in Table 2. In particular, the *Norwegian Earth System Model* (*NorESM*) [73] is used to obtain global climate model simulations at 20 Ma with pCO_2 of 420 ppm (denoted as `EarlyMio_Nor_420`) and 560 ppm (denoted as `EarlyMio_Nor_560`), and at 10 Ma with pCO_2 of 350 ppm (denoted as `LateMio_Nor_350`), for the Early Miocene, the Middle Miocene and the early Late Miocene, respectively. The same mid-Pliocene (3 Ma) simulation (`MidPlio_405`), obtained using the *Community Climate System version 4* (*CCSM4*) [74] with data provided by *ecoClimate* [75] (https://www.ecoclimate.org) is selected for both the late Late Miocene and the Pliocene.

To probe the impact of the choice of paleoclimate model and parameters on the results, we considered additional simulations, including using the *ECHAM5/MPIOM* global climate model [76] and the *COSMO-CLM* regional climate model [44, 77]. In particular, *COSMO-CLM* was run with different elevations of the Asian mountain ranges, to perform sensitivity experiments. The elevation was varied from low (250 m altitude) to high (present-day height, above 5000 m altitude) either for all Asian mountain ranges, to represent their bulk uplift through the geological periods, or for selected mountain ranges (e.g. southern Tibetan Plateau, central Tibetan Plateau, Zagros and Tianshan-Altai Mountains), to represent the outward growth of the Tibetan Plateau. In these experiments, both the Late Miocene and the present-day global climate data were used as lateral boundary conditions to better cover the uncertainties of global climate forcing in simulating the Miocene and Pliocene climate in Asia. In total, we considered twenty-nine paleoclimate model simulations (cf. [24] for further details).

We downscale and calibrate the climate model simulations following a procedure similar to previous studies (e.g., [72, 75, 78]), aiming to limit and correct potential biases. First, we compute the anomalies between each simulation and the baseline climate simulation (present-day control run) for each variable at its original spatial resolution. Second, we bilinearly regrid the anomalies to the 10×10 min grid used by the baseline present climate data of *WorldClim2* (http://www.worldclim.com/version2) with tools provided by *Earth System Modelling Framework* (ESMF

version 8.0.0, https://earthsystemmodeling.org, https://github.com/esmf-org/esmf/releases/tag/ESMF_8_0_0). Third, we apply the regridded anomalies to the baseline present-day climate data from *Worldclim2*.

Next, the nineteen bioclimatic variables are derived from the resulting monthly climate data, using the `biovars` function of the `dismo` R package (version 1.1-4, https://cran.r-project.org/web/packages/dismo/). Finally, the bioclimatic variables are interpolated to each fossil locality based on the nearest neighbors.

Redescription mining

In this study, we are looking for associations between dental traits and climate conditions. In this context, a redescription provides two sets of constraints, called rules or queries, expressed as thresholds over dental traits variables and over bioclimatic variables, respectively. Given a dataset, each query is associated with the collection of sites where the constraints are satisfied. A pair of queries, respectively in terms of dental traits and of climate conditions, that are satisfied roughly at the same sites, can be interpreted as indicating the existence of a local pattern of association between the variables, and is called a redescription. The more similar the two collections of sites, the stronger the pattern. This is the intuition behind the redescription mining task, which aims at identifying and statistically evaluating such patterns of associations between variables in a dataset. Different algorithms have been proposed [79] for this task. Here, as in our previous work [21, 22], we perform the analysis with the `Siren` interface [80], using the `ReReMi` algorithm [81]. This greedy algorithm constructs the queries step by step. It generates conditions by selecting a variable and setting the associated thresholds, using as candidates the values that appear in the data and aiming to maximize the redescription accuracy. For more information about these tools and about the connections between redescription mining and more classical methods such as regression and clustering, in the context of biogeography, we refer the reader to our previous work [22].

In the context of this study, each redescription consists of a set of constraints over dental traits variables, the dental traits query denoted as $q_{\mathbf{D}}$, and a set of constraints over bioclimatic variables, the climate query denoted as $q_{\mathbf{C}}$. An example redescription is

$$q_{\mathbf{D}} = [\mathrm{SF} \leq 0.222] \text{ AND } [\mathrm{BU} \leq 0.357] \qquad q_{\mathbf{C}} = [\mathrm{TMeanY} \leq 15.4]\,,$$

which reads as (note that redescriptions are bidirectional)

> sites where the fraction of species with structural fortification of cusps (SF) is lower than 22.2% and the fraction of bunodont species (BU) is lower than 35.7% also often have a mean annual temperature (TMeanY) lower than 15.4°C, and vice versa.

In the present-day dataset, 3497 grid cells satisfy both the dental traits constraints and the climate constraints. In other words, the redescription holds true at 3497 gird cells (depicted as purple dots in Figure 1 rA). This represents 54.61% of the 6406 grid cells in the dataset, which we call the relative support of the redescription and denote as $\mathrm{supp}\,\%$. In this dataset, a further 160 grid cells satisfy the dental traits constraints but not the climate constraints (depicted as red dots), and 170 grid cells satisfy the climate constraints but not the dental traits constraints (depicted as blue dots).

The accuracy of a redescription is measured by the Jaccard coefficient, denoted as J, that is, the number of sites where both queries of the redescription hold, divided by the number of sites where either of the queries hold. In this case it is equal to 3497/(3497+160+170) = 0.914, which is quite close to the maximum value of one. Note that for a redescription to be highly accurate, the two sets of constraints do not have to be satisfied in a large proportion of the sites, but there should be few sites where one is satisfied but not the other.

In summary, our example redescription captures a strong pattern that holds in the northern half of studied area of the present-day dataset, linking low fractions of species with structural fortifications as well as of species without any lophs, on one hand, and relatively low annual temperatures, on the other hand.

Analysis protocol

Our analysis proceeds in two main steps. First, we extract and select redescriptions from the present-day dataset, following Galbrun *et al.* [22]. Second, we evaluate the redescriptions on the past dataset, examining which queries hold where at each of the different time intervals.

Rather than reusing the redescriptions from Galbrun *et al.* [22], we rerun the mining process on the present-day dataset. We do this in order to update the dataset with respect to the study area and the considered dental traits (cf. Section 8). Running the mining process on the updated present-day dataset, we obtained 151 redescriptions. The run took about half an hour on a commodity laptop. We ranked the resulting redescriptions by decreasing accuracy and filtered out redundant ones. That is, we removed any redescription having more than 90% of its support in common with a more accurate one. This left us with 34 redescriptions. Then we selected the most accurate redescriptions that offer a good coverage of the study area as well as of the different dental traits variables. Except for dental trait OT, that is not represented among accurate redescriptions, all dental traits appear in at least one of the nine most accurate redescriptions. By clustering and visualizing these redescriptions, we checked that they provide a good coverage of the study area. Henceforth, we thus focus our analysis on these nine selected redescriptions, denoted as rA to rI, and the corresponding patterns in the fossil data.

We put our dataset for the past together by selecting a collection of fossil localities and collating the corresponding values for the dental traits variables and the bioclimatic variables, as explained in Section 8). The selected redescriptions were then evaluated on this dataset, to determine the status of each query at every fossil locality. In particular, for a given fossil locality and redescription, we can determine whether the distribution of dental traits over the mammalian community at the locality satisfies the constraints of the dental query, on one hand, and whether the modelled bioclimatic variables at the locality satisfy the climate query, on the other hand. Hence, for a chosen time interval and redescription, we can plot the status of the redescription at each of the fossil localities of the time interval on a map, using the same color code as explained above (i.e. purple, red, blue and gray representing localities where both the dental traits and climatic queries, where only the dental traits query, where only the climatic query, and where neither queries are satisfied, respectively).

Data availability

All prepared input data and results analyzed during the current study are available in [24], where they can also be visualized as maps.

Data sources:

- Present-day climate: *WorldClim2* (http://www.worldclim.com/version2) [72]
- Present-day species occurrence: International Union for Conservation of Nature (IUCN, https://www.iucn.org/), processed by [18]
- Fossil occurrence data: *New and Old Worlds* (NOW) database (https://nowdatabase.org/)

Code availability

To allow full reproducibility of our analysis, the scripts that fully automate the processing pipeline as described in the manuscript, along with a copy of the source data that we used, are available in [24].

Tools used for preparing paleoclimate model simulations:

- *Norwegian Earth System Model* (*NorESM*) [73]
- *Community Climate System version 4* (*CCSM4*) [74] with data provided by *ecoClimate* [75] (https://www.ecoclimate.org)
- *Earth System Modelling Framework* (ESMF version 8.0.0, https://earthsystemmodeling.org, https://github.com/esmf-org/esmf/releases/tag/ESMF_8_0_0)
- `dismo` R package (version 1.1-4, https://cran.r-project.org/web/packages/dismo/)

Tools used for data analysis:

- `Siren` interface [80], using the `ReReMi` algorithm [81] (https://gitlab.inria.fr/egalbrun/siren, commit f154c53b)

REFERENCES

[1] Wallace, A. R. *The Geographical Distribution of Animals, with a Study of the Living and Extinct Faunas, as Elucidating the Past Changes of the Earth's Surface* Vol. 1 (Harper, 1876). URL https://www.nature.com/articles/014186a0

[2] Olson, D. M. *et al.* Terrestrial ecoregions of the world: A new map of life on earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience* **51** (11), 933–938 (2001). https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

[3] Fu, C. *et al.* in *Regional-global interactions in east asia* (eds Tyson, P. *et al.*) *Global-Regional Linkages in the Earth System* Global Change — The IGBP Series (closed), 109–149 (Springer, 2002)

[4] Kreft, H. & Jetz, W. A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography* **37** (11), 2029–2053 (2010). https://doi.org/10.1111/j.1365-2699.2010.02375.x

[5] He, J., Lin, S., Li, J., Yu, J. & Jiang, H. Evolutionary history of zoogeographical regions surrounding the tibetan plateau. *Communications Biology* **3** (1), 1–9 (2020). https://doi.org/10.1038/s42003-020-01154-2

[6] Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403** (6772), 853–858 (2000). https://doi.org/10.1038/35002501

[7] Qiu, Z. & Li, C. Evolution of chinese mammalian faunal regions and elevation of the qinghai-xizang (tibet) plateau. *Science in China Series D: Earth Sciences* **48** (8), 1246–1258 (2005). https://doi.org/10.1360/03yd0523

[8] Fortelius, M. *et al.* Fossil mammals resolve regional patterns of eurasian climate change over 20 million years. *Evolutionary Ecology Research* **4** (7), 1005–1016 (2002). URL http://evolutionary-ecology.com/abstracts/v04/1439.html

[9] Sun, X. & Wang, P. How old is the asian monsoon system?—palaeobotanical records from china. *Palaeogeography, Palaeoclimatology, Palaeoecology* **222** (3), 181–222 (2005). https://doi.org/10.1016/j.palaeo.2005.03.005

[10] Fortelius, M. *et al.* Late miocene and pliocene large land mammals and climatic changes in eurasia. *Palaeogeography, Palaeoclimatology, Palaeoecology* **238** (1), 219–227 (2006). https://doi.org/10.1016/j.palaeo.2006.03.042

[11] Li, S.-F. *et al.* Orographic evolution of northern tibet shaped vegetation and plant diversity in eastern asia. *Science Advances* **7** (5), eabc7741 (2021). https://doi.org/10.1126/sciadv.abc7741

[12] Liu, L., Eronen, J. T. & Fortelius, M. Significant mid-latitude aridity in the middle miocene of east asia. *Palaeogeography, Palaeoclimatology, Palaeoecology* **279** (3), 201–206 (2009). https://doi.org/10.1016/j.palaeo.2009.05.014

[13] Favre, A. *et al.* The role of the uplift of the qinghai-tibetan plateau for the evolution of tibetan biotas. *Biological Reviews* **90** (1), 236–253 (2015). https://doi.org/10.1111/brv.12107

[14] Mosbrugger, V., Favre, A., Muellner-Riehl, A., Päckert, M. & Mulch, A. in *Cenozoic evolution of geo-biodiversity in the tibeto-himalayan region* (eds Hoorn, C., Perrigo, A. & Antonelli, A.) *Mountains, Climate and Biodiversity* 429–449 (Wiley-Blackwell, 2018)

[15] Feijó, A. *et al.* Mammalian diversification bursts and biotic turnovers are synchronous with cenozoic geoclimatic events in asia. *Proceedings of the National Academy of Sciences* **119** (49), e2207845119 (2022). https://doi.org/10.1073/pnas.2207845119

[16] Eronen, J. T. *et al.* Precipitation and large herbivorous mammals i: estimates from present-day communities. *Evolutionary Ecology Research* **12** (2), 217–233 (2010). URL http://www.evolutionary-ecology.com/abstracts/v12/2538.html

[17] Žliobaitė, I. *et al.* Herbivore teeth predict climatic limits in kenyan ecosystems. *Proceedings of the National Academy of Sciences* **113** (45), 12751–12756 (2016). https://doi.org/10.1073/pnas.1609409113

[18] Oksanen, O., Žliobaitė, I., Saarinen, J., Lawing, A. M. & Fortelius, M. A humboldtian approach to life and climate of the geological past: Estimating palaeotemperature from dental traits of mammalian communities. *Journal of Biogeography* **46** (8), 1760–1776 (2019). https://doi.org/10.1111/jbi.13586

[19] Liu, L. *et al.* Dental functional traits of mammals resolve productivity in terrestrial ecosystems past and present. *Proceedings of the Royal Society B: Biological Sciences* **279** (1739), 2793–2799 (2012). https://doi.org/10.1098/rspb.2012.0211

[20] Žliobaitė, I. Concept drift over geological times: predictive modeling baselines for analyzing the mammalian fossil record. *Data Mining and Knowledge Discovery* **33** (3), 773–803 (2019). https://doi.org/10.1007/s10618-018-0606-6

[21] Galbrun, E., Tang, H., Fortelius, M. & Žliobaitė, I. Computational biomes: The ecometrics of large mammal teeth. *Palaeontologia Electronica* **21** (1), 1–31 (2018). https://doi.org/10.26879/786

[22] Galbrun, E., Tang, H., Kaakinen, A. & Žliobaitė, I. Redescription mining for analyzing local limiting conditions: A case study on the biogeography of large mammals in china and southern asia. *Ecological Informatics* **63**, 101314 (2021). https://doi.org/10.1016/j.ecoinf.2021.101314

[23] Saarinen, J. *et al.* Pliocene to middle pleistocene climate history in the guadix-baza basin, and the environmental conditions of early homo dispersal in europe. *Quaternary Science Reviews* **268**, 107132 (2021). https://doi.org/10.1016/j.quascirev.2021.107132

[24] Liu, L. *et al.* The emergence of modern zoogeographic regions in asia examined through climate–dental trait association patterns (2024). URL https://github.com/zliobaite/redescription-asia-neogene. Companion Github repository

[25] Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 ma to present. *Science* **292** (5517), 686–693 (2001). https://doi.org/10.1126/science.1059412

[26] Farnsworth, A. *et al.* Past east asian monsoon evolution controlled by paleogeography, not CO2. *Science Advances* **5** (10), eaax1697 (2019). https://doi.org/10.1126/sciadv.aax1697

[27] Mittelbach, G. G. *et al.* Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters* **10** (4), 315–331 (2007). https://doi.org/10.1111/j.1461-0248.2007.01020.x

[28] Mannion, P. D. A deep-time perspective on the latitudinal diversity gradient. *Proceedings of the National Academy of Sciences* **117** (30), 17479–17481 (2020). https://doi.org/10.1073/pnas.2011997117

[29] Chen, L., Song, Y. & Xu, S. The boundary of palaearctic and oriental realms in western china. *Progress in Natural Science* **18** (7), 833–841 (2008). https://doi.org/10.1016/j.pnsc.2008.02.004

[30] Flower, B. P. & Kennett, J. P. The middle miocene climatic transition: East antarctic ice sheet development, deep ocean circulation and global carbon cycling. *Palaeogeography, Palaeoclimatology, Palaeoecology* **108** (3), 537–555 (1994). https://doi.org/10.1016/0031-0182(94)90251-8

[31] Fenton, I. S., Aze, T., Farnsworth, A., Valdes, P. & Saupe, E. E. Origination of the modern-style diversity gradient 15 million years ago. *Nature* **614** (7949), 708–712 (2023). https://doi.org/10.1038/s41586-023-05712-6

[32] Fortelius, M. *et al.* Evolution of neogene mammals in eurasia: Environmental forcing and biotic interactions. *Annual Review of Earth and Planetary Sciences* **42** (1), 579–604 (2014). https://doi.org/10.1146/annurev-earth-050212-124030

[33] Strömberg, C. A. Evolution of grasses and grassland ecosystems. *Annual Review of Earth and Planetary Sciences* **39** (1), 517–544 (2011). https://doi.org/10.1146/annurev-earth-040809-152402

[34] Hui, Z. *et al.* Miocene vegetation and climatic changes reconstructed from a sporopollen record of the tianshui basin, NE tibetan plateau. *Palaeogeography, Palaeoclimatology, Palaeoecology* **308** (3), 373–382 (2011). https://doi.org/10.1016/j.palaeo.2011.05.043

[35] Barbolini, N. *et al.* Cenozoic evolution of the steppe-desert biome in central asia. *Science Advances* **6** (41), eabb8227 (2020). https://doi.org/10.1126/sciadv.abb8227

[36] Janis, C. in *An evolutionary history of browsing and grazing ungulates* (eds Gordon, I. J. & Prins, H. H. T.) *The Ecology of Browsing and Grazing* Ecological Studies, 21–45 (Springer, 2008)

[37] Janis, C. Evolution of horns in ungulates: Ecology and paleoecology. *Biological Reviews* **57** (2), 261–318 (1982). https://doi.org/10.1111/j.1469-185X.1982.tb00370.x

[38] MacFadden, B. J. *Fossil Horses: Systematics, Paleobiology, and Evolution of the Family Equidae* (Cambridge University Press, 1992)

[39] Barry, J. C. *et al.* Faunal and environmental change in the late miocene siwaliks of northern pakistan. *Paleobiology* **28**, 1–71 (2002). https://doi.org/10.1666/0094-8373(2002)28[1:FAECIT]2.0.CO;2

[40] Ivanov, D., Ashraf, A. R., Mosbrugger, V. & Palamarev, E. Palynological evidence for miocene climate change in the forecarpathian basin (central paratethys, NW bulgaria). *Palaeogeography, Palaeoclimatology, Palaeoecology* **178** (1), 19–37 (2002). https://doi.org/10.1016/S0031-0182(01)00365-0

[41] Böhme, M. The miocene climatic optimum: evidence from ectothermic vertebrates of central europe. *Palaeogeography, Palaeoclimatology, Palaeoecology* **195** (3), 389–401 (2003). https://doi.org/10.1016/S0031-0182(03)00367-5

[42] Jiménez-Moreno, G. & Suc, J.-P. Middle miocene latitudinal climatic gradient in western europe: Evidence from pollen records. *Palaeogeography, Palaeoclimatology, Palaeoecology* **253** (1), 208–225 (2007). https://doi.org/10.1016/j.palaeo.2007.03.040

[43] An, Z., Kutzbach, J. E., Prell, W. L. & Porter, S. C. Evolution of asian monsoons and phased uplift of the himalaya–tibetan plateau since late miocene times. *Nature* **411** (6833), 62–66 (2001). https://doi.org/10.1038/35075035

[44] Tang, H., Micheels, A., Eronen, J. T., Ahrens, B. & Fortelius, M. Asynchronous responses of east asian and indian summer monsoons to mountain uplift shown by regional climate modelling experiments. *Climate Dynamics* **40** (5), 1531–1549 (2013). https://doi.org/10.1007/s00382-012-1603-x

[45] Yang, S. & Ding, Z. Seven million-year iron geochemistry record from a thick eolian red clay-loess sequence in chinese loess plateau and the implications for paleomonsoon evolution. *Chinese Science Bulletin* **46** (4), 337–340 (2001). https://doi.org/10.1007/BF03187199

[46] Kou, X.-Y., Ferguson, D. K., Xu, J.-X., Wang, Y.-F. & Li, C.-S. The reconstruction of paleovegetation and paleoclimate in the late pliocene of west yunnan, china. *Climatic Change* **77** (3), 431–448 (2006). https://doi.org/10.1007/s10584-005-9039-5

[47] Sun, B.-N. *et al.* Reconstructing neogene vegetation and climates to infer tectonic uplift in western yunnan, china. *Palaeogeography, Palaeoclimatology, Palaeoecology* **304** (3), 328–336 (2011). https://doi.org/10.1016/j.palaeo.2010.09.023

[48] Su, T. *et al.* Post-pliocene establishment of the present monsoonal climate in SW china: evidence from the late pliocene longmen megaflora. *Climate of the Past* **9** (4), 1911–1920 (2013). https://doi.org/10.5194/cp-9-1911-2013

[49] Li, S. *et al.* Uplift of the hengduan mountains on the southeastern margin of the tibetan plateau in the late miocene and its paleoenvironmental impact on hominoid diversity. *Palaeogeography, Palaeoclimatology, Palaeoecology* **553**, 109794 (2020). https://doi.org/10.1016/j.palaeo.2020.109794

[50] Huang, Y. *et al.* Vegetation diversity and distribution in the pliocene of the southern hengduan mountains region. *Biodiversity Science* **30** (11), 22295 (2022). https://doi.org/10.17520/biods.2022295

[51] Li, Q. *et al.* Vertebrate fossils on the roof of the world: Biostratigraphy and geochronology of high-elevation kunlun pass basin, northern tibetan plateau, and basin history as related to the kunlun strike-slip fault. *Palaeogeography, Palaeoclimatology, Palaeoecology* **411**, 46–55 (2014). https://doi.org/10.1016/j.palaeo.2014.06.029

[52] Deng, T. *et al.* Out of tibet: Pliocene woolly rhino suggests high-plateau origin of ice age megaherbivores. *Science* **333** (6047), 1285–1288 (2011). https://doi.org/10.1126/science.1206594

[53] Wang, X., Tseng, Z. J., Li, Q., Takeuchi, G. T. & Xie, G. From 'third pole' to north pole: a himalayan origin for the arctic fox. *Proceedings of the Royal Society B: Biological Sciences* **281** (1787), 20140893 (2014). https://doi.org/10.1098/rspb.2014.0893

[54] Deng, T. & Ding, L. Paleoaltimetry reconstructions of the tibetan plateau: progress and contradictions. *National Science Review* **2** (4), 417–437 (2015). https://doi.org/10.1093/nsr/nwv062

[55] Renner, S. S. Available data point to a 4-km-high tibetan plateau by 40 ma, but 100 molecular-clock papers have linked supposed recent uplift to young node ages. *Journal of Biogeography* **43** (8), 1479–1487 (2016). https://doi.org/10.1111/jbi.12755

[56] Su, T. *et al.* No high tibetan plateau until the neogene. *Science Advances* **5** (3), eaav2189 (2019). https://doi.org/10.1126/sciadv.aav2189

[57] Spicer, R. A. *et al.* Why 'the uplift of the tibetan plateau' is a myth. *National Science Review* **8** (1), nwaa091 (2021). https://doi.org/10.1093/nsr/nwaa091

[58] Wang, Y. *et al.* Stable isotopic variations in modern herbivore tooth enamel, plants and water on the tibetan plateau: Implications for paleoclimate and paleoelevation reconstructions. *Palaeogeography, Palaeoclimatology, Palaeoecology* **260** (3), 359–374 (2008). https://doi.org/10.1016/j.palaeo.2007.11.012

[59] Botsyun, S. *et al.* Revised paleoaltimetry data show low tibetan plateau elevation during the eocene. *Science* **363** (6430), eaaq1436 (2019). https://doi.org/10.1126/science.aaq1436

[60] Deng, T., Wang, W. & Ming Yue, L. Recent advances of the establishment of the shanwang stage in the chinese neogene. *Vertebrata Palasiatica* **41** (4), 314–323 (2003). URL http://www.vertpala.ac.cn/EN/abstract/abstract924.shtml

[61] Bai, B., Meng, J., Janis, C. M., Zhang, Z.-Q. & Wang, Y.-Q. Perissodactyl diversities and responses to climate changes as reflected by dental homogeneity during the cenozoic in asia. *Ecology and Evolution* **10** (13), 6333–6355 (2020). https://doi.org/10.1002/ece3.6363

[62] Qiu, Z. & Guan, J. A lower molar of pliopithecus from tongxin, ningxia hui autonomous region. *Acta Anthropologica Sinica* **5** (3), 201–207 (1986). URL http://www.anthropol.ac.cn/EN/abstract/abstract256.shtml

[63] Harrison, T., Delson, E. & Jian, G. A new species of pliopithecus from the middle miocene of china and its implications for early catarrhine zoogeography. *Journal of Human Evolution* **21** (5), 329–361 (1991). https://doi.org/10.1016/0047-2484(91)90112-9

[64] Wu, W. Y., Meng, J. & Ye, J. The discovery of pliopithecus from northern junggar basin, xinjiang. *Vertebrata Palasiatica* **41** (1), 76–86 (2003). URL http://www.vertpala.ac.cn/EN/Y2003/V41/I01/76

[65] Deng, T. New material of hispanotherium matritense (rhinocerotidae, perissodactyla) from laogou of hezheng county (gansu, china), with special reference to the chinese middle miocene elasmotheres. *Geobios* **36** (2), 141–150 (2003). https://doi.org/10.1016/S0016-6995(03)00003-2

[66] Zhang, Z. & Harrison, T. A new middle miocene pliopithecid from inner mongolia, china. *Journal of Human Evolution* **54** (3), 444–447 (2008). https://doi.org/10.1016/j.jhevol.2007.09.005

[67] Sun, J. & Zhang, Z. Palynological evidence for the mid-miocene climatic optimum recorded in cenozoic sediments of the tian shan range, northwestern china. *Global and Planetary Change* **64** (1), 53–68 (2008). https://doi.org/10.1016/j.gloplacha.2008.09.001

[68] Zan, J., Fang, X., Yan, M., Zhang, W. & Lu, Y. Lithologic and rock magnetic evidence for the mid-miocene climatic optimum recorded in the sedimentary archive of the xining basin, NE tibetan plateau. *Palaeogeography, Palaeoclimatology, Palaeoecology* **431**, 6–14 (2015). https://doi.org/10.1016/j.palaeo.2015.04.024

[69] Song, Y. *et al.* Mid-miocene climatic optimum: Clay mineral evidence from the red clay succession, longzhong basin, northern china. *Palaeogeography, Palaeoclimatology, Palaeoecology* **512**, 46–55 (2018). https://doi.org/10.1016/j.palaeo.2017.10.001

[70] Steinthorsdottir, M. *et al.* The miocene: The future of the past. *Paleoceanography and Paleoclimatology* **36** (4), e2020PA004037 (2021). https://doi.org/10.1029/2020PA004037

[71] Fortelius, M. Ungulate cheek teeth: Developmental, functional, and evolutionary interrelations. *Acta Zoologica Fennica* **180**, 1–76 (1985)

[72] Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37** (12), 4302–4315 (2017). https://doi.org/10.1002/joc.5086

[73] Zhang, Z. *et al.* Aridification of the sahara desert caused by tethys sea shrinkage during the late miocene. *Nature* **513** (7518), 401–404 (2014). https://doi.org/10.1038/nature13705

[74] Rosenbloom, N. A., Otto-Bliesner, B. L., Brady, E. C. & Lawrence, P. J. Simulating the mid-pliocene warm period with the CCSM4 model. *Geoscientific Model Development* **6** (2), 549–561 (2013). https://doi.org/10.5194/gmd-6-549-2013

[75] Lima-Ribeiro, M. S. *et al.* EcoClimate: a database of climate data from multiple models for past, present, and future for macroecologists and biogeographers. *Biodiversity Informatics* **10** (2015). https://doi.org/10.17161/bi.v10i0.4955

[76] Micheels, A. *et al.* Analysis of heat transport mechanisms from a late miocene model experiment with a fully-coupled atmosphere–ocean general circulation model. *Palaeogeography, Palaeoclimatology, Palaeoecology* **304** (3), 337–350 (2011). https://doi.org/10.1016/j.palaeo.2010.09.021

[77] Tang, H., Micheels, A., Eronen, J. & Fortelius, M. Regional climate model experiments to investigate the asian monsoon in the late miocene. *Climate of the Past* **7** (3), 847–868 (2011). https://doi.org/10.5194/cp-7-847-2011

[78] Brown, J. L., Hill, D. J., Dolan, A. M., Carnaval, A. C. & Haywood, A. M. PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Scientific Data* **5** (1), 180254 (2018). https://doi.org/10.1038/sdata.2018.254

[79] Galbrun, E. & Miettinen, P. *Redescription Mining* SpringerBriefs in Computer Science (Springer, Cham, 2017)

[80] Galbrun, E. & Miettinen, P. Mining redescriptions with siren. *ACM Transactions on Knowledge Discovery from Data (TKDD)* **12** (1), 6:1–6:30 (2018). https://doi.org/10.1145/3007212

[81] Galbrun, E. & Miettinen, P. From black and white to full color: Extending redescription mining outside the boolean world. *Statistical Analysis and Data Mining* **5** (4), 284–303 (2012). https://doi.org/10.1002/sam.11145

ACKNOWLEDGMENTS

We thank Thomas Denk from the Swedish Museum of Natural History for critical reading of this manuscript and insightful comments. This work was supported by grants from the Research Council of Finland (314803 I.Ž., 316799 A.K., 341620 L.L., 341622 A.K., 341623 I.Ž.).

AUTHOR CONTRIBUTIONS STATEMENT

E.G., H.T., A.K. and I.Ž. conceived the study; L.L. and A.K. selected and curated fossil localities. L.L. coded fossil herbivore dental traits, L.L. and I.Ž. coded modern herbivore dental traits. H.T., Z.Z. and Z.Z. provided the data from paleoclimate model simulations. E.G. carried out the algorithmic analysis of the data and visualized results. L.L.

712 and E.G. led the writing of the manuscript. I.Ž. and H.T. contributed critically to drafts of the manuscript. All authors
713 finalized the text.

714 COMPETING INTERESTS STATEMENT

715 The authors declare no competing interests.

TABLES

Table 1 List of the dental traits and bioclimatic variables. Temperature and precipitation are measured respectively in degrees Celsius (°C) and in millimeters (mm).

Dental traits variables	
HYP	Average ordinated hypsodonty
AL	Fraction of taxa with acute lophs
OL	Fraction of taxa with obtuse lophs
SF	Frac. of taxa with structural fortification of cusps
OT	Frac. of taxa with flat occlusal topography
OO	Frac. of taxa with exclusively obtuse lophs
BU	Frac. of taxa without any lophs (bunodonts)

Bioclimatic variables		
TMeanY	bio1	Mean Annual Temperature
TMeanRngD	bio2	Mean Diurnal Range
TIso	bio3	Isothermality
TSeason	bio4	Temperature Seasonality
TMaxWarmM	bio5	Max Temperature of Warmest Month
TMinColdM	bio6	Min Temperature of Coldest Month
TRngY	bio7	Annual Temperature Range
TMeanWetQ	bio8	Mean Temperature of Wettest Quarter
TMeanDryQ	bio9	Mean Temperature of Driest Quarter
TMeanWarmQ	bio10	Mean Temperature of Warmest Quarter
TMeanColdQ	bio11	Mean Temperature of Coldest Quarter
PTotY	bio12	Annual Precipitation
PWetM	bio13	Precipitation of Wettest Month
PDryM	bio14	Precipitation of Driest Month
PSeason	bio15	Precipitation Seasonality
PWetQ	bio16	Precipitation of Wettest Quarter
PDryQ	bio17	Precipitation of Driest Quarter
PWarmQ	bio18	Precipitation of Warmest Quarter
PColdQ	bio19	Precipitation of Coldest Quarter

Table 2 Time intervals and paleoclimate model simulations.

Time interval		Nb. localities	Climate simulation	Model	Time	pCO_2
Early Miocene	23.03–15.97 Ma	23	`EarlyMio_Nor_420`	*NorESM*	20 Ma	420 ppm
Middle Miocene	15.97–11.63 Ma	42	`EarlyMio_Nor_560`	*NorESM*	20 Ma	560 ppm
early Late Miocene	11.63–7.246 Ma	27	`LateMio_Nor_350`	*NorESM*	10 Ma	350 ppm
late Late Miocene	7.246–5.333 Ma	56	`MidPlio_405`	*CCSM4*	3 Ma	405 ppm
Pliocene	5.333–2.58 Ma	17	`MidPlio_405`	*CCSM4*	3 Ma	405 ppm

FIGURE CAPTIONS