



Archaic DNA or the words of the past: an example with horse domestication

Madeleine Murray, Université d'Aix-Marseille, December 2019

Abstract

Since its domestication, the horse has played an important role in human societies. On the other hand, humans have reshaped the horse's genome by acting on selection, migration and introgression. This review presents the main observations on the domestication of horses together with the methods applied. Ancient DNA has become the driving force of these recent findings as it can be summoned to build phylogenetic trees and to unravel past population dynamics and ancestry. A lot has been achieved in the last two decades, yet ancient DNA has not reached its full potential.

Introduction

Horses were first utilised for their meat and milk, but quickly revolutionised transportation and warfare. Their speed and stamina eased human travels, promoted cultural exchange (religion, science and art), spread languages, but also carried diseases and our genes (Librado *et al.*, 2016; Orlando, in publication). After the introduction of the horseshoes and the horse collar, the horse became widely used in the agricultural sector, increasing crop yield in mediaeval Europe. Although Europe now relies on machinery, the horse remains essential in food production in least-developed countries (Librado *et al.*, 2016). For better or worse, the horse has also revolutionised our ways to make war. The impact of horses on warfare was such that key historical figures such as Alexander the Great, Genghis Kahn or Napoleon cannot be depicted without a horse (Orlando, in publication).

As the horse transformed human societies, humans also reshaped the horse's genome through selection and introgression (Librado *et al.*, 2017). The three milestones in horse genomics are the assembly of its reference

genome in 2009, the development and constant improvement of DNA-hybridisation microarrays since 2012, and the DNA sequencing of ancient fossils from extinct horse species (Librado *et al.*, 2016).

Selection modelled by humans

Similarly to dogs or agricultural crops, humans have favoured particular traits during the domestication of horses (Librado *et al.*, 2016). Genes and alleles associated with increased speed such as *DMRT3*, *MSTN*, *PDK4* and *ACN9* rose in frequency in multiple horse species in the last one thousand years. This supports that speed capacity was positively selected for in the last millennium (Fages *et al.*, 2019).

Fifteen Scythian horses from the Iron Age were sequenced by Librado and colleagues in 2017. In addition to alleles associated with speed performances, increased robustness was identified with the overexpression of genes in link with the carpal bones, tibia, clavicle and radius bone. However, racing performances were not the only traits selected for by humans. Breast and mammary glands as well as the posterior

pituitary also appeared functionally overrepresented increasing contraction, lactation and water retention in those Scythian horses. This study suggests that in addition to speed, Scythians also selected horses for milking and adapted to the dry steppe region of central Asia (Librado *et al.*, 2017).

More and more studies show that coat colour was also selected, and the leopard-spotting complex is one such example with heterozygous being white with black spots. Homozygous are mostly white but present night blindness and poor vision. Negative selection is therefore expected for this allele in the wild, so the high frequency found in domestic horses are likely due to human influence (Integrated Case Study II: Horse Domestication and the Origins of Pastoralism in Central Asia, 2019).

Sequencing of ancient fossils

In 2003, a preserved metapodial horse fossil dating from the Middle Pleistocene was found in an ice wedge in Canada. Its DNA was extracted and sequenced by Orlando and colleagues in 2013 and sequencing quality was controlled with the following tests:

- sequencing results from two different labs were consistent, ruling out post-purification contamination
- autosomal, Y-chromosome and mitochondrial DNA confirmed the fossil carbon dating and specie
- cytosine deamination at overhangs was strikingly higher compared to younger fossils supporting its older age (cytosine deamination is the conversion of cytosine to uracil overtime at single stranded overhangs where DNA is more exposed to degradation).

This paper hence established the world's oldest full genome sequence to about 560-780 thousand years old (Orlando *et al.*, 2013).

Fossil sequencing is a powerful tool in archaeology as it can assign a specie to bone remains when the full skeleton cannot be reconstructed or when morphology alone is not enough. Indeed, contrary to pigs or dogs, skeletons from the *Equus* genus are not very variable. For example, bone remains from mules and horses cannot be distinguished but in 2019 Fages and colleagues were able to assign 27 archaic specimens to mules with the help of ancient DNA. Mules are the sterile offspring of a

horse mare and a donkey jack. These findings prove that mules have been bred for at least 2,200 years despite their inability to produce offspring and the obvious cost it implies. Mules may have sometimes been preferred over horses and they are more sure-footed, more resistant to diseases and harder working (Fages *et al.*, 2019).

However, one should be cautious when extrapolating information from ancient DNA. Additional power is gained but past population size and bottleneck events should always be considered. For example, sampling a population pre-dating and post-dating a bottleneck would considerably influence genetic markers, especially for a small population effective size (Leonardi *et al.*, 2016).

Sex-biased gene flow

Non-recombinant Y-chromosome (NRY) and mitochondrial DNA (mtDNA) studies suggest a tremendous difference in population effective size between stallions and mares. Indeed, NRY regions show poor variation suggesting a low effective population size in stallions, whilst the horse is the domesticated animal with the most diverse mtDNA indicating a high effective population size in mares (Librado *et al.*, 2016). This may be explained by the selection of a few males in domesticated populations and the introgression of females from wild populations, or by the horse's natural polygamous behaviour, or a mixture of both these hypotheses as they are not mutually exclusive (Lippold *et al.*, 2011). However, the difference in population size between stallions and mares is estimated to 2,000 years ago when the overall nucleotide diversity of stallions was halved. Interestingly this corresponds to modern practices with breeders selecting only a few stallions for mating. It is thought that horse breeders in the Byzantine and the Great Mongolian Empires followed similar practices (Orlando, in publication). The reasoning behind this strategy is the more docile nature of mares compared to stallions, and their ability to produce milk (Integrated Case Study II: Horse Domestication and the Origins of Pastoralism in Central Asia, 2019). In opposition, breeders in the Roman Empire did not select particular stallions over others, resulting in a homogenous reproductive rate across them (Orlando, in publication).

The selective sweep on the Y-chromosome caused by human breeding practices may be the origin of the low genetic diversity observed in present domesticated horses. Indeed, substantial amounts of genetic diversity was found on the Y-chromosome of archaic horses (Lippold *et al.*, 2011).

Mutational loads can be described as the genetic burden of a population and is expected to have increased in horses due to the reduction in breeding stallion effective size. Mutational loads were calculated from differential diversity patterns at non-synonymous and synonymous sites, and from sites classified as deleterious and non-deleterious, and were found to correlate with reduced selection. This analysis further supports that implemented breeding strategies led to the accumulation of deleterious variants and reduced the efficacy of purifying selection (also called negative selection). The horse is a good illustration of the genomic cost inflicted by modern breeding (Fages *et al.*, 2019).

NRV and mtDNA studies are therefore incredibly useful, but unfortunately a number of papers often focus on NRV and mtDNA alone without taking into account autosomal DNA (Leonardi *et al.*, 2016). Furthermore, mtDNA studies are usually preferred over NRV as a large number of mitochondria are present per cell. Moreover, degradation time is slower for mtDNA because its circular form protects its DNA overhangs. In addition, the Y-chromosome constitutes a high number of repeat elements which complicates analysis. What is more, Y-chromosomes can only be found in male individuals whereas mtDNA is present in all specimens (Lippold *et al.*, 2011).

Phylogenetic trees

As mentioned above, NRV and mtDNA can be used to estimate and compare stallion to mare effective population size, but also to elucidate genetic relationships intra and inter species. Indeed, principal component analysis (PCA) from modern and archaic horses revealed that modern horses do not form a monophyletic group suggesting that they are descendant of two different archaic groups. Both extant horse lineages, Domestic and Przewalski, were expected to descend from the archaic Botai horse, but it now appears to only be the case of the latter. Moreover, Botai horses were confirmed as

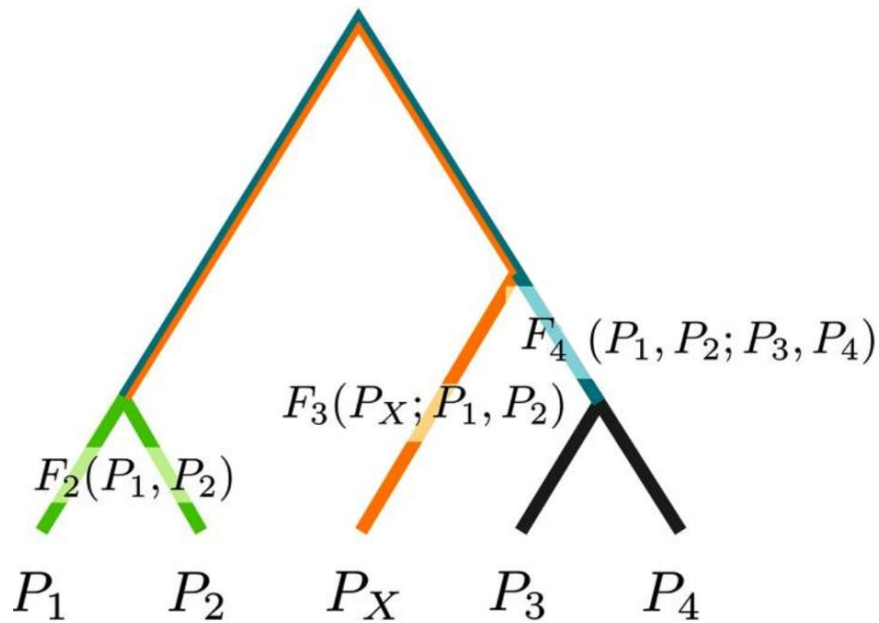
the direct ancestors of Borly4 horses. PCA results in general were confirmed by f3-statistics (Gaunitz *et al.*, 2018). F-statistics (f2-, f3- and f4-statistics) measure shared genetic drift between populations to test for admixture and ancestry (Leonardi *et al.*, 2016). F-statistics can be interpreted as branch length as detailed below on Figure 1. In summary, f2-statistics measures genetic drift between two populations, whilst f3-statistics and f4-statistics test whether the tested population P_x is admixed with P_1 or P_2 . Obviously f4-statistics is more precise than f3-statistics as it incorporates an additional reference population but requires broader knowledge of past population dynamics (Peter, 2016).

Gaunitz and colleagues demonstrated that all Domestic horses have arisen from an unknown group of horses (Figure 2). Although Botai horses are not the ancestors of Domestic horses, they have contributed to a portion of their genome as identified by f4-statistics. This may be explained by massive introgression events leading to the near complete replacement of Botai horse ancestry by ancient Domestic horse ancestry (Gaunitz *et al.*, 2018). Altogether, the integration of ancient DNA into phylogenetic trees often demonstrates unsuspected population dynamics and helps to better understand past migrations, expansions and declines (Leonardi *et al.*, 2016).

In addition to the two extant horse lineages of Domestic and Przewalski's horses, an extinct lineage was hypothesised based on archaic DNA after the discovery of three bones in Siberia which showed great morphological differences with extant species, but especially showed very distinctive mtDNA haplogroups. Another horse specimen was later found with mtDNA so divergent that it brought up the past existence of a fourth ghost lineage in Iberia (Fages *et al.*, 2019).

Past population migration

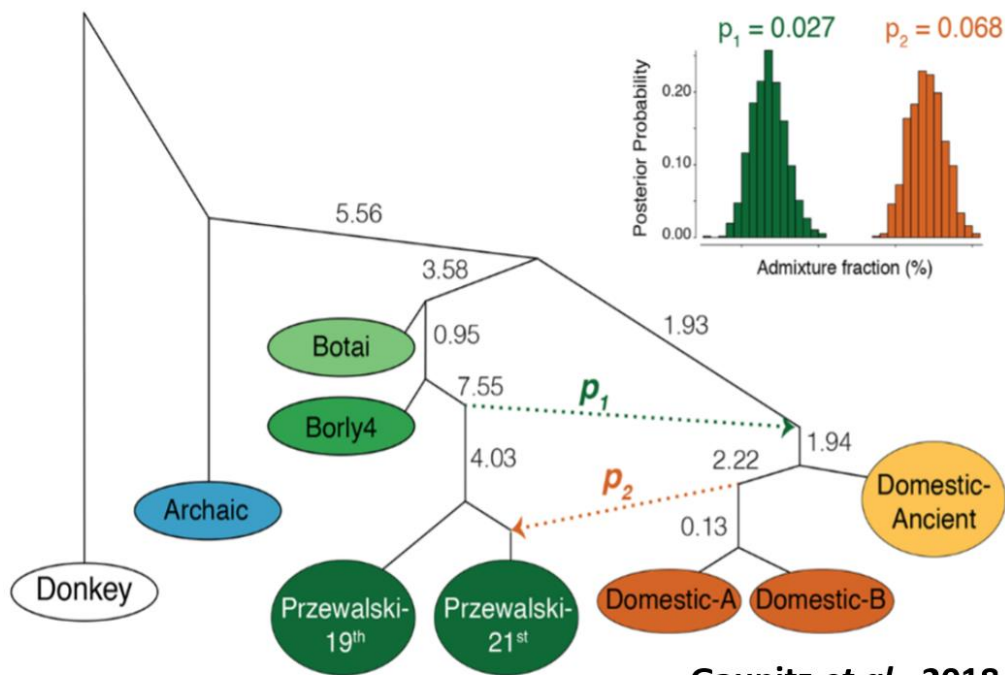
Even though Botai horses may not be the direct ancestor of domesticated horses, it is still widely accepted that domestication originated on the Eurasian steppes about 6,000 years ago. The range of domesticated horses broadened to the rest of the continent and its genome was modified through introgression events. Although Gaunitz and colleagues did not find introgression to be dependent on sex, Yang and colleagues uncovered past horse migration patterns through



Peter, 2016

Figure 1: Simple representation of f2, f3 and f4 statistics

Five populations are displayed on this phylogenetic tree: P_1 , P_2 , P_3 , P_4 and P_X (P_X often stands for the tested admixture population). F_2 corresponds to the path from P_1 to P_2 (in green). F_3 corresponds to the path from P_X to the most recent common ancestor of P_1 and P_2 (in orange). F_4 corresponds to the branch length from the common ancestor of P_1 and P_2 to the common ancestor of P_3 and P_4 (in blue). Side note: P_1 and P_2 are interchangeable in F_3 , as well as P_1 with P_2 , and P_3 with P_4 in F_4 .



Gaunitz *et al.*, 2018

Figure 2: Phylogenetic tree of the horse

Posterior distributions of admixture proportions along p_1 and p_2 branches are shown in the top right corner

the introgression of maternal lineages from wild herds (Gaunitz *et al.*, 2018; Yang *et al.*, 2018).

About seven major maternal lineages are present in modern horses, ranging from A to G. Sequencing modern horses around Asia demonstrated that sub-lineage F3 for example gradually decreased in frequency with distance to Inner Mongolia, suggesting that the sub-lineage likely originated from there. This study supports that horse domestication originated from Central Asia as it is where the highest genetic diversity can be found. Interestingly, unexpected levels of genetic diversity can be found on the ancient Tea-Horse Road, translating the exchange of tea and other goods including horses (Yang *et al.*, 2018).

Although domestication emerged in central Asia, the equid family originated in North America about 55 million years ago. It later expanded to the rest of the world and diversified into several genera. Only the *Equus* genus remains today comprising of horses, donkeys, asses and zebras. The mass extinction of the megafauna in the Americas about 12,000 years ago is yet to be answered, and the extinction of most of the equid families remains unexplained (Librado *et al.*, 2016).

Conclusion

Fossils exhibit a considerable diversity of the equid family, especially at its birthplace in the Americas, but most of its diversity is extinct today. Only the *Equus* genus remains extant, with two lineages for horses for example. However, ancient DNA reveals the existence of additional lineages in the past and higher diversity in archaic horses.

Domestication may be the cause of the reduced genetic diversity in horses as humans maintained low effective population size in stallions leading to the accumulation of deleterious alleles, especially on the Y-chromosome. Humans have also acted on selection preferring alleles associated with speed and influencing coat colour, sometimes leading to genetic disabilities. In addition, the horse's past migrations are also intertwined with historical human trade routes.

In conclusion, ancient DNA has increased statistical power of genetic studies as it allows the inclusion of samples from varied timepoints. However, scientists should remain wary of sampling effects and are advised to always estimate past population effective size as well as

major events such as bottlenecks before interpretation.

In addition, constructing phylogenetic trees to picture past events usually presents unsuspected complications. With or without ancient DNA, the incorporation of autosomal studies together with mtDNA and NRY studies may shed more light on admixture or introgression for example.

References

- Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., McCrory Constantz, C., Gamba, C., Al-Rasheid, K., Albizuri, S., Alfarhan, A., Allentoft, M., Alquraishi, S., Anthony, D., Baimukhanov, N., Barrett, J., Bayarsaikhan, J., Benecke, N., Bernáldez-Sánchez, E., Berrocal-Rangel, L., Biglari, F., Boessenkool, S., Boldgiv, B., Brem, G., Brown, D., Burger, J., Crubézy, E., Daugnora, L., Davoudi, H., de Barros Damgaard, P., de los Ángeles de Chorro y de Villa-Ceballos, M., Deschler-Erb, S., Detry, C., Dill, N., do Mar Oom, M., Dohr, A., Ellingvåg, S., Erdenebaatar, D., Fathi, H., Felkel, S., Fernández-Rodríguez, C., García-Viñas, E., Germonpré, M., Granado, J., Hallsson, J., Hemmer, H., Hofreiter, M., Kasparov, A., Khasanov, M., Khazaeli, R., Kosintsev, P., Kristiansen, K., Kubatbek, T., Kuderna, L., Kuznetsov, P., Laleh, H., Leonard, J., Lhuillier, J., Liesau von Lettow-Vorbeck, C., Logvin, A., Lõugas, L., Ludwig, A., Luis, C., Arruda, A., Marques-Bonet, T., Matoso Silva, R., Merz, V., Mijiddorj, E., Miller, B., Monchalov, O., Mohaseb, F., Morales, A., Nieto-Espinet, A., Nistelberger, H., Onar, V., Pálsdóttir, A., Pitulko, V., Pitskhelauri, K., Pruvost, M., Rajic Sikanjic, P., Rapan Papeša, A., Roslyakova, N., Sardari, A., Sauer, E., Schafberg, R., Scheu, A., Schibler, J., Schlumbaum, A., Serrand, N., Serres-Armero, A., Shapiro, B., Sheikhi Seno, S., Shevnina, I., Shidrang, S., Southon, J., Star, B., Sykes, N., Taheri, K., Taylor, W., Teegen, W., Trbojević Vukičević, T., Trixl, S., Tumen, D., Undrakhbold, S., Usmanova, E., Vahdati, A., Valenzuela-Lamas, S., Viegas, C., Wallner, B., Weinstock, J., Zaibert, V., Clavel, B., Lepetz, S., Mashkour, M., Helgason, A., Stefánsson, K., Barrey, E., Willerslev, E., Outram, A., Librado, P. and Orlando, L. (2019). Tracking Five Millennia of Horse

- Management with Extensive Ancient Genome Time Series. *Cell*, 177(6), pp.1419-1435.e31.
- Gaunitz, C., Fages, A., Hanghøj, K., Albrechtsen, A., Khan, N., Schubert, M., Seguin-Orlando, A., Owens, I., Felkel, S., Bignon-Lau, O., de Barros Damgaard, P., Mittnik, A., Mohaseb, A., Davoudi, H., Alquraishi, S., Alfarhan, A., Al-Rasheid, K., Crubézy, E., Benecke, N., Olsen, S., Brown, D., Anthony, D., Massy, K., Pitulko, V., Kasparov, A., Brem, G., Hofreiter, M., Mukhtarova, G., Baimukhanov, N., Lõugas, L., Onar, V., Stockhammer, P., Krause, J., Boldgiv, B., Undrakhbold, S., Erdenebaatar, D., Lepetz, S., Mashkour, M., Ludwig, A., Wallner, B., Merz, V., Merz, I., Zaibert, V., Willerslev, E., Librado, P., Outram, A. and Orlando, L. (2018). Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science*, 360(6384), pp.111-114.
- Integrated Case Study II: Horse Domestication and the Origins of Pastoralism in Central Asia. (2019). *Subsistence and Society in Prehistory*, pp.163-194.
- Leonardi, M., Librado, P., Der Sarkissian, C., Schubert, M., Alfarhan, A., Alquraishi, S., Al-Rasheid, K., Gamba, C., Willerslev, E. and Orlando, L. (2016). Evolutionary Patterns and Processes: Lessons from Ancient DNA. *Systematic Biology*, p.syw059.
- Librado, P., Fages, A., Gaunitz, C., Leonardi, M., Wagner, S., Khan, N., Hanghøj, K., Alquraishi, S., Alfarhan, A., Al-Rasheid, K., Der Sarkissian, C., Schubert, M. and Orlando, L. (2016). The Evolutionary Origin and Genetic Makeup of Domestic Horses. *Genetics*, 204(2), pp.423-434.
- Librado, P., Gamba, C., Gaunitz, C., Der Sarkissian, C., Pruvost, M., Albrechtsen, A., Fages, A., Khan, N., Schubert, M., Jagannathan, V., Serres-Armero, A., Kuderna, L., Povolotskaya, I., Seguin-Orlando, A., Lepetz, S., Neuditschko, M., Thèves, C., Alquraishi, S., Alfarhan, A., Al-Rasheid, K., Rieder, S., Samashev, Z., Francfort, H., Benecke, N., Hofreiter, M., Ludwig, A., Keyser, C., Marques-Bonet, T., Ludes, B., Crubézy, E., Leeb, T., Willerslev, E. and Orlando, L. (2017). Ancient genomic changes associated with domestication of the horse. *Science*, 356(6336), pp.442-445.
- Lippold, S., Knapp, M., Kuznetsova, T., Leonard, J., Benecke, N., Ludwig, A., Rasmussen, M., Cooper, A., Weinstock, J., Willerslev, E., Shapiro, B. and Hofreiter, M. (2011). Discovery of lost diversity of paternal horse lineages using ancient DNA. *Nature Communications*, 2(1).
- Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., Johnson, P., Fumagalli, M., Vilstrup, J., Raghavan, M., Korneliussen, T., Malaspinas, A., Vogt, J., Szklarczyk, D., Kelstrup, C., Vinther, J., Dolocan, A., Stenderup, J., Velazquez, A., Cahill, J., Rasmussen, M., Wang, X., Min, J., Zazula, G., Seguin-Orlando, A., Mortensen, C., Magnussen, K., Thompson, J., Weinstock, J., Gregersen, K., Røed, K., Eisenmann, V., Rubin, C., Miller, D., Antczak, D., Bertelsen, M., Brunak, S., Al-Rasheid, K., Ryder, O., Andersson, L., Mundy, J., Krogh, A., Gilbert, M., Kjær, K., Sicheritz-Ponten, T., Jensen, L., Olsen, J., Hofreiter, M., Nielsen, R., Shapiro, B., Wang, J. and Willerslev, E. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, 499(7456), pp.74-78.
- Peter, B. (2016). Admixture, Population Structure, and F-Statistics. *Genetics*, 202(4), pp.1485-1501.
- Yang, L., Kong, X., Yang, S., Dong, X., Yang, J., Gou, X. and Zhang, H. (2018). Haplotype diversity in mitochondrial DNA reveals the multiple origins of Tibetan horse. *PLOS ONE*, 13(7), p.e0201564.