

Revising the biogeography of livestock animal domestication

Indrė Žliobaitė

Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland

Running heads

Left = Žliobaitė

Right = Revising the biogeography of livestock animal domestication

Correspondence: I. Žliobaitė, Department of Geosciences and Geography, University of Helsinki, PO Box 64, Helsinki 00014, Finland. email: indre.zliobaite@helsinki.fi

Consult the copyright statement on the inside front cover for non-commercial copying policies.

© 2019 Indrė Žliobaitė

ABSTRACT

Background and aim: Human society relies on four main livestock animals: sheep, goats, pigs, and cattle. All were domesticated at nearly the same time and place. But why only these few and not others? And why are livestock domestication rates not uniform across biogeographic realms?

Perspective and dataset: I survey the global occurrence of large mammalian herbivore genera around 15,000 to 5000 years before present (BP), and compile a dataset characterizing their ecology, habitats, and dental traits.

Methods of analysis: Predictive modelling of the probability of domestication using machine learning (decision trees and logistic regression). I extract patterns from the resulting models to highlight ecological differences between domesticated and non-domesticated genera. I analyse the estimated probabilities of domestication across biogeographic realms and in the context of local climatic conditions.

Conclusions: The most suitable genera for domestication appear to be generalists adapted to persistence in marginal environments of low productivity, largely corresponding to cold, semi-arid climate zones. Although domestication rates varied across continents, potentially suitable candidate animals were rather uniformly distributed across continents. I propose that the rates of domestication across biogeographic realms largely depend upon how much intersection between hot and cold semi-arid climatic zones was available on each continent.

Keywords: biogeography, domestication, ecology, hypsodonty, large mammals, teeth.

INTRODUCTION

Human society today relies on four main livestock animals – sheep, goats, pigs, and cattle – all of which were domesticated at nearly the same time and place, starting about 10,500 years ago in the Fertile Crescent. The Fertile Crescent is a region in the Middle East that lies across the territories of present-day Iraq, Israel, Syria, Lebanon, Egypt, and Jordan; it also reaches some parts of Turkey and Iran, and Cyprus is sometimes included. The name was popularized by James H. Breasted (1914).

Why has *Homo sapiens* domesticated so few animal and plant groups (Diamond, 1997)? Even more intriguing, those species did not evolve in or near the Fertile Crescent. Instead, their wild progenitors immigrated mostly from central Asia (Diamond, 1997).

Curiously, domestication of those livestock animals happened repeatedly at the same place, with ‘genetic analyses detecting multiple domestic lineages for each species’ (Zeder, 2008, p. 11597). Many arguments have been put forward to explain why the place and time of these animals were suitable for domestication (Diamond, 1997; Larson *et al.*, 2014; Zeder, 2017; MacHugh *et al.*, 2017; Crosby, 2006). Why were animals from elsewhere more suitable for domestication than the native animals? And why were animals local to the Fertile Crescent not domesticated there instead of, or in addition to, the ones that were?

The main objective of this study is to investigate to what extent opportunities for early livestock domestication have been exhausted. To analyse options for animal domestication, I ask what potential candidate species were available around the time of first known domestications of livestock (see Table 1). I analyse the ecological, biogeographic, and climatic contexts of large plant-eating mammals contemporary to the time of those early domestications. I aim to deduce common patterns of their physiology, ecology, and dietary preferences, as well as their inferred behaviour. Computationally, I model the probability of domestication as a function of those factors.

To the best of my knowledge, this is the first quantitative analysis of domestication patterns at the global scale. The datasets I compiled for this study are available online (evolutionary-ecology.com/data/3215Appendix.pdf). The scripts for predictive modelling along with the datasets in csv format are on GitHub (<https://github.com/zliobaite/teeth-domestication>).

MATERIALS AND METHODS

The study region is the world at the time of early livestock animal domestication. I consider a time range from around 15,000 to 5 000 years before present (BP).

Domestication has many definitions and interpretations (Décory, 2019), which vary across groups of organisms and contexts of usage. Here I consider domesticated those species upon which humans exert a significant degree of influence over reproduction and care, and which have undergone significant genetic, behavioural, and/or morphological changes compared with their wild progenitors. I consider only livestock animals and do not cover birds.

The study is in two parts: (1) analysis of ecological characteristics of candidate animals; and (2) the biogeographic context of their domestication. For the latter, I divide the world into six regions following designation of biogeographic realms at the present day (Olson *et al.*, 2001; but see Rosenzweig *et al.*, 2012 for alternative zonations): Nearctic (mainly North America), Palearctic (mainly Eurasia), Afrotropic (mainly Africa), Indomalaya (South-East Asia), Australasia, and

Neotropic (South and Central America). Mapping of those realms is given in the Appendix. The realms distinguish large terrestrial areas within which organisms have been evolving in relative isolation over long periods of time, separated by geographic features, such as oceans, broad deserts, and high mountain ranges. I use realms as natural units for analysing and comparing domestication rates across the globe.

The dataset of candidate genera

I worked with a list of 68 genera that might be candidates for domestication. I compiled the list from several sources. First, I used the lists of living mammalian species (Wilson and Reeder, 2005; Nowak, 2018) by assuming that genera alive today were present in the same biogeographic realms within the selected time frame of the past. To these candidates, I added fossil genera from the list in the NOW database of fossil mammals (The NOW Community, 2018). I excluded genera that went extinct before 5000 BP, assuming that they were rare at the times of possible first domestications (e.g. *Hippidion* in South America). Even though such a taxon could have been domesticated just before it went extinct, while going towards extinction it would have comprised fewer and fewer individuals (Žliobaite *et al.*, 2017); also, perhaps, its persistence in native environments would have become more challenging, even if domestication were to have been attempted. I report analysis at the level of genus rather than species, since genus is a more robust unit of analysis for the past (Eronen *et al.*, 2011); congeneric species rarely co-occur (Levin *et al.*, 2012); and patterns of diversity of genera closely follow the patterns of species (Rosenzweig *et al.*, 2013).

I compiled the candidate list from scratch rather than building on any existing regional list because I aimed for consistency of treatment and I knew of no existing publicly available lists that had aimed for that. A regional compilation of candidate species for southwest Asia by Garrad (1984) exists, but it covers only a small part of the study's scope. Diamond (1997) compiled a global list, which he used for computing domestication rates at a continental scale, but the list of candidate taxa was not made available in his publication.

I focused on herbivorous genera within the body mass range from 40 to 1000 kg, which covers the body mass classes of early domesticated livestock animals. Having livestock animals large enough saves the need to herd countless numbers of individuals. Yet domestic animals cannot be too large, as they would be too challenging to handle and would take too much time and effort to raise to maturity. Therefore, neither small mammals nor megaherbivores, such as elephants, are practical for domestication. I did not include carnivores, but did include omnivores. In retrospect, the most suitable seem to be large herbivores, from about human size

to, at most, an order of magnitude larger. This is in line with carnivore–prey body size relationships in mammals, where preferred prey is from about half to about twice size of the carnivore (Owen-Smith and Mills, 2008). I assume that the lower bound for humans is the same as for predators hunting in the wild, but the upper bound for humans can be higher because of the tools and technology available to humans for managing and killing livestock animals. Thus the selected upper bound is roughly an order of magnitude higher than the body mass of a human.

I characterized each candidate genus by its habitat, dietary characteristics, and functional dental traits (judging by their molars). That resulted in a list of 15 features intended to capture the way of life of a genus. Each feature was coded as a binary or an ordinal variable; I list them in Table 2.

I collected habitat and dietary characteristics from numerous academic sources and animal databases. I compiled this dataset in 2017 and did not record individual sources for the habitat and dietary variables, and cannot provide referencing for each individual data point thereof. Given my choice to work at the resolution of the genus, habitat and dietary variables can be only approximate, since a genus may include species that have different diets or habitats, or their ways of life may have changed between 15,000 years BP and now. Usually, I resolved within-genus variations by using the characteristics of the most common species. In rare cases, when a genus had species that were apparently equally common, I assigned values to the genus’s characteristics that were about half-way between the species. If information about the behaviour now and in the past diverged, I assigned the values of the past. I characterized dental traits following the functional crown type scoring scheme described by Žliobaitė *et al.* (2016), with one modification allowing selenodonts to have acute lophs, following the reasoning of Oksanen *et al.* (2019).

All the variables (except for hypsodonty) were coded on a scale from 0 to 1. I kept hypsodonty with its original ordinate scale (Fortelius *et al.*, 2002) so as not to introduce extra complications in interpretation. A detailed description of the scoring scheme is given in the Appendix.

To estimate domestication rates, I counted domestication of a genus on each continent separately. For example, *Sus* was domesticated in both the Palearctic and Indomalaya; I counted the two as separate events and separate domestication candidacies. If different species of the same genus were domesticated in the same realm, I assigned fractional candidacy for each case. For example, two species of camels were domesticated on the same continent, so each species received 0.5 candidacy. The same applied for *Bos* and *Equus*. In contrast, llamas

and alpacas in the Neotropic realm belong to different genera, so I counted each as a separate candidacy with value 1.

A full dataset along with the list of genera excluded from the candidate list and the reason for doing so are also given in the Appendix. As a summary of the dataset and a sanity-check, Fig. 1 visualizes the first two (scaled) principal components of the dataset. We can see from the plot and the rotation matrix (given in the Appendix) that the primary axis is mostly hypsodonts (open habitat or grazing taxa) versus brachydonts. The second axis is primarily about arid environments vs. wetlands.

By and large, domesticated animals occur in all spaces of this visualization, but the density of coverage is not uniform. In particular, the reader will notice that the highest density of domesticated genera occurs at the arid (bottom-left) end where camelids and goats are located.

Computational methods for predictive modelling

My main computational task is to build predictive models to explain patterns of domestication. In machine learning terms, I need a probabilistic classifier where the contributions of each variable will be easy to interpret. The inputs for the model can be any or all variables given in Table 2. The target variable is a binary-class label indicating whether the genus has been domesticated or not. One of the challenges faced by this predictive modelling inquiry is the relatively small sample size, in combination with a high imbalance of classes (13 domesticated genera vs. 55 undomesticated genera).

I used two types of models: logistic regression and decision tree (for an introductory text to machine learning, see Witten *et al.*, 2016). Logistic regression models a linear relationship between the input variables and the class label. The output is passed through an s-shaped (sigmoid) function to make sure that the prediction falls between 0 and 1. In this case, the output can be interpreted approximately as the probability of domestication.

A number of alternative procedures for estimating the model parameters exist. I chose a combination of LASSO and Ridge optimization [$\alpha = 0.5$ (for details, see Hastie *et al.*, 2009)]. This selection keeps the weights from becoming too large, and at the same time minimizes the number of variables selected to be included in the final model. The number of variables included depends on assessment of variance through the course of model-fitting.

Decision tree is a non-linear model, which, conceptually similarly to LASSO, first selects the variable that explains the class the best, then second best to explain the residual, and so on. The main difference is that LASSO does this globally for the whole dataset whereas

decision tree partitions the data and provides local explanations for the subsets. And, LASSO, in principle, is linear (if we disregard the sigmoid), while decision tree is non-linear.

For fitting a tree, I used the standard CART algorithm with the Gini coefficient as the splitting criterion (for details, see Witten *et al.*, 2016). I produced two decision tree models. In one, all 15 variables are candidates. In the second, the variables describing environment and habitat are not used. The second model aimed to capture domestication patterns purely from the perspective of the organism.

I summarize the modelling choices in Table 3. Modelling was done in R software, using packages `glmnet` (for LASSO) and `rpart` (for tree fitting). The code used for the analysis is available on the GitHub repository (<https://github.com/zliobaite/teeth-domestication>), although it is not very well documented.

SUITABILITY FOR DOMESTICATION AND DOMESTICATION RATES

I tried to accomplish two things with my computational analysis: to identify which ecological and environmental characteristics best explain domestication of livestock animals; and to describe quantitatively each biogeographic realm in terms of the estimated suitability of its candidate animals for livestock domestication 15,000 to 5000 years BP.

Which ecological and environmental characteristics best explain domestication

Two inferred decision tree models are shown in Fig. 2. The first tree internally selected the most informative variables from all 15 candidate variables in Table 2. The second tree used only the 10 candidate variables that describe the animals themselves (grazer, specialized, water dependent, dental characteristics), and not the variables that describe the environment or habitat (open habitat, highland habitat, migration, high endemism, tropical).

The first tree can explain the majority of early livestock animal domestication cases (8 out of 13) using only two variables: coming-not-from-tropical-environments and being moderately migratory. This selection suggests that environmental criteria have more explanatory power than physiological criteria. As the principal component analysis already hinted, domesticated animals appear throughout most of the space described by physiological traits, but the counts of successful domestication vary across that space. The highest densities of successful domestication – as the principal component and the tree analysis together suggest – must be at the arid end of non-tropical environments. The fact that moderate migration has such high explanatory power suggests a potential impact of cold and seasonality.

The second tree (the one without environmental characteristics) gives a more entangled, but interesting, pattern. Domesticatable animals described by this tree are those that have high hypsodonty but not those that are exclusively grazers. They are expected to be moderately or highly dependent on water sources and not highly specialized (e.g. not aquatic animals). Yet, only half of the animals at the terminal node ‘Maybe (9/20)’ were actually domesticated, which suggests that physiological characteristics alone, at least those included in my analysis, are not sufficient to explain the domestication patterns. Which animals actually become livestock animals appears to be a combination of biogeographic and ecological factors.

Decision trees work well for visual and structural explanation. But without further modification, they are too coarse for a probabilistic assessment of the suitability of candidate animals for domestication. I wish to assign a suitability score for each candidate genus such that I can rank them to analyse realized vs. potential domestication across biogeographic realms.

The inferred logistic regression model for estimating the probability of domestication is:

$$P(\text{Domestication}) = 1/(1 + \exp(-Z)),$$

where $Z = 0.42 \text{ HYP} - 0.35 \text{ AL} + 0.50 \text{ Water} + 0.40 \text{ Highlands} + 0.14 \text{ Open} - 0.70 \text{ Endemic} - 2.92$

(1)

The model automatically selected the six most informative variables from 14 candidate variables (all variables in Table 2 except for ‘tropical’; a complementary model considering all 15 variables as candidates is given in the Appendix). The regression components can be interpreted similar to those of a regular linear regression, where regression coefficients act as weights and their signs indicate the directionality of the relationship. The model predicts a high probability of domestication if an animal is hypsodont, does not have acute lophs, is strongly associated with highlands, is weakly associated with open environments, is not highly endemic, and is somewhat dependent on the availability of drinking water.

Judging from all three models, I propose that the dominant, but not exclusive, characteristics that make animals domesticatable are their adaptedness to consume abrasive food (hypsodonty) in combination with somewhat marginal, seasonally variable

environmental conditions. This follows from non-tropical generalists having had the highest propensity for livestock domestication.

Estimated suitability for domestication of candidate genera

Now I rank the 68 candidate genera according to their potential suitability for domestication. Are there any taxa that – compared with other domesticated animals – must have been very unlikely to be domesticated? And do any of the non-domesticated taxa look particularly suitable?

Table 4 ranks the candidates from the highest to the lowest probability of domestication. Probabilities are estimated using the logistic regression model given in equation (1). For the reader's information, I also include scores given by both decision trees.

Antilocapra and *Ovibos* are the non-domesticated genera that rank highest on the list. Both are North American taxa, which hints that non-domestication of those animals might have had more to do with the place than the animals themselves. Quite close comes the saiga antelope, which is a mysterious animal of semi-deserts, grasslands, and steppe. Today they are critically endangered and found only on the central Asian steppes. Saiga have been hunted for a long time, and I cannot think why they have not been domesticated.

The next non-domesticated cohort includes *Hemitragus*, *Pseudois*, *Ammotragus*, *Budorcas*, and *Redunca*. All of these except *Redunca* are close taxonomic relatives of goats and sheep, and all are from the Palearctic. Since goats and sheep have already been domesticated in the same region, perhaps people have had no strong need to domesticate similar wild genera instead of breeding and improving already domesticated species.

Among domesticated genera, *Bos* and *Bubalus* rank rather low owing, perhaps, to their being relatively more specialized and relatively more sensitive to food quality compared with the higher cohorts, including sheep, goats, Bactrian camels and horses.

Sus has been domesticated but ranks very, very low. In fact, *Sus* has been domesticated a number of times, which is somewhat surprising. *Sus* is an omnivore with a diet relatively close to that of humans.

Realized and potential rates of domestication across biogeographic realms

Now that we have estimates for suitability for domestication, next we analyse how potential suitability is distributed across biogeographic realms, and how it relates to observed rates of domestication across the realms.

Table 5 gives the observed domestication rates for each realm. It also reports potential domestication rates, which I estimated as follows. Suppose the ten genera given the highest suitability scores by the logistic regression model were domesticated instead of the ten that were actually domesticated. I take those ten most suitable and see in which realms they actually live. Finally, I determine what the domestication rates would be if those ten genera were those actually domesticated. Note that the top ten genera would include *Antilocapra*, *Ovibos*, *Saiga*, and *Hemitragus*, although none of those were domesticated. But *Bos*, *Bubalus*, *Camelus*, and *Sus* would not be included because they are lower in the suitability ranking.

The estimated potential domestication rate assumes that the same number of genera would be potentially domesticated as were domesticated in reality. I used the value ‘ten genera’ to obtain Table 5. However, I also used other values to see how the potential rates of domestication would change. Figure 3 plots potential domestication rates across all the ranking of candidate genera, first assuming that the genus with the highest estimate is domesticated, then, in the second scenario, two genera with the highest scores are domesticated no matter where they are geographically, and so onwards with more genera. Each such scenario corresponds to one integer on the horizontal axis in Fig. 3. This is conceptually similar to received operating characteristics (ROC) curves often used in assessing predictions in medical diagnostics and many other disciplines.

From the vertical alignment of the lines in Fig. 3, we can discern three groups – Palearctic, Nearctic, and Neotropic – which host the largest shares of potentially suitable genera. Their three lines begin and proceed closely together for the first third of the way along the horizontal axis. A second set of realms (Afrotropic and Indomalaya) is a bit lower. Finally, Australasia hosts the fewest potentially suitable genera. All lines meet on the right because, when all 68 genera are potentially domesticated, then the fraction of domesticated genera in each realm is 100%.

Interestingly, while the logistic regression model used for estimating suitability scores does not include any latitudinal information, latitudinal patterns emerge from the analysis. Let us focus the interpretation on the left-hand third of the plot, where potential domestication rates are close to actual domestication rate globally. There, the temperate realms appear to offer roughly four times more potentially suitable taxa than the tropical realms. The Neotropic realm includes a mixture of tropical and non-tropical environments, and appears in between the temperate and tropical groups. Australasia has only one genus large enough to be a candidate and it receives a low estimate, thus Australasia lags well behind the other groups.

The analysis suggests that, perhaps with the exception of Australia, the potential suitability is surprisingly consistent across the biogeographic realms: around 30% in the temperate realm, less than 10% in the tropical realm, and around 20% in the temperate-tropical combination that is the Neotropic realm.

Most interesting is the relation between actual and potential domestication rates. Figure 3 shows that the actual rates (circles and crosses) of domestication in both the Palearctic and Neotropic realms are quite close to their potential rates. The same is true in the Australasia and Afrotropic realms. But Indomalaya has a higher actual rate than its potential rate. And the Nearctic realm (North America) has an actual rate of zero, although its potential rate is as high as that of the Palearctic.

Diamond (1997) argued that uneven rates of domestication across the globe were primarily due to animal characteristics rather than geographic location. My analysis agrees only partially with that. Perhaps the difference arises from the fact that Diamond (1997) counted domestication at the species level, combined South and North America, and did not separate Indomalaya, which is climatically very different from the rest of Asia. He calculated domestication rates of 18% for Eurasia, 4% for the Americas, and 0% elsewhere. The present analysis, which uses biogeographic realms, suggests, in contrast, that while the total number of candidates in South America was very small, South America (Neotropic realm) had very similar actual domestication rates as Eurasia (Palearctic). North America (Nearctic), on the other hand, with many candidates, has had no actual domestications.

The distribution of actual and potential rates of domestication across the realms suggest that while each realm consistently hosted a pool of suitable animals, other factors beside the suitability of animals themselves must have driven domestication rates. But I am baffled at this time by the large North American mismatch between actual and potential rates. As for availability of potentially suitable livestock animals, I hypothesize that a large part of the explanation might be biogeographic in nature, requiring a particular intersection of cold and hot semi-arid climatic zones. In the next section, I discuss evidence for this intersection, along with my interpretations.

BIOGEOGRAPHIC ANALYSIS

To explain suitability scores for domestication across biogeographic realms, I suggest an interplay between climatic zones primarily suitable for plant domestication and animal domestication. My argument goes as follows: a hot semi-arid climate was most suitable for plant domestication, whereas a cold semi-arid climate selected for animals most suitable for

domestication as livestock. The realized rates of domestication, with the exception being North America, reflect how much hot and cold semi-arid climatic zones intersected on each continent.

Biogeography of plant domestication

Plant domestications by humans began at least half a millennium earlier than the first known animal domestications (Larson *et al.*, 2014). Many arguments have been put forward as to why domestication of plants started when it did. Early arguments, such as the Oasis hypothesis (Childe, 1928), favoured climate change. Recently, greater emphasis has been placed on socio-anthropological causes, such as depletion of wild food resources (Diamond, 1997).

The reasons for the time and place of initial plant domestications will undoubtedly turn out to be multiple. Among them is likely to be the warm and seasonal climate of the Fertile Crescent. It favoured a bimodal way of life for plants and animals: feasting during the rich seasons, dormancy during the lean ones. Such bimodality requires effective storage of nutrients, as well as a fast onset and growth at the start of the rainy season. Plants can cope with seasonal aridity, for example, by producing large protein-rich seeds (Diamond, 1997) that are nutritious and preserve well. Such seeds make a nutritious storable material considered to be favourable for domestication.

Assuming that climate stress selects for domestication-friendly plant material, the next question is where in the world such climates could be found at the time of early domestication. Based on the climate of the Fertile Crescent today, Diamond (1997) argued that a Mediterranean climate presents the most favourable climatic conditions for plant domestication. In the Köppen-Geiger climate classification system (Kottek *et al.*, 2006), the Mediterranean climate is characterized by dry summers and mild wet winters with the average temperature of the coolest month between 0°C and 22°C and the cumulative precipitation in the driest summer month less than 30–40 mm. Figure 4a shows areas of Mediterranean climate today. Diamond (1997) argued that out of all the Mediterranean zones, the Fertile Crescent was most favourable because it was farthest from the sea.

Mediterranean climate zones today are ideal for wine production. Their dominant vegetation is scrubby and dense: broad-leaved evergreen shrubs, bushes, and small trees. Plants of this kind are not ideal for plant domestication, and not only because it takes several years for their first fruits to appear. Such vegetation cover would have been labour-intensive for early agriculture, as the land would have needed to be cleared of bushes and shrubs

continuously. A slightly harsher (drier) climate, on the other hand, would have kept large parts of the land naturally clear from woody vegetation.

Indeed, the first domesticated plants were grasses (wheat and barley). Maybe grasses could have been domesticated in a climate harsher than the Mediterranean? The next harsher zone is the semi-arid zone, characterized by low and very seasonal annual precipitation. The precipitation threshold for semi-arid zones depends on whether the main precipitation comes in winter or in summer and how it interplays with the temperature and the length of daylight. Vegetation in these zones is short or scrubby, dominated by grasses or small shrubs. The coverage is depicted in Fig.e 4b.

Semi-arid zones in Fig. 4b coincide with many areas of early agriculture. These areas are shown in Fig. 4c (after Zeder, 2017): the Fertile Crescent (#9), sub-Sahara (#7), Nile Valley (#8), India (#11), North China (#12), Mesoamerica (#3), and Andes (#6). In all these areas, grass plants or herbaceous plants were domesticated.

Several areas of early plant domestication are outside the proposed semi-arid zone pattern, in particular: rice domesticated in South China; bananas, yams, and taro in New Guinea; millet, bean, and burdock in Japan, as well as some parts of the Americas. Of those, rice from China and bananas from New Guinea are perhaps most widely cultivated today. Both places have quite seasonal (highlands in New Guinea) but not particularly arid climates. In contrast, two hot semi-arid areas, the Kalahari in Africa and Australian coastlands, hosted no known plant domestications.

Despite the open questions, the global match between semi-arid climatic zones and areas of early plant domestication is quite striking, suggesting that hot semi-arid zones, denoted as BSh in the Köppen-Geiger climate classification system (Kottek *et al.*, 2006), rather than Mediterranean zones (CSa), might have been the climate that saw the birth of agriculture.

Cold seasonal environments host suitable livestock animals

To exploit fully the advantage of plant domestication required early humans to settle in places suitable for agriculture. Humans who settled in such places would naturally look for domesticable animals nearby. Diamond (1997) pointed out a curious pattern: whereas the first livestock animals were domesticated in the Fertile Crescent, all of them appear to have evolved in Central Asia, suggesting their initial adaptation to much colder seasonal environments. Indeed, my computational analysis (reported in the previous section) hinted

that seasonal semi-arid zones, such as cold steppe, are likely to have hosted animals most suitable for livestock domestication.

Strong seasonality prescribes two modes of existence: the prosperous season during which the ecosystem flourishes, and the lean season during which the ecosystem shuts down to the bare essentials. While small mammals can hibernate or hide (Liow *et al.*, 2009) during the lean season, large mammals living in seasonal environments have little alternative but to migrate in order to survive the lean months. For successful migration, it is critical for the group to follow the leader (Guttal and Couzin, 2010), as commonly observed, for example, in birds. Therefore, seasonal environments must select for a follow-the-leader social structure, particularly in large mammals.

Seasonality requires the accumulation of energy during the rich months in order to survive the lean ones. Therefore, seasonal environments must also select for rapid growth of offspring, ready to migrate when the time comes, and having accumulated sufficient resources to survive the lean season.

Hypsodonty, which characterizes durability of teeth and tolerance of abrasive food, appears to be one of the strongest explanatory variables for domestication in my analysis. Food for domestic animals needs to be common in human habitats and available in bulk, since specialized diets such as meat, fish, nuts, fruits, or aquatic plants would take a lot of effort to collect and would not be economically viable. More importantly, the diet of domestic animals needs to be different from that preferred by humans, since otherwise a more energetically efficient solution for humans would be to eat that food themselves. Grass is available in bulk and requires little maintenance in semi-arid climatic zones, while humans cannot utilize it directly because we are unable to digest cellulose. Not surprisingly, most livestock animals appear to be mixed-feeders that, owing to hypsodonty and digestive specializations, tolerate or even prefer grass.

Fast growth might be associated with seasonality and migration as I have pointed out. But animals experience seasonality in both tropical and temperate environments and, in my analysis, tropical habitats have a strong negative association with domestication. Why is it that some tropical groups, such as antelopes, have never been successfully domesticated? Since the tropical savanna animals have been hunted for a long time, they must be edible. Many African antelopes come from marginal and very seasonal environments and presumably have similar migratory and growth strategies for coping with the lean season, but this time a hot one rather than a cold one.

The main difference is that in cold seasonal environments temperature is the limiting factor, while in hot seasonal environments it is a lack of precipitation. Thus, cold seasonal environments are likely to select for fast growth of fat to help to withstand changing temperatures, while tropical seasonal savannas are likely to prioritize selection to minimize loss of water. Indeed, most grazing antelopes native to such conditions, such as *Oryx*, can survive without drinking water for long periods. [A similar observation comes from isotope analysis of two grazing fossil pig genera that lived in Africa savannas: *Metridiochoerus*, which evolved locally, is thought to have been much less water-dependent than *Kolpochoerus* (Rannikko *et al.*, 2017), which originally came from Asia.] Indeed, in the present analysis water dependency is one of the main explanatory variables for domestication (along with hypsodonty, migration, and non-tropical habitat).

For domestication to be successful, animals should have some sort of an incentive to remain domesticated. An animal could escape predators for years, survive without food for weeks, but a water-dependent animal would find it difficult to survive without water for more than a couple of days. Since humans are water-dependent themselves, they are nearly always in reach of drinking water, and may have the technology for finding and storing water. That could be the main incentive for animals that regularly need drinking water to stay close to humans.

Other variables in my analysis, such as migration, grazing, and highlands have their strongest association with domestication in the middle of their range of values. Such animals can be characterized as ‘migratory, but not too much’, ‘grazer, but not to an extreme’, and ‘highland habitat, but not too high up’. This links back to observations from analyses of evolutionary processes, suggesting that the frontier for speciation, the so-called ‘species factory’ (Bernor *et al.*, 1996; Fortelius *et al.*, 2015), is often not at the extreme ends of the environmental range, but right behind the extreme frontier. Semi-arid environments, which have produced most of the species for domestication (plants and animals), are somewhat like this – they are close to deserts, but not yet deserts.

Following my computational analysis, I propose that cold seasonal habitats, particularly alpine environments, are highly appropriate for producing suitable candidates for livestock domestication. My analysis also showed that animals from all over the trait space have been domesticated, but those of marginal environments (such as goats, llamas, camels, and sheep) were the most often domesticated in my character space. Animals from marginal environments are unlikely to be highly specialized but are efficient processors of low-quality

vegetation of various kinds. This facilitates the conversion of low-quality fibrous vegetation, which humans otherwise could not utilize, into energy for human use.

Crossroads at the Fertile Crescent

While the hot semi-arid climate zones in Fig. 4b appear to have been the most common for plant domestication, the cold semi-arid zones highlighted in the same figure appear often to have produced animals highly suited for domestication. I propose that places where hot semi-arid and cold semi-arid zones meet were the most suitable for early settled human life.

Indeed, the Fertile Crescent is one of several such zones. While the match in Fig. 4b is based on present-day semi-arid climatic zones, the final question for my analysis is how such zones were distributed around 10,000 years ago at the time of the first domestications, and whether their global distribution can potentially explain different rates of domestication across continents.

The best currently accessible global climatic map of the past depicts vegetation at the Last Glacial Maximum, that is 15,000 to 25,000 years BP (Ray and Adams, 2001). Figure 5a depicts my projection of Ray and Adams' (2001) climate reconstruction to the Köppen-Geiger climatic classes, and Fig. 5b shows present-day semi-arid zones (Kottek *et al.*, 2006) along with deserts for a direct comparison. During the early domestication of plants and animals, about 10,000 years BP, the geographic distribution of the climatic zones must have resembled those in Figs. 5a and 5b. Judging from these maps, the most notable changes in the hot semi-arid areas occurred in the Fertile Crescent, Northeast Africa, and Central America. In those locations, hot semi-arid areas were more widespread than today. Two other major changes in semi-arid areas have occurred in South America and Australia. In South America cold semi-arid areas have shrunk whereas in Australia hot semi-arid areas have expanded.

Looking at the distribution of zones around 20,000 years BP (Fig. 5b), only two zones, marked by blue ellipses, appear to provide large enough interfaces between hot and cold marginal environments. By and large it seems that the Fertile Crescent and Central South America areas were the largest interfaces of suitable environments, and those are the continents where the highest rates of domestication occurred.

Climatically speaking, temperate desert mainly forms in Central Europe and North America due to topographic effects and a monsoon climate (Hui Tang, personal communication, January 2018). A monsoon climate makes the summer drier in Central Eurasia, the Arabian Peninsula, and the Mediterranean. These regions are mostly covered by a cold steppe climatic zone (Bsk). North America does not have much in way of a subtropical region, so there is

little room for subtropical desert there. Only Eurasia and Africa have large areas of subtropical aridity. As a result, Arabia becomes the main region where the two climatic zones meet.

North America also seems to have had a reasonably large suitable interface, but at least visually the extreme frontier (hot or cold deserts) is lacking in Fig. 5a, disqualifying the North American semi-arid areas (highlighted in green). Yet, extreme areas are notable in present-day North America. Thus, lack of domestication in North America remains puzzling, particularly given that my analysis has identified about the same average suitability of species as in Eurasia (Palearctic). Indomalaya domestications are few, and do not quite fit the common pattern of cold environments. Perhaps humid and hot tropical settings required locally adapted animals, and the two domesticated ones, *Bubalus* and *Bos* (zebu), were the most suitable locally.

CONCLUDING REMARKS

Human society today relies on a small number of livestock animals. What made these taxa particularly suitable for domestication? Was it the animals themselves, the environment, the geography, the timing, or perhaps the interplay of each of these? As Alfred Crosby (2006) phrased it, domestication was often discontinuous and did not always work. The patterns I analysed are the results of what did work.

The present analysis has shown that the most suitable animals for domestication were generalists adapted to persistence in marginal environments of low productivity, largely corresponding to cold semi-arid climate zones. Such animals would have inhabited temperate steppe or grasslands. At the time of the first domestications, the global distribution of such areas should have been somewhat in between the maps shown in Fig. 5a and Fig. 5b. Yet by far the largest, and likely the only area for such animals to emanate from, is likely to have been Central Asia, and that is where the big four domestic animals are thought to have originated. The Fertile Crescent has been and is the major intersection of hot and cold semi-arid climate zones. If hot semi-arid zones offered suitable plants and cold semi-arid zones offered suitable animals, the intersection of the two zones must have been the ideal place to domesticate both.

Many questions remain. One of them is why animals from marginal hot environments apparently were unsuitable for domestication. Even if animals from cold environments might be preferable due to their growth rates, growth rates do not vary that much with respect to body mass, and this should not automatically disqualify animals from hot climates. The

Sahara Desert is not very likely to have been a barrier either. The difference must be related to adaptations to the tropical habitats. I have hypothesized that water dependency and predator density are important for the follow-the-leader social structure, but some of the hot arid climate grazing antelopes, like waterbuck, are water-dependent, and some, like African buffalo, are quite resistant to predators. I hope for interesting studies to follow.

ACKNOWLEDGEMENTS

I thank Mikael Fortelius, Jared Diamond, Marcus Clauss, Miikka Tallavaara, and Hui Tang for discussions on early versions of the manuscript. Any errors, omissions, or misinterpretations are my own. An earlier version of this paper was submitted to and rejected by *Environmental Archaeology*. I thank three of their anonymous reviewers for feedback, many of whose suggestions I have incorporated here. Finally, I thank Michael Rosenzweig for clinical editing of the text – I really like the result. This is a contribution from the Valio Armas Korvenkontio Unit of Dental Anatomy in Relation to Evolutionary Theory.

REFERENCES

- Bernor, R., Fahlbusch, V., Andrews, P., de Bruijn, H., Fortelius, M., Rogl, F. *et al.* 1996. The evolution of western Eurasian Neogene mammal faunas: a chronologic, systematic, biogeographic, and paleoenvironmental synthesis. In *The Evolution of Western Eurasian Neogene Mammal Faunas* (R.L. Bernor, V. Fahlbusch, and H.-W. Mittmann, eds.), pp. 449–470. New York: Columbia University Press.
- Bibi, F. 2013. A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. *BMC Evol. Biol.*, **13**: 166 [<https://doi.org/10.1186/1471-2148-13-166>].
- Breasted, J.H. 1914. Earliest Man, the Orient, Greece, and Rome. In *Outlines of European History*, Part I by J.H. Breasted and J.H. Robinson. Boston, MA: Ginn & Co.
- Childe, V.G. 1928. *The Most Ancient Near East*. London: Norton.
- Crosby, A. 2006. *Children of the Sun: A History of Humanity's Unappeasable Appetite for Energy*. New York: Norton.
- Décory, M. 2019. A universal definition of ‘domestication’ to unleash global animal welfare progress. *dA.Derecho Animal (Forum of Animal Law Studies)*, **10**(2): 39–55.
- Diamond, J. 1997. *Guns, Germs, and Steel: The Fates of Human Societies*. New York: Norton.

- Diamond, J. 2002. Evolution, consequences and future of plant and animal domestication. *Nature*, **418**: 700–707.
- Eronen, J., Evans, A., Fortelius, M. and Jernvall, J. 2011. Genera are often better than species for detecting evolutionary change in the fossil record: a reply to Salesa *et al.* *Evolution*, **65**: 1514–1516.
- Fortelius, M., Eronen, J., Jernvall, J., Liu, L., Pushkina, D., Rinne, J. *et al.* 2002. Fossil mammals resolve regional patterns of Eurasian climate change over 20 million years. *Evol. Ecol. Res.*, **4**: 1005–1016.
- Fortelius, M., Getitz, S., Gyllenberg, M., Raia, P. and Toivonen, J. 2015. Modeling the population-level processes of biodiversity gain and loss at geological timescales. *Am. Nat.*, **86**: 742–754.
- Garrad, A. 1984. The selection of South-West Asian animal domesticates. In *Animals and Archaeology*, Vol. 3: *Early Herders and Their Flocks* (J. Clutton-Brock and C. Grigson, eds.), pp. 117–132. Oxford: BAR Publishing.
- Guttal, V. and Couzin, I.D. 2010. Social interactions, information use, and the evolution of collective migration. *PNAS*, **107**: 16172–16177.
- Hastie, T., Tibshirani, R. and Friedman, J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edn. New York: Springer.
- Jarman, O. 1974. The social organisation of antelope in relation to their ecology. *Behaviour*, **48**: 215–267.
- Kottek, M., Grieser, J., Beck, C., Rudolf, B. and Rubel, F. 2006: World map of the Köppen-Geiger climate classification updated. *Meteorol. Z.*, **15**: 259–263.
- Larson, G., Piperno, D.R., Allaby, R.G., Purugganan, M.D., Andersson, L., Arroyo-Kalin, M. *et al.* 2014. Current perspectives and the future of domestication studies. *PNAS*, **111**: 6139–6146.
- Levin, S., Carpenter, H., Godfray, Ch., Kinzig, A., Loreau, M., Losos, J. *et al.* 2012. *Princeton Guide to Ecology*. Princeton, NJ: Princeton University Press.
- Liow, L.H., Fortelius, M., Lintulaakso, K., Mannila, H. and Stenseth, N.C. 2009. Lower extinction risk in sleep-or-hide mammals. *Am. Nat.*, **173**: 264–272.
- MacHugh, D.E., Larson, G. and Orlando, L. 2017. Taming the past: ancient DNA and the study of animal domestication. *Annu. Rev. Anim. Biosci.*, **5**: 6.1–6.23.
- Nowak, R. 2018. *Walker's Mammals of the World*. Baltimore, MD: Johns Hopkins University Press.

- Oksanen, O., Žliobaitė, I., Saarinen, J., Lawing, A. and Fortelius, M. 2019. A Humboldtian approach to life and climate of the geological past: estimating palaeotemperature from dental traits of mammalian communities. *J. Biogeogr.*, **46**: 1760–1776.
- Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C. *et al.* 2001. Terrestrial ecoregions of the world: a new map of life on Earth. *Bioscience*, **51**: 933–938.
- Owen-Smith, N. and Mills, M.G.L. 2008. Predator–prey size relationships in an African large-mammal food web. *J. Anim. Ecol.*, **77**: 173–183.
- Rabanus-Wallace, T., Wooller, M.J., Zazula, G.D., Shute, E., Jahren, A.H., Kosintsev, P. *et al.* 2017. Megafaunal isotopes reveal role of increased moisture on rangeland during Late Pleistocene extinctions. *Nature Ecol. Evol.*, **1**: 0125 [<https://doi.org/10.1038/s41559-017-0125>].
- Rannikko, J., Žliobaitė, I. and Fortelius, M. 2017. Relative abundances and palaeoecology of four suid genera in the Turkana Basin, Kenya, during late Miocene to Pleistocene. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, **487**: 187–193.
- Ray, N. and Adams, J. 2001. A GIS-based vegetation map of the world at the Last Glacial Maximum (25,000–15,000 BP). *Internet Archaeol.*, **11** [https://intarch.ac.uk/journal/issue11/rayadams_index.html].
- Rosenzweig, M.L., Drumlevitch, F., Borgmann, K.L., Flesch, A.D., Grajeda, S.M., Johnson, G. *et al.* 2012. An ecological telescope to view future terrestrial vertebrate diversity. *Evol. Ecol. Res.*, **14**: 247–268.
- Rosenzweig, M.L., Buzzard, V., Donoghue, J., II, Lehr, G., Mazumdar, N., Rasmussen, H.M. *et al.* 2013. Patterns in the diversity of the world’s land vertebrate genera. *Evol. Ecol. Res.*, **15**: 869–882.
- The NOW Community. 2018. *New and Old Worlds Database of Fossil Mammals (NOW)*. Licensed under CC BY 4.0 [last retrieved on 1 October 2018 from: <http://www.helsinki.fi/science/now/>].
- Wilson, D.E. and Reeder, D.M., eds. 2005. *Mammal Species of the World: A Taxonomic and Geographic Reference*, 3rd edn. Baltimore, MD: Johns Hopkins University Press.
- Witten, I.H., Frank, E., Hall, M.A. and Pal, C.J. 2016. *Data Mining: Practical Machine Learning Tools and Techniques*, 4th edn. Amsterdam: Elsevier.
- Zeder, M. 2008. Domestication and early agriculture in the Mediterranean Basin: origins, diffusion, and impact. *PNAS*, **105**: 11597–11604.

- Zeder, M.A. 2017. Domestication as a model system for the extended evolutionary synthesis. *Interface Focus*, **7**: 20160133 [<https://doi.org/10.1098/rsfs.2016.0133>].
- Žliobaitė, I., Rinne, J., Toth, A., Mechenich, M., Liu, L., Behrensmeyer, A.K. *et al.* 2016. Herbivore teeth predict climatic limits in Kenyan ecosystems. *PNAS*, **113**: 12751–12756.
- Žliobaitė, I., Fortelius, M., Stenseth, N.C.. 2017. Reconciling taxon senescence with the Red Queen's hypothesis. *Nature*, **552**: 92–95.

Table 1. Domestication of livestock animals

Domestic animal	Ancestor	Place of domestication	Years BP	Digestive type
Sheep	Asiatic mouflon	The Fertile Crescent	10,000	Ruminant
Goat	Bezoar ibex	The Fertile Crescent	10,000	Ruminant
Pig	Eurasian wild boar	The Fertile Crescent	10,000	Non-ruminant
		China	8000	
Cattle (taurine)	Aurochs	The Fertile Crescent	10,000	Ruminant
(zebu)		South Asia	8000	
Horse		Central Asia	6000	Non-ruminant
Donkey	<i>Equus africanus</i>	North Africa	5000	Non-ruminant
Llama		South America	5000	Pseudoruminant
Alpaca	<i>Vicugna</i>	South America	4000	Pseudoruminant
Water buffalo	<i>Bubalus arnee</i>	South Asia	4000	Ruminant
Yak	<i>Bos mutus</i>	Central Asia	4000	Ruminant
Camel (Bactrian)	<i>Camelus ferus</i>	Central Asia	4000	Pseudoruminant
Camel (Dromedary)		North Africa	3000	Pseudoruminant

Note: Domestication time is an approximate time of transition between management of animals and morphological changes associated with domestication, rounded to the nearest thousand years. The list of animals is derived from Diamond (1997), Larson *et al.* (2014), and Zeder (2008).

Table 2. Variables describing characteristics of the candidate genera

Variable	Possible values
1. Grazer	1 = grazer, 0.67 = grass-dominated mixed-feeder, 0.33 = mixed-feeder that takes some grass, 0 = no grass
2. Highly specialized	1 = highly specialized (e.g. aquatic) diet, 0.5 = somewhat specialized, 0 = common grazing or browsing
3. Water dependence	1 = drinking water daily, 0.5 = every 2–3 days, 0 = a week or more without water
4. Open habitat	1 = grassland, open savanna, steppe, 0.5 = woody savanna, bushland, 0 = dense forest
5. Highland habitat	1 = mountains or hills, 0.5 = foothills, 0 = steppe or forests
6. Migration	1 = regular seasonal migration, 0.5 = seasonal change in range, 0 = no migration
7. Highly endemic	1 = highly endemic, 0 = widely distributed
8. Tropical	1 = tropical, 0 = non-tropical (temperate or mountainous) habitats
9. Hypsodonty (HYP)	3 = hypsodont, 2 = mesodont, 1 = brachydont
10. Acute lophs (AL)	1 = acute lophs present in molars, 0 = absent
11. Obtuse lophs (OL)	1 = obtuse lophs present, 0 = absent
12. Fortification (SF)	1 = structural fortification of molar cusps, 0 = no fortification
13. Flatness (OT)	1 = occlusal topography flat, 0 = not flat
14. Goat-like (GO)	1 = obtuse lophs are present, but acute lophs, fortification, and flatness are absent
15. Bunodont (BU)	1 = no lophs, 0 = lophs of any type are present

Table 3. Predictive models used for analysis and modelling choices

Model	Model fit	Variable selection	Parameter selection	Purpose
Decision tree	CART, Gini split	Automated from all 15 variables	Cross-validation	Identifying factors that render livestock animals suitable for domestication
Decision tree	CART, Gini split	Automated from 10 variables	Cross-validation	Explaining domestication only by animals, no geography
Logistic regression	LASSO and Ridge	Automated from 14 variables	Cross-validation	Numeric estimates for suitability for domestication (ranking)

Table 4. Ranking of candidate taxa from the highest to the lowest probability of domestication estimated by the logistic regression model

Genus	Domesticated	Logistic regression	Tree all	Tree organismal
<i>Vicugna</i>	Yes	0.35	0.25	0.12
<i>Antilocapra</i>	No	0.35	0.09	0.33
<i>Ovibos</i>	No	0.35	0.25	0.33
<i>Capra</i>	Yes	0.35	0.89	0.33
<i>Ovis</i>	Yes	0.35	0.89	0.33
<i>Saiga</i>	No	0.31	0.09	0.33
<i>Equus</i> (donkey)	Yes	0.3	0.25	0.8
<i>Lama</i>	Yes	0.3	0.89	0.33
<i>Camelus</i> (Bactrian)	Yes	0.3	0.09	0.33
<i>Hemitragus</i>	No	0.3	0.89	0.33
<i>Pseudois</i>	No	0.3	0.09	0.33
<i>Ammotragus</i>	No	0.3	0.25	0.33
<i>Redunca</i>	No	0.29	0.05	0.2
<i>Budorcas</i>	No	0.28	0.09	0.33
<i>Equus</i> (horse)	Yes	0.27	0.89	0.8
<i>Equus</i> (zebra)	No	0.27	0.05	0.12
<i>Bison</i>	No	0.27	0.09	0.12
<i>Capricornis</i>	No	0.27	0.05	0.33
<i>Connochaetes</i>	No	0.26	0.05	0.12
<i>Cervus</i>	No	0.25	0.09	0.04
<i>Syncerus</i>	No	0.25	0.05	0.8
<i>Bos</i> (taurine)	Yes	0.25	0.89	0.8
<i>Bos</i> (yak)	Yes	0.25	0.89	0.14
<i>Kobus</i>	No	0.25	0.05	0.2
<i>Aepyceros</i>	No	0.25	0.05	0.33
<i>Oreamnos</i>	No	0.25	0.09	0.14
<i>Bubalus</i>	Yes	0.24	0.05	0.2
<i>Bos</i> (zebu)	Yes	0.24	0.05	0.8
<i>Hydrochoerus</i>	No	0.24	0.05	0.2
<i>Phacochoerus</i>	No	0.22	0.05	0.12
<i>Camelus</i> (Dromedary)	Yes	0.22	0.89	0.33
<i>Damaliscus</i>	No	0.22	0.05	0.12
<i>Hippotragus</i>	No	0.21	0.05	0.12
<i>Taurotragus</i>	No	0.2	0.05	0.14
<i>Oryx</i>	No	0.18	0.05	0.14
<i>Alcelaphus</i>	No	0.18	0.05	0.12
<i>Hylochoerus</i>	No	0.17	0.05	0.04
<i>Axis</i>	No	0.17	0.05	0.04
<i>Boselaphus</i>	No	0.17	0.05	0.14
<i>Myotragus</i>	No	0.15	0.25	0.33
<i>Potamochoerus</i>	No	0.14	0.05	0.04
<i>Hippocamelus</i>	No	0.14	0.09	0.04

<i>Pseudoryx</i>	No	0.14	0.05	0.2
<i>Litocranius</i>	No	0.13	0.05	0.04
<i>Pan</i>	No	0.12	0.05	0.04
<i>Sus</i>	Yes	0.12	0.89	0.04
<i>Tragelaphus</i>	No	0.12	0.05	0.04
<i>Beatragus</i>	No	0.12	0.05	0.14
<i>Rangifer</i>	No	0.11	0.09	0.04
<i>Cephalophus</i>	No	0.11	0.05	0.04
<i>Pongo</i>	No	0.1	0.05	0.04
<i>Addax</i>	No	0.1	0.05	0.14
<i>Macropus</i>	No	0.09	0.05	0.04
<i>Tapirus</i>	No	0.09	0.05	0.04
<i>Dicerorhinus</i>	No	0.09	0.05	0.04
<i>Dama</i>	No	0.09	0.25	0.04
<i>Blastocerus</i>	No	0.09	0.05	0.04
<i>Alces</i>	No	0.09	0.25	0.04
<i>Hippo (Madagascar)</i>	No	0.09	0.05	0.04
<i>Hexaprotodon</i>	No	0.09	0.05	0.04
<i>Gorilla</i>	No	0.08	0.05	0.04
<i>Rusa</i>	No	0.08	0.05	0.04
<i>Okapia</i>	No	0.07	0.05	0.04
<i>Odocoileus</i>	No	0.07	0.09	0.04
<i>Babyrousa</i>	No	0.06	0.05	0.04
<i>Elaphurus</i>	No	0.06	0.05	0.04
<i>Hippo (Crete)</i>	No	0.06	0.25	0.04
<i>Rucervus</i>	No	0.05	0.05	0.04

Note: Probabilities estimated by decision trees are also provided. Grey shading represents taxa that have been domesticated.

Table 5. Actual and estimated potential rates of domestication across biogeographic realms

Realm	Number of candidate genera	Genera domesticated	Actual domestication rate (%)	Estimated potential domestication rate (%)
Palearctic	22	6	27	30
Nearctic	10	0	—	30
Neotropic	7	2	28	24
Indomalaya	14	2	14	7
Afrotropic	24	0	—	1
Australasia	1	0	—	0
All	68	10	15%	—

Note: In some realms, several species or subspecies of the same genera were domesticated (e.g. *Bos*); in such cases, a genus is counted only once. See text for method of calculating the potential rate of domestication. The last row ('All') is not a sum of the preceding rows as some genera occur in several realms..

Fig. 1. Visualization of the two components with the highest variance resulting from single-value decomposition (of 15 input variables listed in Table 2). Domestication status is not included in the inputs for the projection.

Fig. 2. Decision tree models for domestication. (a) A model built using all the candidate variables listed in Table 2; (b) a model excluding the environmental variables. The trees are to be read from top to the bottom: for a given candidate animal, each elliptical node indicates a particular variable, and the values recorded indicate which path to take until a square terminal node is reached, which gives an estimate of whether the animal is suitable for domestication. The numbers in a terminal node indicate how many animals with these characteristics have been domesticated and how many animals there are with these characteristics in the dataset.

Fig. 3. Potential domestication rates at different hypothetical values of the number of best candidates (see text). Circles denote the observed rates of domestication. The vertical line denotes the number of genera (10) that were actually domesticated. Horizontal lines are derived from estimated ranking of genera for domestication.

Fig. 4. Present-day Mediterranean and semi-arid climate zones by Köppen-Geiger climate class. (a) Mediterranean climate zones (light yellow: Csa, hot-summer Mediterranean climate; dark yellow: Csb, warm-summer Mediterranean climate). (b) Semi-arid (steppe) climate zones (dark yellow: BSh, hot semi-arid zone; light yellow: BSk, cold semi-arid zone). In dark yellow areas, it does not freeze; in light yellow areas, the coldest month does see freezing temperatures. (c) Areas of domestication of plants and animals with examples for each: (1) eastern North America: chenopods, squash, sunflower, maygrass; (2) Southwest US: turkeys; (3) Mesoamerica: maize, squash, beans, turkeys; (4) northern Peru/Ecuador: squash, lima beans; (5) Amazonia: manioc, yams, peanuts, Muscovy duck; (6) Andes: oca, potato, quinoa, amaranth, llama, alpaca, guinea pigs; (7) sub-Saharan Africa: pearl millet, sorghum, African rice; (8) Horn of Africa/Nile Valley: asses, tef; (9) Near East: wheat, barley, lentils, peas, sheep, goats, taurine cattle, pigs; (10) Central Asia: horses, golden hamster; (11) South Asia: browntop millet, water buffalo, zebu cattle; (12) North China: foxtail, broomcorn millet; (13) South China/South-East Asia: rice, chickens; (14) Japan: barnyard millet, mung bean, burdock; (15) New Guinea: bananas, yams, taro. The list of domesticates is not exhaustive. Map (c) is drawn after Zeder (2017).

Fig. 5. Climate zones in the past and at the present time. (a) Vegetation cover at the Last Glacial Maximum (15,000 to 25,000 years BP) adapted from Ray and Adams (2001). (b) Present-day desert and semi-arid climate zones (Köppen-Geiger climate class B). Blue ellipses indicate intersections of hot and cold semi-arid zones that have arid zones nearby. The green ellipse of 5a over North America indicates an area that doesn't have more extreme zones nearby.

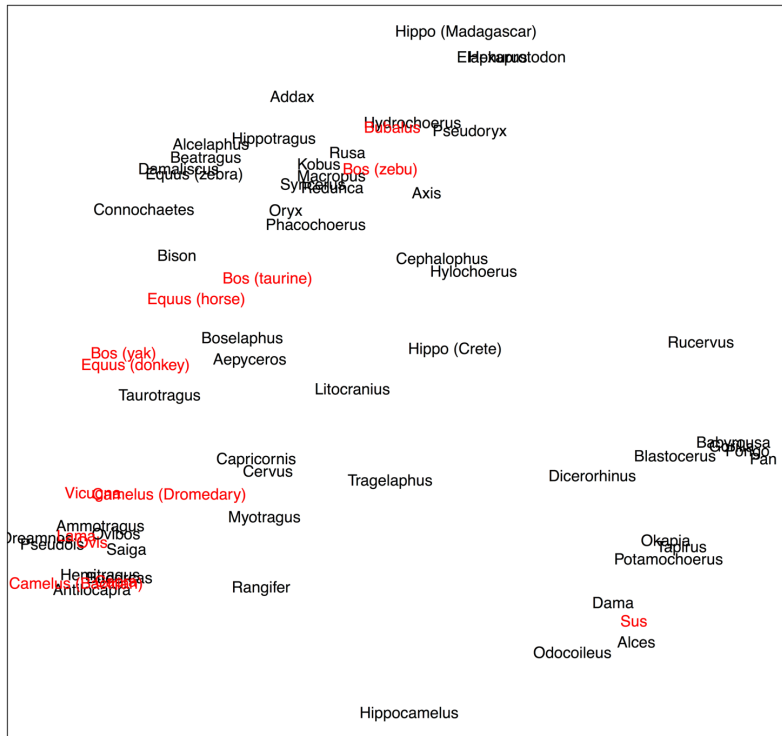
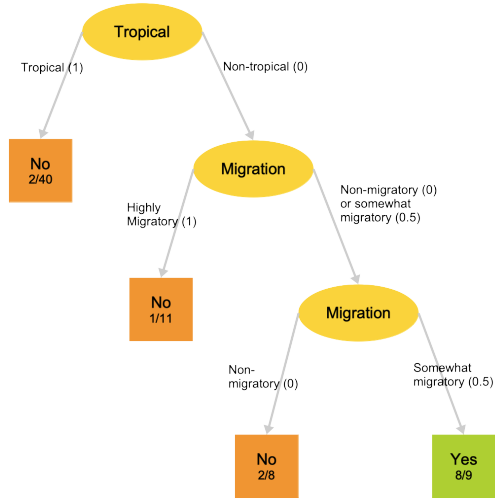
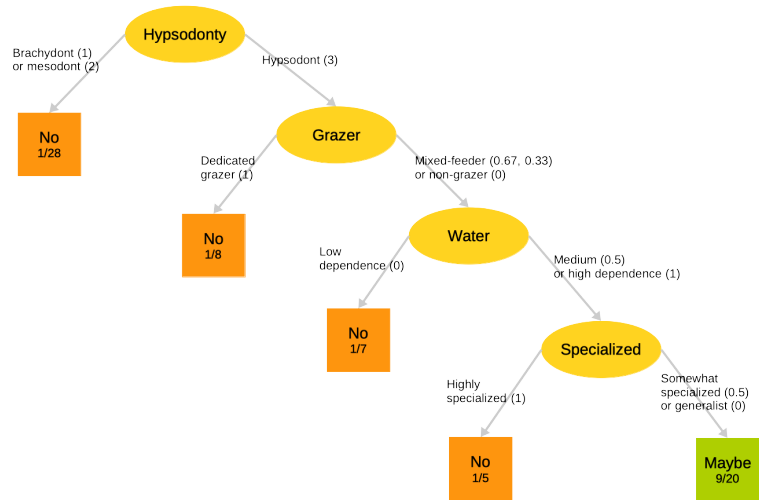


Figure 1. Visualization of two components with the highest variance resulting from singular value decomposition (of 15 input variables listed in Table 2). Domestication status is not included in the inputs for the projection.



(a)



(b)

Figure 2. Decision tree models for domestication. (a) A model built using all the variables given in Table 2 as candidates. (b) A model built excluding environmental variables. The trees are read from top to the bottom: given a candidate animal, each yellow node indicates a variable to look at, and the values for a particular animal indicates which path to take until a terminal square node is reach, which gives an estimate of whether the animal is suitable for domestication. The number in the terminal nodes indicate how many animals with these characteristics have been domesticated and how many animals with these characteristics are in total in our dataset.]

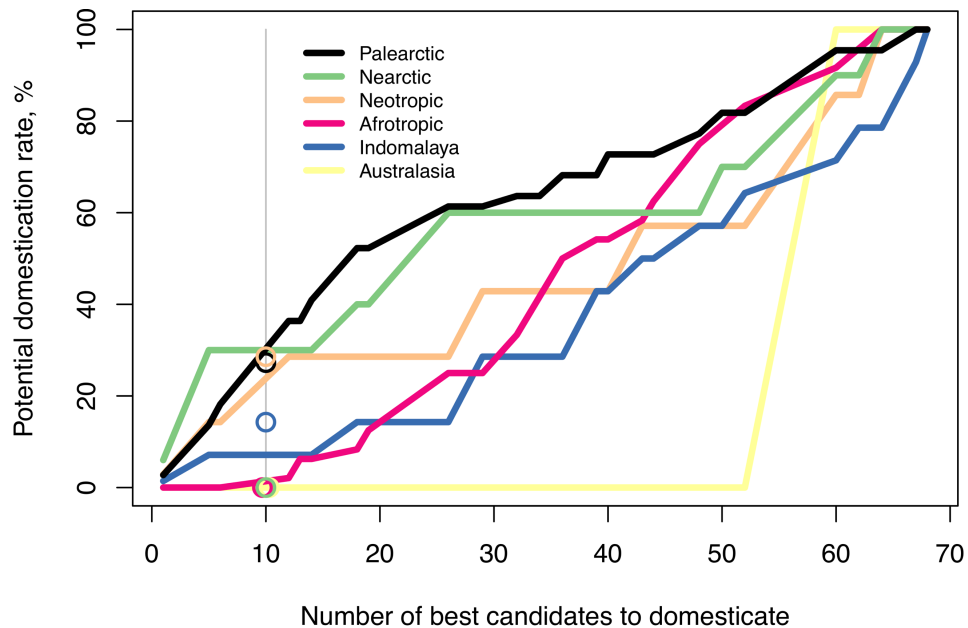


Figure 3. Potential domestication rates considering different number of best candidates. Circles denote real observed rates of domestication. The vertical line denotes the number of genera that were actually domesticated. The horizontal lines are derived from estimated ranking of genera for domestication.

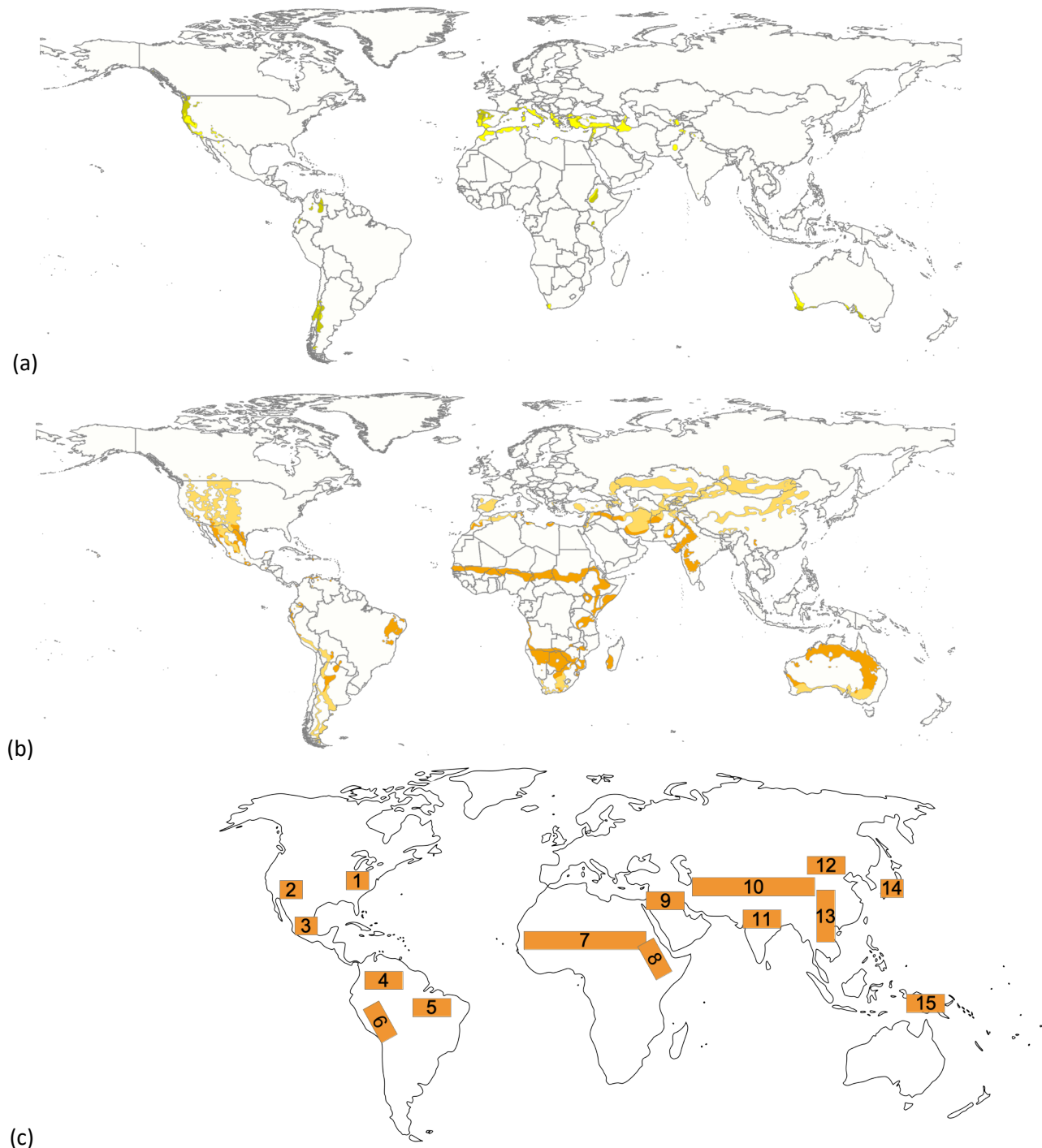


Figure 4. Mediterranean and semi-arid climate zones today by Koppen climate classes. (a) Mediterranean climate zones today (light yellow – Csa, hot-summer Mediterranean climate; dark yellow - Csb, warm-summer Mediterranean climate). (b) Semi-arid (steppe) climate zones today (dark yellow – BSh, hot semi-arid zone; light yellow – BSk, cold semi-arid zone). In dark yellow it does not freeze, in light yellow the coldest month has freezing temperatures. (c) Areas of domestication of plants and animals with examples for each centre: (1) Eastern North America: chenopodia, squash, sunflower, and maygrass; (2) Southwest US: turkeys; (3) Mesoamerica: maize, squash, beans, and turkeys; (4) northern Peru/Ecuador: squash and lima beans; (5) Amazonia: manioc, yams, peanuts, and Muscovy duck; (6) Andes: oca, potato, quinoa, amaranth, llama, alpaca, and guinea pigs; (7) sub-Saharan Africa: pearl millet, sorghum, and African rice; (8) Horn of Africa/Nile Valley: asses, tef; (9) Near East: wheat, barley, lentils, peas, sheep, goats, taurine cattle, and pigs; (10) Central Asia: horses, golden hamster; (11) South Asia: browntop millet, water buffalo, and zebu cattle; (12) North China: foxtail and broomcorn millet; (13) South China/Southeast Asia: rice and chickens; (14) Japan: barnyard millet, mung bean, burdock; (15) New Guinea: bananas, yams, and taro. The list of domesticates not exhaustive. Map (c) is drawn after Zeder (2017).

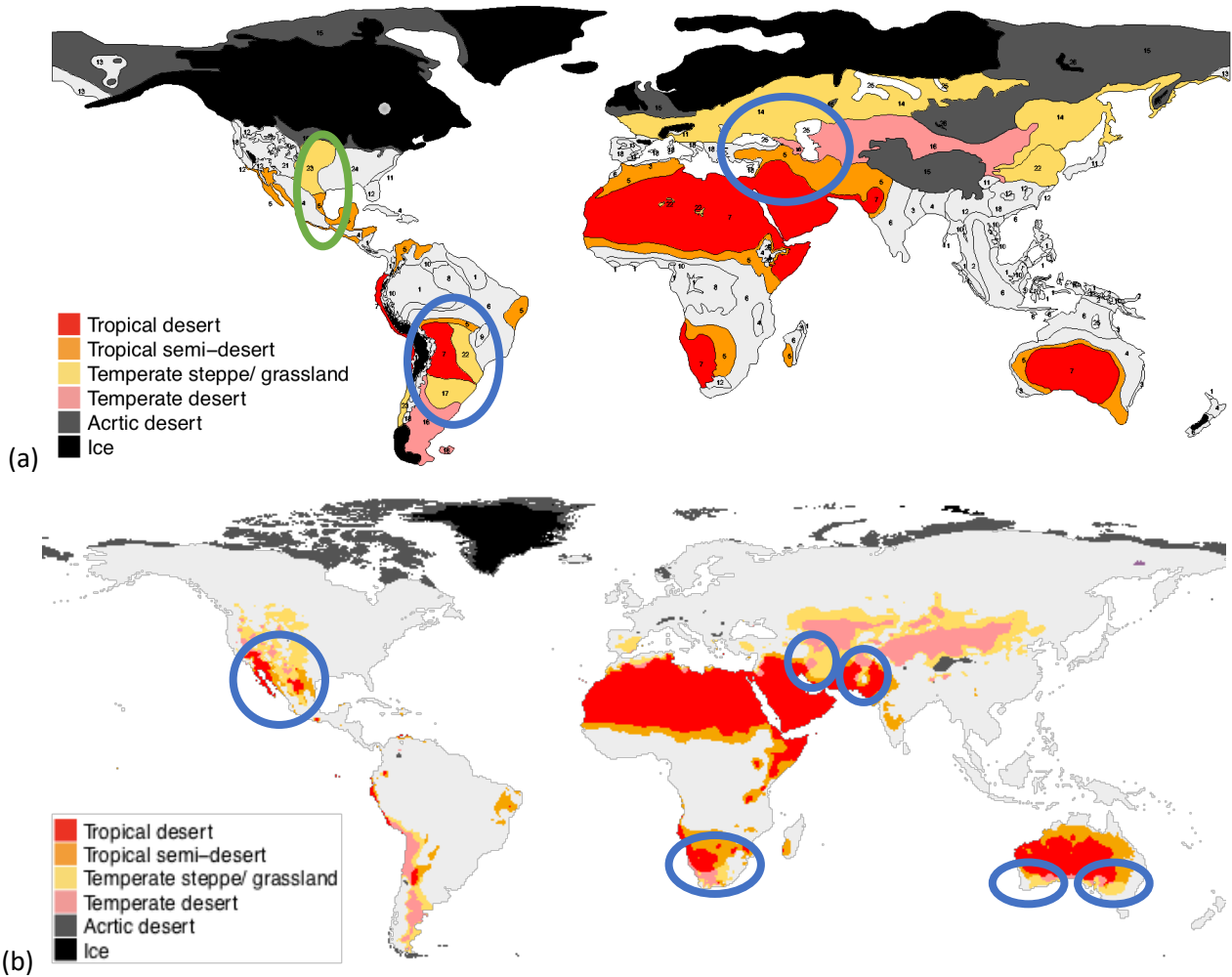


Figure 5. Climate zones in the past and today. (a) Vegetation cover at the Last Glacial Maximum (15 000-25 000), the plot is adapted from Ray and Adams (2001). (b) Desert and Semi-arid climate zones today (Koppen climate class B). Blue ellipses indicate intersections of hot and cold Semi-arid zones with arid zones nearby. Green ellipses indicate the areas without apparent more extreme zones nearby.