

Pascale Gerbault^{a, *}, Michela Leonardi^{b, *}, Adam Powell^{a, c}, Christine Weber^b, Norbert Benecke^d, Joachim Burger^b, Mark G. Thomas^{a, e}

Domestication and migrations: Using mitochondrial DNA to infer domestication processes of goats and horses

* Corresponding Authors: p.gerbault@ud.ac.uk; leonardm@uni-mainz.de (ML and PG have contributed equally to this work, ML for horses, and PG for goats)

a Research Department of Genetics, Evolution and Environment, University College London, United Kingdom

b Palaeogenetics Group, Institute of Anthropology, Johannes Gutenberg-University Mainz, Germany

c AHRC Centre for the Evolution of Cultural Diversity, Institute of Archaeology, University College London, United Kingdom;

d German Archaeological Institute, Scientific Department, Berlin, Germany

e Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Sweden

Abstract

The complex process of animal domestication will have led to important biological changes, including a reshaping of patterns of genetic diversity. Studying those patterns permits inferences to be made on the demographic and trait-selection components of the domestication process, and different mod-

els arising from archaeological and archaeozoological data to be compared. In order to contribute to current understanding of the domestication process, we analysed both modern and ancient mtDNA-sequence data in two domestic species, goats and horses, to compare different models of their demographic histories. The first archaeological evidence for goat domestication, on the southern slopes of the Zagros and Taurus mountains, dates back to between 10,000 and 11,500 BP, while the first evidence for horse domestication seems to date to around 7,000 BP, in modern day Ukraine and Kazakhstan. We examine two sets of increasingly complex demographic models, five for goats and three for horses, based on prior knowledge from archaeological and genetic data, and use coalescent simulation and observed genetic-diversity patterns to compare these models. We discuss the implications of the results and propose further domestication-process hypotheses that could be considered for each species.

Keywords

Domestication, goat, horse, coalescent, mitochondrial DNA, ancient DNA

Introduction

Domestication and archaeology

The southern slopes of the Zagros and Taurus mountains host the earliest archaeological evidence, dating back to between 10,000 and 11,500 BP, for the domestication of sheep, goats, cattle and pigs. It seems that animals were initially managed in these centres and only later develop features of the domesticated forms we see today (Zeder, 2008). Reconstructing the demographic histories of these species is a challenging task and, to date, most of our knowledge comes from archaeozoological investigation of skeletal remains. In this article we show that the additional inferences that can be made on the basis of genetic and palaeogenetic data provide complementary information about the demographic history of human and domesticate populations.

In order to contribute to current understanding of domestication processes, we focus on two of these species, namely goats and horses. The natural distribution of *Capra aegagrus*, the closest wild ancestor of domestic goat (*Capra hircus*) is restricted to Anatolia, the Near and the Middle East, a region that encompasses the Zagros and Taurus mountains, where the earliest evidence for animal domestication is found. As *Capra hircus* is now widely distributed throughout Eurasia and Africa, goat domestication must have first occurred in these regions; and domesticated stock must have spread from there.

Horse domestication differs in at least two important ways from that of goat, sheep, cattle and pigs. First, it seems to have occurred at least 3,000 years later (Anthony and Brown, 2003), and second, archaeological evidence suggests the existence of distinct domestication centres, located mainly in modern day Ukraine and Kazakhstan; some distance from domestication centres for other species. Indeed, isotopic and osteological analyses provide direct evidence of horse milking and riding in a site from the Botai Culture, Northern Kazakhstan, dated to the mid-sixth millennium BP (Outram et al., 2009).

Goat and horse genetic diversity

As domesticates are numerous worldwide, it is expected that the range expansions and migrations from local wild populations would have left some signatures on patterns of genetic diversity in these animals. Mitochondrial DNA (mtDNA) has some features that are well suited to address such questions. First, its uniparental (i.e., maternal) inheritance means all copies in a population are related through a single-gene genealogy that is relatively easy to infer. Second, its mutation rate is high enough (particularly in hypervariable regions I and II) to allow distinct types (haplotypes) to evolve over relatively short time spans. Third, because it is present in many copies per cell, it is considerably easier to obtain from archaeological samples, where DNA degradation can be considerable.

One of the first studies on goat genetic diversity (Luikart et al., 2001) defined 4 highly divergent mtDNA haplogroups in *C. hircus* populations. Two alternative hypotheses were proposed to explain the presence of such high divergence in those domesticates: either (i) they represented four distinct maternal origins (and therefore, four distinct domestication events); or (ii) such diversity already existed in the ancestral population. The latter hypothesis was discarded because the estimated effective-population size necessary to maintain such diversity (between 38,000 and 82,000 effective females) would have been, *a priori*, too high. However, an ancient DNA study has shown that two of these haplogroups, A and C, were already present within an ancient sample from south-eastern France dated to 7,000 years BP (Fernandez et al., 2006). Furthermore, the genetic diversity of domestic goats (Naderi et al., 2007) appears to be a subset of the diversity observed in wild goats (Naderi et al., 2008), at least in terms of presence/absence of mtDNA haplogroups. The principal differences between wild and domestic, and within domestic, populations remain at the level of haplogroup frequencies, with haplogroup A being the more frequent in domestic goats. Therefore, if these highly divergent haplogroups are represented in wild populations, and the genetic diversity observed in domestic goats can be considered a subset of *C. aeg-agrus* diversity, then the “one haplogroup equals one domestication event” approach to inferring population histories should be reconsidered. Indeed, it is now clear that equating haplogroups – which are arbitrarily designated – with ancestral populations or demographic episodes, such as domestication events, is highly problematic (Barbujani et al., 1998; Belle et al., 2006; Burger and Thomas, 2011; Goldstein and Chikhi, 2002; Nielsen and Beaumont, 2009).

For horses, the only wild species still alive is Przewalski’s horse (*Equus przewalskii*). However, it is genetically very different from domestic horse and cannot be considered as a proxy for the ancestral horse population(s) from which domestication occurred. One attempt to investigate horse domestication on the basis of genetic diversity was made by Vilà and colleagues (2001). They analysed a portion of the mtDNA control regions in 191 modern horses from 10 different breeds. They also sequenced ancient DNA from 8 Pleistocene Alaskan horses and 8 northern European archaeological samples dated between 1,000 and 2,000 years ago. They found high diversity and deep coalescence time in mtDNA.

Unsurprisingly, analysis of autosomal microsatellites revealed a higher divergence than for mtDNA. These data indicate that domestic horse mtDNA lineages have an ancient origin that is anterior to the beginning of domestication, and further suggest either multiple domestication events, or domestication from a genetically diverse population, or introgression from many different wild lineages subsequent to domestication. The wild ancient Alaskan group shows low levels of genetic diversity, and even though haplotype diversity differs significantly between modern breeds, each shows usually high genetic diversity. Moreover, there is little evidence of strong mtDNA phylogeographic or breed structuring. In contrast, individuals from the same breed did tend to cluster together in a tree based on 15 hypervariable autosomal microsatellites. These observations led Vilà and colleagues to favour post-domestication introgression of wild horses from a large number of distinct populations and predominantly maternal gene flow in breeding practices (Vilà et al., 2001).

This hypothesis was supported by another study that considered a wider set of sequences (Jansen et al., 2002). In their article, 652 mtDNA sequences from domestic and Przewalski horses were interpreted from a median-joining network representation of the data. The diversity pattern did not show any notable phylogeographic structuring. By eliminating the proportion of mtDNA types that are likely to have arisen in their sample – by the fastest conceivable mutation rate – between the earliest likely date for horse domestication and the present, they estimated that a minimum of 77 wild mares contributed to domestic horse populations. They finally addressed the possibility that these females were domesticated from a single wild horse population. To do so, they used the contemporaneous geographical distribution of mtDNA haplotypes and compared their frequency and diversity with those of wild Przewalski and Alaskan horses. Jansen and colleagues concluded that the most likely scenario is that more than one wild horse population was recruited for domestication. Finally, recent work from Lira and colleagues (2010) suggested a certain degree of continuity between wild Neolithic and modern domesticated Iberian horses.

The coalescent and inference of population demographic histories

The coalescent is a retrospective and probabilistic model of how lineages coalesce, or join, to form the genealogy of a sample of genes under a particular demographic history. When combined with a mutation model, it can be used to make predictions about patterns of genetic variation. In simplified terms, coalescent theory considers the ancestral relationships among gene copies for a sample of genes. Rather than considering the genetic history of a whole population going forward in time, the coalescent looks backwards in time, building gene genealogies only for the samples under consideration. At each step going back one generation in time, two samples “pick” a parent randomly in the population, and whenever they both “pick” the same parent, they coalesce into a single lineage (Rosenberg and Nordborg, 2002). All the lineages eventually coalesce into one single common ancestor, the most recent common ancestor (MRCA) of all the gene copies in a sample. Because selectively neutral mutations do not affect reproduction, they can be superimposed, or ‘sprinkled’ at random, onto the branches of a simulated genealogy afterwards. (Note that an important assumption here is that the mitochondrial variation is selectively neutral.) When the MRCA is found, mutations are added through a Poisson process forwards along the branches of the generated gene genealogy to produce simulated genetic data. Crucially, because one need only consider the ancestors of the sampled gene copies, there is no need to simulate the entire population. This feature makes coalescent simulation computationally very efficient and so an appropriate and powerful tool for investigating domestication processes.

Coalescent simulation does not lead a researcher directly from empirical data to an inference about population-demographic history. Rather, it allows a researcher to examine a range of different demographic history hypotheses. This can be performed by simulating genetic datasets under a range of demographic scenarios and then identifying those scenarios (hypotheses) for which the simulated data are most similar to the observed data (see Nielsen and Beaumont (2009) for a review of some of the methods utilized). This is in sharp contrast to phylogeographic approaches, which assume that population histories can be directly deduced from estimated genealogies (as represented by phylogenetic trees or networks). The coalescent approach has a number of advantages, including the ability to examine a wide range of different hypotheses and accommodation of the fact that genealogies sampled from distinct individuals are random realizations of a stochastic, population-level process. But perhaps most importantly, the coalescent simulation approach, in combination with statistical methods such as Approximate Bayesian Computation (ABC) (Fagundes et al., 2007; Ray et al., 2009), allows different hypotheses to be compared in a robust statistical framework.

In this paper we use coalescent simulation and observed genetic diversity patterns to compare a number of models, of different complexity, for the domestication of goats and horses. The models we consider here are necessarily simplifications of the true domestication processes and are only a small subset of all plausible models. They are, however, informed by archaeological data and allow us to address the following questions concerning these domestication processes: (1) Can the mtDNA diversity pattern observed in wild and domestic goat populations be explained by those populations having originated from a single metapopulation (without significant structuring between them), or are more complex (multi-population) models required? (2) Assuming continuity between wild and domestic horse populations, are patterns of modern and ancient horse mtDNA diversity best explained by domestication of a small proportion of wild horses, or by domestication of most wild horses present at the time?

Material and Methods

Samples

Goats: We used published sequences of domestic (Naderi et al., 2007) and wild goats (Naderi et al., 2008) from Europe and the Middle East, as well as published sequences of ancient domestic goats (Fernandez et al., 2006). When using the term “Middle East” for both *C. hircus* and *C. aegagrus* samples, we refer to samples coming from five countries, namely Azerbaijan, Dagestan, Iran, Pakistan or Turkey. We restricted our Middle-Eastern *C. hircus* samples to the same geographic region as *C. aegagrus* samples for simplicity of comparison. We used the following sequences: 1046 mtDNA sequences of *C. hircus* from all over Europe and an ancestral sample of 19 sequences, 368 mtDNA sequences of *C. hircus* from the Middle-East and 471 sequences of *C. aegagrus* (Fig. 1). We considered 130 base pairs of the mtDNA HVR region, the maximum length available for the 19 published ancient samples (Fernandez et al., 2006).

Horses: We analysed 246 base pairs (bp) from the mtDNA HVR in ancient and modern horses. We obtained sequences for 10 ancient wild samples from Central Europe, dated to the Late Glacial period (17,000–11,600 BP) (Weber, 2005). We then compared those to 1096 published modern horse sequences from different regions and breeds throughout the world (Lira et al., 2010).

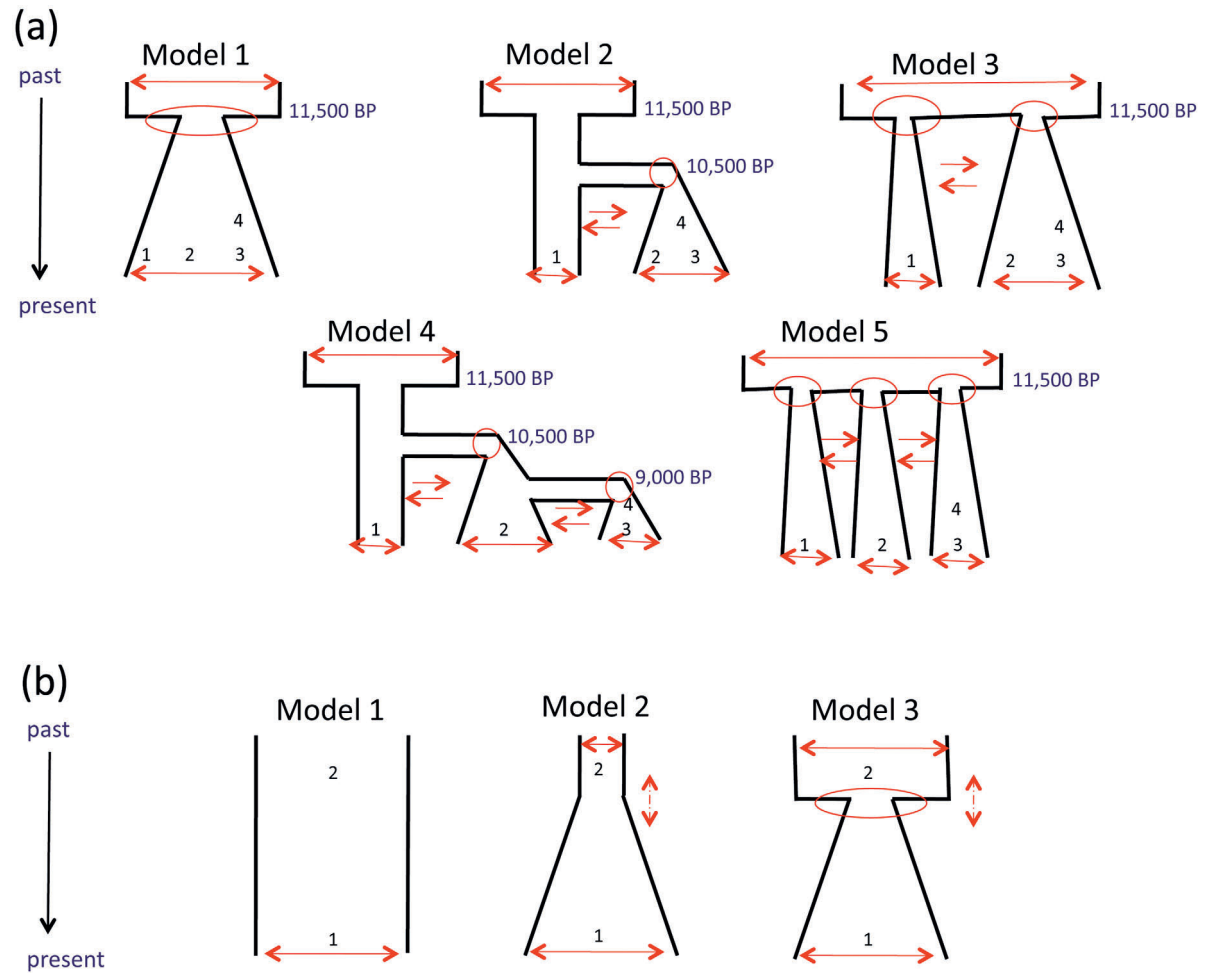


Fig. 1 | Schematic representation of the demography simulated under the coalescent to infer goat (a) and horse (b) domestication processes using mtDNA diversity. Note that time goes backward (arrows on the left) as these models have been simulated under the coalescent. Parallel lines means the population size remains constant through time, while non-parallel lines are used for expanding populations. Arrows and circles highlight the unknown parameters for which ranges of possible values were used. (a) Regarding the five models simulated for the goat samples, in each model three modern populations were sampled: 1: a sample of *C. aegagrus* population, 2: a sample of Near East *C. hircus* population, and 3: a sample of Europe *C. hircus* population. Additionally, sample 4, corresponding to the ancient DNA sample from Southern France, was sampled from the coalescent process at 7,000 years BP. Depending on the model considered, unknown parameters estimated at the end of the simulations are the ancestral population size (parameter range of [100–50,000] in all models), the modern population size(s) (parameter ranges of [10,000–1,000,000] in Model 1, [1,000–10,000] for the wild and [10,000–1,000,000] for the domestic populations in Model 2 to 5), the bottleneck sizes (parameter range of [10–10,000] in all models) that founded the respective populations simulated, and migration rates (parameter range of [0–0.001] in Model 2 to 5) when several populations were considered. Note that for model 1, the modern effective female population size estimated would represent the entire goat population, both wild and domestic; while in model 2 and 3, two modern population sizes are estimated, one for the wild population, and another one estimated for the domestic population of both Near Eastern and European goats. Finally in model 4 and 5, three modern effective population sizes are considered as distinct from one another, with an estimation of the wild effective population size, another of the Near-Eastern domesticate population and the last one of the European domesticate populations. (b) Regarding the three models simulated for the horse samples, two different genetic groups were sampled, indicated by the following numbers: 1: Late Glacial (17,000–11,600 BP); 2: modern horses. Depending on the model, unknown parameters that are estimated at the end of the simulations are the ancestral population size (parameter range of range [100–10,000] in Model 3), the modern population size (parameter range of [1,000–500,000] in all models), the bottleneck size (parameter range of [10–10,000] in Model 2), and the time of domestication (parameter range of [700–1,700] in Model 2 and 3)

Summary statistics

Distinct indices can be calculated from a set of DNA sequence samples to describe various aspects of their genetic diversity. These measures are called “summary statistics” and can be obtained for any sample of DNA sequences. They contain indirect information on the demographic parameters – ancestral population size(s), bottleneck size(s) and population structures – that have shaped the observed data in those samples. The software ARLEQUIN v. 3.1 (Excoffier et al., 2005) was used to calculate a number of summary statistics describing the genetic diversity within and among the samples used. Five summary statistics (number of haplotypes, number of segregating sites, haplotype diversity, Tajima’s D and average number of pairwise differences) were calculated for each sample individually, and four summary statistics (number of private haplotypes to one sample, number of private haplotypes to the other sample, average number of pairwise differences and pairwise F_{st}) were calculated for pairwise comparisons of samples (e.g. between ancient and modern samples, or between modern samples from different regions). In total, 44 summary statistics were used to describe and compare the genetic diversity within and between the four goat populations samples (Fernandez et al., 2006; Naderi et al., 2008; Naderi et al., 2007). Likewise, a total of 14 summary statistics were calculated to describe the relationships within and between the two horse populations samples.

Coalescent simulations

The coalescent approach allows both the evaluation of competing demographic models and the estimation of the model parameters. First, genealogies are simulated for a given demographic model, with its associated parameter values. Once the genealogies have been created, simulated genetic data are generated through the genealogy. Finally, summary statistics describing the genetic diversity of the simulated data under that model are calculated. These simulations were performed with the software Bayesian Serial Simcoal (Anderson et al., 2005), as it allows the incorporation of both modern and ancient DNA samples. The software is based on a two-step modelling process. If n_1 and n_2 are the sample sizes for the modern and ancient sample, respectively, and t_1 is the age (in generations) of the ancient sample, coalescence of n_1 sequences is modelled backwards in time from the present, with n_2 sequences added to the genealogy after t_1 generations.

Genetic data are generated for each simulated genealogy, and those data are used to calculate summary statistics (i.e., the simulated summary statistics). The model(s) that generate the statistics the ‘closest’ to those observed (i.e., the ‘best-fitting’) are retained. For this purpose, a group of techniques called approximate Bayesian computation (ABC) have been widely used in population genetic studies (Beaumont et al., 2002; Fagundes et al., 2007; Nielsen and Beaumont, 2009; Wegmann et al., 2009). These techniques require a very large number of simulations to be performed. The basic idea underlying the use of ABC is that the probability of obtaining the observed summary statistics is proportional to the number of simulated data sets producing summary statistics that are a small distance from the observed summary statistics. Euclidean distances between the observed and simulated summary statistics allow us to determine which of the simulated model(s) best explain the observed patterns of genetic diversity. These Euclidean distances are standardized by common mean and standard deviation of the simulated summary statistics, calculated over all the simulations, regardless the model they belong to (Ray et al., 2009). Note that there are some issues about how small these Euclidean distances should be

and about the number of statistics to be used. These issues are discussed in detail elsewhere (Joyce and Marjoram, 2008; Wegmann et al., 2009). Once the best fitting models have been determined, inferences of some unknown parameters drawn from prior distributions (e.g. ancestral effective-population sizes, modern effective-population sizes, bottleneck or founder event sizes, or time of events) can be made.

Models of goat domestication

Five demographic models were examined to assess how the domestication process has shaped goat genetic diversity (Fig. 1a). The first model considers that all goat samples (the wild, and both modern domestic samples) come from a single population that underwent exponential growth following a bottleneck that occurred some 11,500 years ago. This bottleneck corresponds to a founder event, when animal management in the wild began to affect genetic diversity of the ancestral goat population. This is a simplistic way of seeing the domestication process, but our aim is to examine the simplest possible scenarios first, before adding further complexity. Our reasoning here is that if patterns genetic variation can be explained better by simple models than more complex ones, then – in the absence of clear evidence for specific components of additional complexity – invoking more complex models cannot be justified. This model tests the hypothesis that contemporaneous domestic and wild goat populations behave as a single, random-mating metapopulation. It is equivalent to a neutral model of population expansion without any population structuring. As we cannot be certain what the real ancestral effective population size may have been, we parameterized this component of the model so that it can take any value between 100 and 50,000 effective females. We also parameterized the bottleneck size, allowing it to take any value between 1,000 and 10,000 effective females, while the modern effective female population size could take any value between 10,000 and 1,000,000. We used the same ranges for both ancestral effective-population and bottleneck sizes in subsequent models.

The second model considers that both modern wild and modern domestic populations were derived from an initially ‘managed’ wild population that originated from an ancestral wild population via a bottleneck 11,500 years ago. This population was assumed to stay constant in size to give rise to modern wild goats. Some 1,000 years later, a founder event from this ‘managed’ wild population occurred, forming what would become the contemporaneous *C. hircus* population. The modern wild and domesticate effective population sizes take, independently, any values between 1,000 and 10,000, and between 10,000 and 1,000,000, respectively. These ranges were also used for the third model, which assumes that *C. hircus* samples both from Europe and the Middle East on the one hand, and the *C. aegagrus* sample, on the other hand, are drawn from a single ancestral population via two distinct founder events some 11,500 years ago. Both populations undergo exponential growth following those founding events. Both Models 2 and 3 assume that domestic populations, whether they come from Europe or the Middle East, represent a single large metapopulation without significant structuring. Note that Models 2 and 3 both allow migrations between domestic and wild goat populations, and migration rates can take any value between 0 and 0.001 females per generation.

The fourth model is similar to the second model except that Middle Eastern and European domestic goats were modelled as separate populations, with the latter assumed to be derived from the former. In this model, as in Model 2, the ancestral wild population underwent a founder event some 11,500 years ago and stayed constant thereafter. The Middle Eastern *C. hircus* population was then formed by a

founder event 10,500 years ago from the wild population which remained constant in size, and 1,500 years later the European *C. hircus* population was formed from the Middle Eastern domesticated population by a third founding event. In the fifth model, each of the three modern populations – European domestic, Middle Eastern domestic, and wild – were assumed to have been formed from the ancestral wild population by distinct founder events, each of which occurred 11,500 generations ago. All three populations were allowed to undergo exponential growth from that time on. In both Models 4 and 5, each modern domesticated and wild effective population size could take any value in the intervals [10,000–1,000,000] and [1,000–10,000], respectively. In both models, migrations between European *C. hircus* and *C. aegagrus* are prohibited. However, bidirectional migration is permitted (in the range [0–0.001]) between European and Middle Eastern *C. hircus*, and between Middle Eastern *C. hircus* and *C. aegagrus*. Both models address the role of continental-scale geographic structuring of goat populations, and whether more population structuring is required to account for the differences between the two domestic samples and the wild sample.

Models of horse domestication

As outlined above, the findings of several studies on horse mtDNA have been interpreted as indicating multiple domestications. However, patterns of genetic diversity in modern horses do not indicate any strong phylogeographic structuring (Jansen et al., 2002; Lira et al., 2010; Villà et al., 2001). This could be explained by extensive gene flow between regions – presumably due to breeding practices and intensive trade – during the centuries since domestication. Nonetheless, we should first consider simple models of horse domestication; one of a single domestication event that founded a single and effectively panmictic worldwide population of domestic horses. We therefore simulated three different demographic models in order to test this single domestication hypothesis (Fig. 1b). The explored ranges for population sizes and domestication times were kept wider than might be considered realistic, in order to accommodate extreme parameter values.

Model 1 can be considered as our “null hypothesis”. We simulated a single population constant in size over time. Effective-population sizes between 1,000 and 500,000 individuals were considered. While such a model might be considered unrealistic, it is important to include it to see if the genetic data available contain sufficient information to justify favouring more complex and realistic models.

Model 2 simulates a wild population whose size remained constant until domestication occurred and then underwent a period of exponential growth lasting until the modern day. The interval of values for the ancestral wild population size was between 10 and 10,000 effective females. We set the time of domestication within a range between 4,200 and 10,200 years BP. The modern day effective female population size could assume any value between 1,000 and 500,000 females, as in Model 1.

In Model 3, the ancestral wild population remains constant, at a size taking any value between 100 to 10,000 individuals, until domestication occurs, within the same time range used in Model 2. At that time, a bottleneck occurs – corresponding to the founder event – of between 10 and 10,000 effective domesticated females, from which the domesticated population grows exponentially until modern times. The range of possible values of modern effective female size is the same as in Models 1 and 2 (between 1,000 and 500,000 females).

Model comparisons

In order to compare the models tested, we used an extension of the Bayes Factor (Kass and Raftery, 1995) for more than two models. We performed 500,000 simulations under each model, providing a total of 2,500,000 and 1,500,000 simulations for goats and horses, respectively. For each species, we standardized the Euclidean distances (d) with common mean ($mean_i$) and standard deviation (std_i) over the total number of simulations using all the summary statistics as follow:

$$d = \sqrt{\sum_i \left(\frac{stat_i - mean_i}{std_i} - \frac{obs_i - mean_i}{std_i} \right)^2},$$

with i varying from 1 to 44, or 1 to 14 summary statistics for goats or horses, respectively. We then ranked those standardized Euclidean distances from the smaller ones (i.e., the ones where the summary statistics of the simulated data were most similar to those of the observed data). Finally, we obtained a ratio of the models that gave the smallest Euclidean distances over all the 2,500,000 and 1,500,000 simulations performed for goats and horses, respectively. This approach provides a relative marginal posterior probability of each model over all those tested. It enables comparisons of different models because it eliminates dependencies on the number of parameters used in each model while integrating overall parameter values of these models. It also includes an automatic penalty for model complexity as models with more parameters have larger parameter spaces to search and so require more simulations in order to identify those generating a good fit between the simulated and the observed statistics. Therefore, for a given number of simulations, additional parameters that fail to improve the fit to the data would produce proportionally fewer simulations that are closer to the empirical data, and the corresponding models receive less marginal support. Figure 2 presents the results of this approximate Bayes Factor analysis for the best 10,000 simulations for both goats (Fig. 2a) and horses (Fig. 2b).

Results

Goats

The simple model (Model 1), under which samples come from a single large expanding population, yielded a poorer overall fit to the data than did any of the other models except Model 5. This implies that demographic structure is required to explain the observed patterns of genetic diversity. The highest posterior probabilities were obtained for Models 2 and 4, which have 5 and 8 parameters, respectively. Both models assumed that the size of the *C. aegagrus* population remained constant size starting at 11,500 years BP, and that *C. hircus* population(s) come from an already ‘managed’ wild population of constant size. From an archaeological and geographic point of view, Model 4 makes the most sense, and despite being relatively complex, receives the best support.

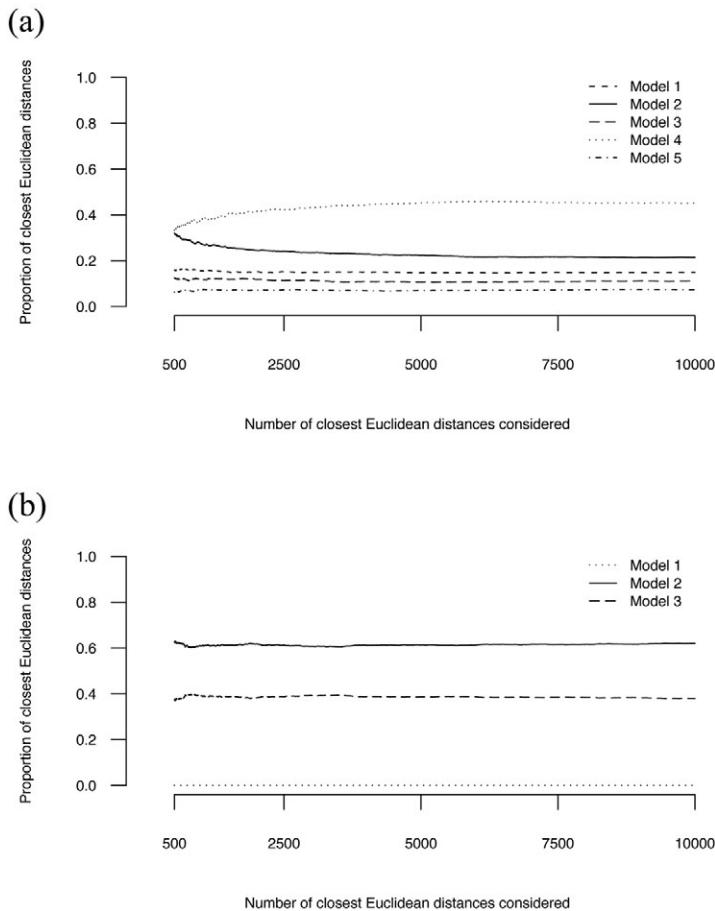


Fig. 2 | Representation of the posterior probabilities of each model tested for goat (a) and horse (b), in function of the number of the smaller Euclidean distances considered. This Bayes factor analogous analysis is based on 500,000 simulations of each model, with a comparison based on the ranking of the closer simulated summary statistics to the observed ones. These representations show the result of this analysis based on the 10,000 smaller Euclidean distances for both goat (a) and horse (b), but note that as there are 5 and 3 models for goat and horse, the total number of simulations taken into account is not the same, i.e. 2,500,000 and 1,500,000 simulations, respectively

Horses

Among the three models proposed for horse domestication, the one that best fit the data is Model 2, which assumes a constant ancestral population size that starts to expand at the time of domestication, without any bottleneck. This result supports the hypothesis that most lineages from the ancestral wild population were incorporated into the domestic one.

Discussion

The Neolithic transition is a cultural process defining the move from a semi-nomadic life-style with an economy based on hunting and gathering to a sedentary culture, in which agriculture and animal exploitation become the dominant subsistence strategies. It is likely to have been a long and complex process involving the acquisition of social behaviours associated with sedentary settlements, the development of new economic strategies (such as animal and plant domestication) and the retention of new skills and technical innovations (e.g. pottery, polished stone tools). This transition entailed diverse changes that shaped cultures (structured societies), environments (farming-like landscapes) and genetic diversity (of both domesticated species and human populations). As such past evolutionary processes

shape the gene pool of modern populations, genetic data can be used to make inferences on those processes of domestication during the Neolithic transition.

Analyses of genetic diversity have attempted to address the relative role of founder events, geographic structure, local admixture with wild progenitors, and migrations, on their genetic diversity. In this context, phylogeographic studies have attempted to identify founding female (using mtDNA) or male (using Y chromosomes) lineages. However, the arrival of a population into a geographical region does not necessarily equate to the age of particular genealogical lineages observed in that region (Barbujani et al., 1998; Barbujani and Goldstein, 2004).

As exemplified above in the case of goat and horse domestication, the coalescent simulation approach enables different evolutionary models to be evaluated for consistency with observed data. The results of this analysis for the goat samples show that serial founder effects from previously managed populations (Model 2 and, in particular, Model 4) appear more likely than distinct founder events for each population (Models 3 and 5). Furthermore, as Model 1 receives little support, it seems that a certain level of population structure between wild and domestic populations is needed to explain the observed mtDNA diversity. Furthermore, as Models 3 and 5 have been shown not to explain the observed summary statistics very well, it seems that any structure is more likely to have arisen after 11,500 BP, probably by way of sequential founding events with subsequent gene flow. In summary, multiple independent domestications do not seem necessary to explain the highly divergent lineages found in today's day goat populations. Further models can be envisaged, such as models that parameterize the times of the founder events to infer when the domesticated goat populations started to diverge genetically from the wild population, and assess how this time correlates with archaeological estimates.

In comparison, the results of our analyses show that of the three models considered, horse mtDNA pattern of diversity are best explained by one of an expanding domesticated population issued from a smaller single wild ancestral population. This seems in accord with previous studies that suggested either multiple domestications, or strong introgression of wild mares in the domestic population (Jansen et al., 2002; Vila et al., 2001). However, we would need to explicitly test further scenarios either by including more ancient samples from different regions (Europe, Central Asia) or by considering more than one ancient wild population as ancestral to modern domestic horses. Indeed, it would be useful to simulate two horse populations ancestral to two domestic horse populations, (e.g. one in Central Asia and one in Europe, with or without gene-flow between them). Horse domestication could also be seen as a bottleneck (or founder event) from both populations and either a subsequent strong gene flow between the two groups, or the convergence in a unique modern population, could occur, as discussed above. However, the fact that no strong phylogeographic structure is found in modern horses suggests that geographically structured models are only likely to receive support with the addition of considerably more ancient DNA data than is considered here.

The best model among a set tested does not necessarily equate to the one reflecting what actually occurred. No model that we could test could capture the true complexity of a population's history. This is exemplified when studying structured populations (Beaumont, 1999; Nielsen and Beaumont, 2009; Wakeley, 1999; Wilson et al., 2003). Furthermore, we must emphasize that it is possible for multiple models to be consistent with both archaeological and genetic data; thus efficient statistical tools must be used to compare alternative evolutionary models. In that context, Approximate Bayesian Computation (ABC) provides a very promising means of comparing models of complex demographic histories of diverse populations (Beaumont et al., 2002; Ray et al., 2009). These methods make use of simplified models as a baseline to compare more complex models against. The reason why an ABC approach was

not used here to estimate unknown parameters is that the number of simulations performed in this study was not sufficient to provide reliable estimates. More simulations need to be run in order to be more conclusive. But the results presented here do provide us with information that can be used to refine models and to sharpen intuitions on goat and horse domestication processes. The coalescent approach is not an end in itself (Currat, this volume; Currat and Excoffier, 2005; Excoffier et al., 2009; Francois et al., 2010; Hofer et al., 2009), but it can provide an initial understanding of the ancestral relationship between contemporaneous and ancient populations (Bramanti et al., 2009; Malmström et al., 2009).

References

- Anderson, C.N., Ramakrishnan, U., Chan, Y.L., Hadly, E.A., 2005. Serial SimCoal: A population genetics model for data from multiple populations and points in time. *Bioinformatics* 21, 1733–1734.
- Anthony, D.W., Brown, D.R., 2003. Eneolithic horse rituals and riding in the steppes: new evidence, in: Levine, M., Renfrew, C., Boyle, K. (Eds.), *Prehistoric Steppe Adaptation and the Horse*. McDonald Institute for Archaeological Research, Cambridge, pp. 55–68.
- Barbujani, G., Bertorelle, G., Chikhi, L., 1998. Evidence for Paleolithic and Neolithic gene flow in Europe. *American Journal of Human Genetics* 62, 488–492.
- Barbujani, G., Goldstein, D.B., 2004. Africans and Asians abroad: Genetic diversity in Europe. *Annual Review of Genomics and Human Genetics* 5, 119–150.
- Beaumont, M.A., 1999. Detecting population expansion and decline using microsatellites. *Genetics* 153, 2013–2029.
- Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian computation in population genetics. *Genetics* 162, 2025–2035.
- Belle, E.M., Landry, P.A., Barbujani, G., 2006. Origins and evolution of the Europeans' genome: evidence from multiple microsatellite loci. *Proceedings of the Royal Society London B: Biological Sciences* 273, 1595–1602.
- Bramanti, B., Thomas, M.G., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M.N., Jankauskas, R., Kind, C.J., Lueth, F., Terberger, T., Hiller, J., Matsumura, S., Forster, P., Burger, J., 2009. Genetic discontinuity between local hunter-gatherers and Central Europe's first farmers, *Science* 326, 138–140.
- Burger, J., Thomas, M.G., 2011. The palaeopopulation-genetics of humans, cattle and dairying in Neolithic Europe, in: Pinhasi, R., Stock, J. (Eds.), *The Bioarchaeology of the Transition to Agriculture*. Wiley Blackwell, Chichester, UK, pp. 371–384.
- Currat, M., Excoffier, L., 2005. The effect of the Neolithic expansion on European molecular diversity. *Proceedings of the Royal Society London B: Biological Sciences* 272, 679–688.
- Excoffier, L., Foll, M., Petit, R.J., 2009. Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics* 40, 481–501.
- Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary bioinformatics Online* 1, 47–50.
- Fagundes, N.J., Ray, N., Beaumont, M., Neuen-schwander, S., Salzano, F.M., Bonatto, S.L., Excoffier, L., 2007. Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Science USA* 104, 17614–17619.
- Fernandez, H., Hughes, S., Vigne, J.D., Helmer, D., Hodgins, G., Miquel, C., Hanni, C., Luikart, G., Taberlet, P., 2006. Divergent mtDNA lineages of goats in an Early Neolithic site, far from the initial domestication areas. *Proceedings of the National Academy of Science USA* 103, 15375–15379.
- Francois, O., Currat, M., Ray, N., Han, E., Excoffier, L., Novembre, J., 2010. Principal component analysis under population genetic models of range expansion and admixture. *Molecular Biology and Evolution* 27, 1257–1268.
- Goldstein, D.B., Chikhi, L., 2002. Human migrations and population structure: what we know and why it matters. *Annual Review of Genomics and Human Genetics* 3, 129–152.
- Hofer, T., Ray, N., Wegmann, D., Excoffier, L., 2009. Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. *Annals of Human Genetics* 73, 95–108.
- Jansen, T., Forster, P., Levine, M.A., Oelke, H., Hurler, M., Renfrew, C., Weber, J., Olek, K., 2002. Mitochondrial DNA and the origins of the domestic horse. *Proceedings of the National Academy of Science USA* 99, 10905–10910.

- Joyce, P., Marjoram, P., 2008. Approximately sufficient statistics and Bayesian computation. *Statistical Applications in Genetics and Molecular Biology* 1, Article 26.
- Kass, R.E., Raftery, A.E., 1995. Bayes Factor. *Journal of the American Statistical Association* 90, 773–795.
- Lira, J., Linderholm, A., Olaria, C., Brandstrom Durling, M., Gilbert, M.T., Ellegren, H., Willerslev, E., Liden, K., Arsuaga, J.L., Gotherstrom, A., 2010. Ancient DNA reveals traces of Iberian Neolithic and Bronze Age lineages in modern Iberian horses. *Molecular Ecology* 19, 64–78.
- Luikart, G., Gielly, L., Excoffier, L., Vigne, J.D., Bouvet, J., Taberlet, P., 2001. Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proceedings of the National Academy of Science USA* 98, 5927–5932.
- Malmström, H., Gilbert, M.T., Thomas, M.G., Brandstrom, M., Stora, J., Molnar, P., Andersen, P.K., Bendixen, C., Holmlund, G., Gotherstrom, A., Willerslev, E., 2009. Ancient DNA reveals lack of continuity between neolithic hunter-gatherers and contemporary Scandinavians. *Current Biology* 19, 1758–1762.
- Naderi, S., Rezaei, H.R., Pompanon, F., Blum, M.G., Negrini, R., Naghash, H.R., Balkiz, O., Mashkour, M., Gaggiotti, O.E., Ajmone-Marsan, P., Kence, A., Vigne, J.D., Taberlet, P., 2008. The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proceedings of the National Academy of Science USA* 105, 17659–17664.
- Naderi, S., Rezaei, H.R., Taberlet, P., Zundel, S., Rafat, S.A., Naghash, H.R., el-Barody, M.A., Ertugrul, O., Pompanon, F., 2007. Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity. *PLoS One* 2, e1012.
- Nielsen, R., Beaumont, M.A., 2009. Statistical inferences in phylogeography. *Molecular Ecology* 18, 1034–1047.
- Outram, A.K., Stear, N.A., Bendrey, R., Olsen, S., Kasparov, A., Zaibert, V., Thorpe, N., Evershed, R.P., 2009. The earliest horse harnessing and milking. *Science* 323, 1332–1335.
- Ray, N., Wegmann, D., Fagundes, N.J., Wang, S., Ruiz-Linares, A., Excoffier, L., 2009. A statistical evaluation of models for the initial settlement of the American continent emphasizes the importance of gene flow with Asia. *Molecular Biology and Evolution* 27, 337–345.
- Rosenberg, N.A., Nordborg, M., 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics* 3, 380–390.
- Vila, C., Leonard, J.A., Gotherstrom, A., Marklund, S., Sandberg, K., Liden, K., Wayne, R.K., Ellegren, H., 2001. Widespread origins of domestic horse lineages. *Science* 291, 474–477.
- Wakeley, J., 1999. Nonequilibrium migration in human history. *Genetics* 153, 1863–1871.
- Weber, C., 2005. Molecular genotyping of a Skythian horse population. Master thesis, Institute of Anthropology, Johannes Gutenberg University, Mainz.
- Wegmann, D., Leuenberger, C., Excoffier, L., 2009. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics* 182, 1207–1218.
- Wilson, I.J., Weale, M.E., Balding, D.J., 2003. Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 166, 155–201.
- Zeder, M.A., 2008. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the National Academy of Science USA* 105, 11597–11604.

