



HAL
open science

The Evolutionary and Historical Foundation of the Modern Horse: Lessons from Ancient Genomics

Ludovic Orlando

► **To cite this version:**

Ludovic Orlando. The Evolutionary and Historical Foundation of the Modern Horse: Lessons from Ancient Genomics. *Annual Review of Genetics*, 2020, 54 (1), pp.563-581. 10.1146/annurev-genet-021920-011805 . hal-03030321

HAL Id: hal-03030321

<https://hal.science/hal-03030321>

Submitted on 30 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Title Page

The evolutionary and historical foundation of the modern horse: lessons from ancient genomics

Ludovic Orlando^{1,2}

¹Laboratoire d'Anthropobiologie Moléculaire et Imagerie de Synthèse, Faculté de Médecine Purpan, Bâtiment A, 37 Allées Jules Guesde, 31000 Toulouse, France

²GLOBE Institute, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen K, Denmark

Email: ludovic.orlando@univ-tlse3.fr

ORCID number: 0000-0003-3936-1850

Correspondence should be sent to Ludovic Orlando, ludovic.orlando@univ-tlse3.fr

Keywords

Horse, Domestication, Genomics, Selection, Conservation, Evolution

Abstract

The domestication of the horse some ~5,500 years ago followed that of dogs, sheep, goats, cattle, and pigs by ~2,500-10,000 years. By providing fast transportation and transforming warfare, the horse had an impact on human history with no equivalent in the animal kingdom. Even though the equine sport industry represents a considerable economic worth today, the evolutionary history underlying the emergence of the modern domestic horse remains contentious. In the last decade, novel sequencing technologies have revolutionized our capacity to sequence the complete genome of organisms, including from archaeological remains. Applied to horses, these technologies have provided unprecedented levels of information and have considerably changed models of horse domestication. This review illustrates how ancient DNA, especially ancient genomes, have contributed to rethink the process by which horses were first domesticated, and then diversified into a variety of breeds showing a whole range of traits that are useful to humans.

1. Introduction

By the end of the 2000s, the sequencing of a reference genome and the development of cost-effective genome-wide genotyping solutions for horses have considerably advanced the genomic mapping of diseases and traits in this species (98). At about the same time, the application of the then-emerging next-generation sequencing technologies truly revolutionized ancient DNA research as the complete genome sequence of ancient individuals, including from extinct species, could now be characterized (see 18, 62 for reviews). With a growing understanding of the genetic basis underlying important traits in horses, such as coat-coloration (e.g. 4, 38, 87), speed (26, 36, 37) and gaitedness (3), the application of DNA sequencing to the molecules that survive in the archaeological bone assemblages held the promise to enhance our understanding of the history of horse breeding, even for characters that rarely fossilize. Predicting the phenotype of ancient horses from the genotypes they carried would certainly help appreciate the individual variation within key past equestrian cultures (58). By providing direct measures of how allelic frequencies changed through time, ancient DNA data could also be expected to provide unprecedented statistical power to catch artificial selection red-handed and identify the (pre-)historical contexts into which past breeders started to propagate specific traits of interest (see 61 for a review of statistical methods to detect selection signatures in time-serial data). Additionally, as ancient DNA analyses can chart population affinities through space and time, they held the potential to unveil the historical processes by which different bloodlines were blended to form the over 600 horse breeds known to the world today (34).

A decade later, many of these promises have been fulfilled and the horse represents, after humans, the species in which the most extensive time-series of ancient genomes has been characterized (Figure 1). The horse is also the organism in which the oldest genome has been sequenced, from a metapodial bone that was frozen in the Yukon permafrost for the last 560,000 to 780,000 years (73). Together with genome-wide data from 117 additional equine specimens, the 166 ancient genomes currently available have revealed that all available models of horse domestication required important revision (76). They also unveiled an unexpectedly large diversity in the past, with many now-extinct lineages still roaming around at the time humans first domesticated the horse, and

provided examples of exceptionally fast changes in the horse gene pool, including in response to recent changes in breeding practices (22), and adaptation to extreme environments (48). Finally, and perhaps surprisingly, ancient DNA did not only improve our understanding of the past but provided important lessons for the future conservation of endangered populations, including the endangered Przewalski's horses (19, 28). This review covers the main findings resulting from a decade of ancient DNA analyses in horses and presents several areas where further research is necessary (see 49, 76 for reviews covering complementary aspects).

2. Horses today

There are almost 60 million horses on the planet today. They are mostly confined to the sport and leisure industries in most developed countries but still fulfill their traditional roles in developing countries where they provide transport, ploughing force, meat, milk, hair and leather. In addition to stud horses or farm animals, many so-called wild horse populations can be found around the world, including the iconic mustangs from the American West, the Sable Island ponies and the Australian brumbies. All these populations but one in fact descend from modern domestic horses that returned feral within the last few centuries. Przewalski's horses represent the only exception and are generally considered as the last truly wild horses on the planet, as they were discovered roaming wild in the second half of the 19th century in the semi-desert steppes of Dzungaria, North-West China (10). They typically show a dun brown-colored coat, an erected black mane continuing as a dorsal black stripe to the tail, and black striped legs. The population of Przewalski's horses is extremely reduced in numbers, and comprises approximately one animal for every ~29,000 domestic horses. The last animal ever captured in the wild was in 1947 and the last confirmed wild-sighting in 1969 (9). The animal was in fact once declared extinct in the wild but was saved from total extinction in captivity, where the stock could be rebuilt from the contribution of 12-16 foundation lineages (19). Its reintroduction in natural reserves in Mongolia, China and Russia represents one of the chief flagship examples for conservation biology.

Before the era of genome sequencing, the phylogenetic relationships between domestic and Przewalski's horses remained contentious. While microsatellite data (9) and Y-chromosome haplotyping (99) clearly separated both groups, other markers such as mitochondrial DNA (72) and multiple gene introns located on the X chromosome and the autosomes showed Przewalski's horses contended within a larger group of domestic variation (45). This could be indicative of incomplete lineage sorting, where the tree reconstructed from single genes could differ from the average population tree due to variation present in the ancestral population at the time of divergence (60). Alternatively, the evidence available may indicate that any clear separation between both groups should be rejected or that two independent lineages formed but remained connected by gene flow post-divergence. The interpretation was further complicated by the recent history of Przewalski's horses, in which domestic genes may have been introduced through uncontrolled admixture in captivity in what would otherwise appears as a divergent lineage. Despite the presence of one additional pair of chromosomes in the karyotype of Przewalski's horses relative to domestic horses ($2n=66$ vs $2n=64$), both lineages can indeed mate and produce fertile offspring ($2n=65$) (10). In fact, whether some foundation Przewalski's horses of the captive pedigree were pure Przewalski's horses, hybrids or even Mongolian domestic horses has been debated.

Bead-chip genotyping of ~54,000 SNP markers spread across the autosomes revealed Przewalski's horses as a sister group to all domestic breeds investigated (63). While this confirmed the presence of two separate lineages, this did not explicitly rule out the possibility of gene flow. The first genome sequence of a Przewalski's horse was obtained in 2013 and was compared to that of five domestic breeds encompassing the whole range of domestic variation (73). D-statistics calculations testing the balance of shared derived variants between a pair of domestic horses and the Przewalski's horses indicated no to marginal significance. This suggested that domestic and Przewalski's horses likely remained isolated after they diverged. The timing of their divergence was estimated to ~38,000-72,000 years ago, leveraging the molecular clock that could be calibrated based on the genome sequence of the oldest genome (73). Further work based on additional data and different modelling approaches refined the divergence dates to ~35,000-54,000 years ago (19) and ~35,000-44,000 years ago (22).

One individual genome may, however, not be representative of the whole Przewalski's horse pedigree and the animals that survived extinction in captivity may not fully encapsulate the whole variation once present in their lineage. This issue could be addressed after further genomes of Przewalski's horses were sequenced, including living descendants of all foundation lineages, along with animal tissues preserved in museum collections, and including: one alleged hybrid called Theodore that took part of the second generation living in captivity and died captive in 1927, and various animals caught in the wild in the 19th century, including the holotype which died in

1878 (19). The presence of almost a quarter of the domestic ancestry within Theodore's genome provided unambiguous evidence of admixture within the first generations of captivity (it, however, suggested that the animal was in fact a second rather than a first-generation hybrid, as previously thought; 113). The genome data also established for the first time that the Przewalski's horses living today showed various excess of domestic ancestry relative to animals caught in the wild in the 19th century. This indicated some heterogeneity in the captive pedigree, with some lineages showing rather limited domestic ancestry (estimated to 1.2-3.9% or 10.4-13.8% depending on the underlying methodology) and others showing significant amounts (10.5-13.7% and 28.4-31.1%, respectively).

Interestingly, statistical modelling suggested that while the lineages of domestic and Przewalski's horses were connected by significant gene flow until roughly the Last Glacial Maximum (LGM, ~19,000-26,000 years ago), they remained almost-to-fully isolated since and until domestic genes were introgressed within Przewalski's horses during captivity (19). This work, thus, revealed that significant changes were introduced to the gene pool of Przewalski's horses while surviving in captivity, including significant admixture with domestic horses, increasing inbreeding and substantial loss of genetic diversity. Further work indicated that the mutational load present in the genome of Przewalski's horses is the largest measured within a panel of 37 horse populations and breeds, except for Friesian horses (74). Altogether, the available evidence indicates that conservation efforts must be continued, despite the now relatively improved endangered IUCN status for the species (International Union for Conservation of Nature).

3. The complexity of early domestication stages

As Przewalski's horses have diverged from the lineage that led to modern domestic horses at least ~30,000 years prior to any archaeological evidence for horse domestication (19, 22, 73), the differences present in their genomes do not inform about the domestication process. In an attempt to identify the first genetic changes underlying early domestication stages and accompanying the domestic transformation of wild horses, Gaunitz and coworkers sequenced the genome of the earliest horses known in the archaeological record (28). More specifically, they sequenced 20 genomes from the osseous assemblages excavated at Botai, a settlement site from the North-Central steppes of Kazakhstan dated to approximately ~5,500 years ago, and where the first evidence for horse harnessing, milking and corralling was reported (78). In addition to five genomes from Borly4, another site located further east in the Kazakh steppes and dated to ~5,000 years ago, they also sequenced 17 genomes from archaeological horse remains spread across Eurasia and spanning the last ~4,100 years. Far from clarifying the evolutionary roots of modern domestic horses, this extensive genome dataset revealed instead that Botai-Borly4 horses in fact belonged to the Przewalski's horse lineage and that all other modern and ancient horses analyzed in the study formed another lineage (hereafter referred to as DOM2, Figure 2). Gene flow from Botai-Borly4 horses was estimated to be minimal, representing only an average 2.7% genome ancestry within the domestic lineage (range = 2.0-3.8%; 28) and while the effective size of the maternal line started to decline from after ~35,000 years ago, Bayesian skyline profiles indicated that the trend suddenly reverted in the third millennium BCE (Before Common Era, ~4,000-5,000 years ago), where it expanded by several orders of magnitude.

One major implication of this study is that the genetic makeup of domestic horses dramatically changed between ~5,000 and ~4,100 years ago. One possibility is that while horses were first domesticated at Botai, they were domesticated a second time independently from a source population radically different genetically. Another possibility invokes introgressive capture (44) as a process leading to the progressive, rapid dilution of the Botai-ancestry in the domestic genome as the population expanded away from its initial geographic range and admixed with local populations. Disentangling both scenarios will require to map the genetic structure of horse populations during and prior to the third millennium BCE. This will not represent a trivial task given the paucity of the horse archaeological record during that time period (91).

It is noteworthy that this time period also played host to large-scale population movements in humans. For example, ancient genomic work on human remains has revealed that the people who developed the Corded-Ware culture some ~4,200-5,000 years ago in the region known today as Germany, trace about two-thirds of their ancestry into the so-called Yamnaya steppe cattle-herders (2, 30, 69). The end of the third millennium BCE is also marked by the emergence of spoke-wheeled chariots drawn by horses within the so-called Sintashta-Petrovka horizons, dated to the period between ~4,100 and ~3,800 years ago (77). For now, whether the Yamnaya and Sintashta-Petrovka phenomena triggered the massive change observed in the genetic makeup of

domestic horses remains unknown. Interestingly, from ~3,900 years ago, archeological bone assemblages start to show a strong over-representation of horse males (23). This suggests that the emergence of a genetically-new type of domestic horse did not just coincide with the development of novel utilizations for the horse. It also extended to the animal domestic sphere the rise of a new, gendered representation of the world in human societies, in which gender inequalities developed (88).

Finding Przewalski's horses nested within Botai and Borly4 horses has another major implication, especially considering that models where the latter are directly ancestral to the former could not be rejected (28). This indicates that those horses that were domesticated at Botai for the first time survived after returning in the wild until their discovery in the 19th century. This supports a prey-domestication pathway, in which the domestication incentive for settled Botai hunter-gatherers was to maintain access to an animal that represented the core of their subsistence economy, at a time of declining resources. This contrasts with directed pathways of domestication, which are most commonly proposed for horses (44, 112), in which the first horse herders directly applied to horses the experience gained by several thousand years of domestication of large herbivores in order to deliberately select specific characteristics that fit their purpose (e.g. speed).

4. Extinction of multiple divergent horse lineages during the last ~5,000 years

Another surprise that arose after sequencing ancient genomes was the discovery that a larger diversity of horse lineages roamed Eurasia at the time of domestication. The first such lineage was identified from two specimens excavated in the Taymir peninsula, North-East Siberia and radiocarbon dated to approximately ~16,000 and ~43,000 years ago (90). Further ancient genome sequencing revealed that this lineage was not confined to North Siberia but likely occupied most of Siberia, as far South as the Tuva Republic (in the Altai-Sayan mountain range; 28), and survived at least in the region of Yakutia until ~5,000 years ago (48). This overlaps the geographic and temporal range of a taxonomic group described in the paleontological record as *Equus lenensis* (6). Statistical modeling including coalescent simulations under an isolation model and Approximate Bayesian Computation (90), Pairwise Sequential Markov Coalescent demographic profiling (47, 48) and demographic fit of multidimensional Site Frequency Spectrum (SFS; 22), indicated split times around ~110,000-130,000 years ago from the lineage ancestral to both modern domestic and Przewalski's horses. This is equivalent to or slightly younger than the coalescent time of all mitochondrial haplotypes segregating amongst modern horse lineages (1, 28, 50, 55).

Finding that *Equus lenensis* occupied an early diverging branch of the horse tree does not rule out per se their possible contribution to the genetic makeup of modern domestic horses. In order to address this, Schubert and coworkers (90) calculated D-statistics assessing whether this lineage shared relatively more derived variants with modern domestic horses than with Przewalski's horses. The unbalance observed was statistically significant, suggesting that a substantial fraction of the *Equus lenensis* genome may still segregate within modern domestic horses. An alternative scenario is that, instead of gene flow into domesticates, this signal reflects the presence of a geographically-structured ancestral population, in which the founder populations of both *Equus lenensis* and modern domesticates originated from similar regions (hence, shared an excess of common variants). Subtle changes along the branch leading to modern Przewalski's horses may also explain this signal (for example, in the case the latter admixed with an unrelated ghost population after it diverged from the lineage leading to modern domestic horses). That D-statistics in the form of (outgroup, *Equus lenensis*; Botai-Borly4, modern DOM2) do not significantly deviate from zero (28) lent support for the latter hypothesis. It is thus likely that subtle genomic changes have affected the Przewalski's horse lineage after ~5,000 years ago, while returning to the wild.

Further genome sequencing of ancient horses that lived in the region that now forms Spain and Portugal identified the presence of another divergent horse lineage in Iberia during the third millennium BCE (some ~3,993 to 4,791 years ago, Figure 1; 22). Climate niche modeling indicated that the environmental conditions present in the region have been compatible with the survival of horses throughout at least the last ~44,000 years (46). It can be, thus, reasonably hypothesized that this lineage represents the native lineage of horses in Western Europe, which survived in the Iberian glacial refugia during the LGM (103). In the absence of specimens that showed DNA preservation levels compatible with genome sequencing before the Iron Age, some 2,500 years ago, it is currently unknown whether this lineage was extirpated from its native range as the domestic horse expanded in the region, or was successfully domesticated or introgressed significant ancestry into the early DOM2 domestic horses that entered the region (22).

Population model fit to the multi-dimensional SFS, however, ruled out direct genetic contribution into the genetic makeup of modern domestic horses. Further analyses deriving the possible genetic contribution of the Iberian divergent lineage from ratios of f_4 -statistics (79) indicated an average genome ancestry generally no larger than ~5% within modern and ancient DOM2 horses, suggesting at best marginal contribution to the genetic makeup of modern domestic horses. Microsatellite data in modern breeds have documented larger heterozygosity and haplotypic diversity within Iberia (103). Rather than supporting successful local domestication in Iberia horses, and genetic continuity until the present times, this most likely indicates that the recent demography of horse breeding in Iberia resulted in more limited losses of genetic diversity. In any case, the presence of at least four genetically differentiated lineages supports the existence of significant geographic structure within pre-domestication horse populations across Eurasia. This is in striking contrast to previous conclusions based on temporal and geographical patterns of mitochondrial variation in horses, which reported no (15, 40) to faint population structure across Eurasia (64). This illustrates both the limitations of demographic inference using single marker loci and opens the possibility for different modes of management between male and female bloodlines during domestication.

5. Resource management during domestication and modern breeding

Modern horses stand as an exception within animal domestic species. They show some of the largest genetic diversity in the mitochondrial genome (40, 97) but extremely limited variation within the Y chromosome (53, 101). This suggested domestication models including the contribution of only a limited number of stallions and extensive restocking from local mares (reviewed in 49). In early studies, analyses were limited to a small number of loci in the absence of an assembly for the Y-chromosome. Genotyping by pyrosequencing of these markers in ancient horses first indicated that this model may be over-simplistic as one Iron Age Scythian stallion from Arzhan (Tuva Republic) was shown to carry one haplotype absent from modern domestic bloodlines (54). Further sequencing of a larger number of loci became possible once longer contigs and scaffolds could be assembled for the Y-chromosome using modern horse samples (39, 100). Ancient DNA data uncovered the presence of previously uncharacterized levels of variation within modern domesticates, in particular in Asia (24, 48, 100), and levels of nucleotide diversity ~10-fold larger than today within 11 Pazyryk Scythian horses from Berel' (Kazakhstan), dated to approximately ~2,500 years ago (48).

The extensive genome time-series characterized by Fages and coworkers (22) included a total of 175 males, most of which were spread across the last ~4,000 years. This provided a unique opportunity to assess the genetic diversity of male bloodlines during the first ~1,500 years of DOM2 management, and during the ~2,500 years following Berel'. This confirmed that the Y-chromosome nucleotide diversity of horses in Asia and Europe was considerably larger in the past than it is today, and revealed parallel losses of diversity in both continents starting ~2,000 years ago and leaving only a remaining half of the diversity ~1,000 years ago (22). This pattern was in line with previous analyses based on a more limited subset of samples and SNP markers (109), and the coalescent time of haplotypes from Arabian-Oriental origins some ~700-1,100 years ago (24, 100). Additionally, autosomal genetic affinities between horse populations associated with individual archaeological sites and/or cultures showed almost all modern domestic breeds grouping together with ancient Persian Sassanid horses (4-5th century AD) and all the other DOM2 domestic horses analysed that post-dated the Arab-Muslim expansion (7-9th century AD) (22). This indicates common origins. The only exception was found in Icelandic horses and Shetland ponies, which grouped within a group including Viking and Roman horses. Altogether, these results ruled out the participation of only a limited number of stallion lineages from the earliest domestication stages. They also suggested that the dramatic lack of genetic diversity in modern horses was only introduced most recently, following the growing influence of Arabian-Oriental bloodlines in medieval and modern studs. This influence spread almost ubiquitously across European breeds, except in Iceland, and some islands from Britain and probably Scandinavia, where opportunities for outbreeding were limited or forbidden.

With increasingly complete genome time-series, it has become possible to go beyond the limitations of single marker analyses and chart patterns of diversity changes through time genome-wide, across all autosomes. For example, nucleotide diversity can be estimated in predefined time bins to inform on the average amount of genetic differences present between pairs of samples. Fages and collaborators (22) implemented such calculations within time-bins of 250 years and found diversity estimates for both Europe and Asia ~12-40% larger in all time bins preceding 1600 AD. Additionally, they measured heterozygosity levels in individual genomes for which sufficient coverage was achieved and confirmed a ~16% drop within the last ~200 years. This drop was mirrored by an increase in the total amount of mutations present in a given genome that were predicted to be deleterious

on the basis on conservation scores (83). This finding was further confirmed extending such calculations of the mutational load outside genomic regions coding from proteins and considering a larger comparative panel of 175 modern genomes (instead of 30; 74). Combined, these results suggest that the development of close studs and purebred lines within the last few centuries, in addition to the deprecation of horsepower caused by mechanization, seem to have impacted on the genetic diversity of the horse more than any other period in the previous four millennia. Close stud breeding represented a massive shift in horse husbandry techniques, which fragmented the horse population in small reproductive isolates (the so-called breeds), and along with the reduced utility of the horse, increased genetic drift and mutational load, by reducing the power of purifying selection to eliminate deleterious variants from the reproductive stock.

It is, however, important to notice that such diversity measurements are not devoid of potential issues. As the statistical power to call heterozygous sites increases with coverage (12), the generally limited coverage achieved in ancient genomics is expected to bias estimates downwards. At the same time, sequencing error rates are generally inflated in ancient DNA sequence data due to post-mortem DNA damage causing systematic nucleotide mis-incorporations, mostly C→T transitions and their complementary G→A (41). This is expected to bias estimates upwards, especially with lower coverage data, as inflated sequencing error rates increase the chances to observe two different alleles at any given position. This effect can be mitigated computationally by conditioning analyses on transversions and/or through enzymatic reactions removing the most prevalent post-mortem DNA damage prior to construction of DNA libraries and sequencing (11). Both such precautions were implemented in the analyses of Fages and collaborators (22), who also controlled that their calculations were robust to variation in coverage and sequencing errors.

The calculation of mutational loads is not devoid of potential issues either, including the fact that it is unknown whether individual mutations are recessive, dominant or co-dominant. Conditioning on homozygous sites as performed by Fages and collaborators (22) seem to have mitigated these effects as simulations returned mutational load values decreasing with increasing selection impact. Interestingly, mutational loads calculated in the early domestication stages were on par with those measured during more recent historical periods (22, 74). This confirmed previous observations on a more limited number of ancient genomes and rejects the so-called '*cost-of-domestication*' hypothesis, positing that early domestication stages result in massive demographic bottlenecks, hence, limit purifying selection and inflate mutational loads. More than a '*cost-of-domestication*' per se, a '*cost-of-modern-breeding*' seems to better reflect the patterns of variation in genomic data currently available for horses.

6. Selection for appearance and pace and other traits

It is clear that changes in horse husbandry and management techniques have had a global impact on the horse genome. However, by selecting animals of particular appearance or pace, past herders and breeders have also affected the horse genome more locally, within those regions underpinning target traits. The time-stamped nature of ancient DNA data allows us to reconstruct individual allelic trajectories through time. This considerably increases the statistical power to measure the strength of selection and to estimate the time when it started (61, 89).

After humans, horses certainly represent the animal where most of such work has been done thus far. Most of the attention has focused on variants either associated with or causative for particular coat coloration phenotypes (22, 58, 59, 84, 107) but significant work has been done on speed (8, 22) and ambling (22, 106). This type of studies developed even before the advent of next-generation DNA sequencing and relied then on multiplex PCR amplifications of short fragments overlapping SNP of interest, followed by pyrosequencing (see 108 for a methodological overview). The main limitation of such approaches was that the available DNA extract could be rapidly exhausted, owing to the necessarily large number of PCR replicates that had to be done to rule out allelic dropout (93). Currently, these approaches have been superseded by whole genome sequencing, which provides allele samples at the whole genome scale (22), or target-enrichment techniques, which retrieve sequence data at all loci of interest in parallel (e.g. ~5,000 in 16).

Relative to their wild counterparts, domestic animals often display a full variety of coat coloration phenotypes (see 52 for a review). Association studies in modern horses has identified the genetic basis of many such phenotypes, including for different colors (e.g. dun, chestnut, white), patterns (e.g. leopard spotting) and dilutions (cream, tobiano, sabino). This provided a unique opportunity to use ancient DNA to predict the

phenotype of ancient horses, and identify time periods when particular colors become particularly popular or were superseded by others. For example, solid chestnut phenotypes have gained considerable attraction during medieval times in Western Europe, while spotted and diluted coats seem to have been preferred during the Iron Age (107). Determining how much this reflects the influence of Christianity, especially the four horse coats depicted in the four riders of the Apocalypse, or other non-religious factors, requires further work. In particular, as medieval specimens are often excavated from the graves of knights or noblemen, the dominance of the chestnut coloration observed may simply reflect social preferences.

Leopard spotted horses seem to have lost attraction after the early Bronze Age, a time when they were particularly abundant in Turkish-Thrace site of Kirklareli-Kanligecit (2,700-2,200 BCE; 59), before they regained attention during the Iron Age (107). Interestingly, Leopard Spotting is associated with Congenital Night Blindness, an incompletely dominant syndrome limiting vision in the dark in homozygous individuals. Leopard Spotting is, thus, likely negatively selected in the wild and the positive selection coefficients estimated at Kirklareli-Kanligecit suggest that the horse population analyzed was domesticated, and not hunted. Whether these horses belonged to the DOM2 or the Botai-Borly4 lineage is presently unknown. However, the presence of the tobiano allele and the one causative for Leopard Spotting observed at Botai suggests that a diversity of coat phenotypes was contended on site, which nicely complements archaeological arguments supporting the domestic status of these horses (28, 107). Finally, even though ancient DNA data clearly established the rise of coat phenotypes with domestication, it is noteworthy that the dun, black and leopard spotted alleles have been documented in wild horses, including during the Upper Paleolithic (4, 38, 58, 84). Interestingly, these phenotypes are also depicted in cave paintings, which may suggest that parietal art could simply express naturalistic and not symbolic representations of the world (84).

Before even coat colors, speed is perhaps the most characteristic trait that comes to mind when picturing a horse. Given the economic importance of the racing industry, important research programmes have endeavored to decipher the genetic basis of fast-distance racing. One important locus involved is within the myostatin gene, in which homozygous (heterozygous) Thoroughbred carriers of a SINE insertion (Short Interspersed Nuclear Element) show better racing performance at short (middle) distances while homozygous non-carriers have greater stamina (82). Other genomic sites, both within and outside the MSTN locus, show strong association with the traits (29, 36, 37). Interestingly, many of the alleles associated with middle-to-short distance racing performance increased in frequency within the last ~1,100-1,600 years, which suggests that stamina was more valued than speed during most of the history of domestication. One such alleles within the MSTN locus appears to have probably entered the Thoroughbred bloodline from local British mares as it was absent from the Darley Arabian stallion whose line sired ~95% of all living Thoroughbreds (8).

In addition to speed, gait also received some attention, especially ambling. This trait corresponds to the four-beat gait in which horses go at a slow pace moving their pairs of legs laterally. It shows a simple genetic basis where a premature stop codon in the DMRT3 gene is permissive for ambling, which provides a more comfortable riding experience, and improves harness racing performance (3). Nonetheless, the DMRT3 allele responsible for ambling could only be detected thus far in samples not older than the 9th century AD (22, 106), and only rose in frequency since in particular breeds and/or contexts. In Iceland, it entered the Icelandic horse stock as part of the foundation stock that the Vikings originally brought on the island from Scandinavia and/or the British Isles (106). Combined, the data available thus far tend to suggest that speed and comfortable gaits have become attractive only recently, probably in the last ~1,000 years.

Other approaches to detect signatures of positive selection in ancient horse genomes have leveraged other statistics than temporal profiles of allele frequencies. One such approach, the Population Branch Statistics (PBS), was originally developed in the context of identifying the genetic basis of life at high-altitude in Tibetans (110). The method considers three populations, including one focal population and one closely related, along with an outgroup. The PBS values are then calculated within non-overlapping genomic window from the F_{st} differentiation indexes between each population pair, and allow to identify candidate loci for selection along the branch leading to the focal population. Other methods, such as GRoSS (86) or LSD (51) make use of different principles and can apply to more complex evolutionary trees including more than three populations. Still other do not assume that populations are isolated and can account for admixture (14).

The application of these approaches to a dataset including the genomes of 11 Iron Age horses excavated at Berel' in the Kazakh range of the Altay mountains, 15 modern breeds, and one outgroup indicated that Scythian Pazyryk nomad herders may have selected a special horse morphotype, showing more robust forelimb

metapodials, as the top-PBS windows were over-represented in genes involved in the forelimb development (50). Linear morphometric measurements confirmed this prediction. Likewise, the inclination for new morphological features may have contributed to the rising influence of Arabian-Oriental bloodlines from the 9th century AD in Europe and Central and South-West Asia, as numerous genes involved in key developmental stages for the formation of the body axis showed top-LSD scores in Byzantine horses, relative to pre-Muslim Roman and Mongolian horses (22). LSD scores along the branch leading to a panel of 45 ancient and modern domestic horses further pointed to a number of genes involved in the development of the neural crest or neural crest derived tissues, which lent support to the '*neural crest hypothesis*' (105). This model was proposed to explain how the application of one selection pressure affecting the development of a temporary population of cells which migrate and differentiate during development in many derivative tissues may have systematically resulted in the emergence of common phenotypes across domestic animals, collectively grouped within the so-called '*domestication syndrome*'.

Combined, and complemented by classical scans for selection signatures in the genomes of modern horses living at high-altitude (56) or the extremely cold region of North-East Siberia (48), these studies have considerably advanced our understanding of how artificial and natural selection have sculpted the horse genetic makeup through time, and grounded the development of entire arrays of horse types and possible modes of exploiting and interacting with the animal. Nonetheless, future work should aim at detecting signatures of polygenic adaptation (85) as most of the complex behavioral and performance traits can be expected to involve minute contributions of a large number of gene variants, rather than single loci with exceptionally large effects.

7. Conclusion and Perspectives

The horse is certainly amongst the animals that most influenced human history (43). It provided the first opportunity for fast long-distance travels, effectively globalizing the world for the first time, and considerably changed warfare. Even though ancient horse genomes have started to document the process of horse domestication, management and improvement with unprecedented details, lots remain to be discovered (see the Future Issues list provided below for an example of pending questions). Perhaps most critically, the geographic source for the expansion of the modern domestic horse lineage during the third millennium BCE remains presently unknown. The exact role that horses played in the expansion of steppe herders then, but also in other, more recent time periods, especially concomitantly with the development of mounted combat during the early Iron Age, some ~2,800 years ago (21), are also unclear.

Importantly, with the exception of one study focused on the Pazyryk Scythian horses of Berel' (50), most of the recent work was aimed at characterizing global patterns of variation in ancient horse populations. Judging from current trends in ancient human genomics (e.g. 67), it can be expected that future work will increasingly focus on more local scales and address how the many (past) cultures and horse husbandry practices impacted on horse genetics. This work is expected to go beyond simple descriptions of sex recruitment bias in various funerary contexts (e.g. 23, 70) but to explore potential segregation of horse traits within social classes, including elites, military chiefs and farmers (81).

All previous genome work has also relied on shotgun sequencing. For economic reasons, this restricted analyses to the fraction of archaeological bone assemblages showing the best DNA preservation levels, and a limited number of specimens per site. Now that this work has framed the genomic variation present in ancient horse populations across Eurasia, polymorphic sites can be selected across the genome and ascertained to reflect variation that is not present or frequent in specific horse bloodlines over-represented in the modern industry. Target-enrichment techniques (16) could then help focus sequencing efforts to such sites (and other variants causative or associated with important phenotypic traits), which will both considerably reduce underlying analytical costs and increase the range of ancient specimens from which ancient DNA data could be obtained. This can prove invaluable in regions such as Turkey, Iran, Greece, Macedonia and Arabia, where specific equestrian practices developed throughout history, but where ancient DNA preservation is generally limited. It can also be anticipated that the enhanced sensitivity provided by these approaches could also benefit the genetic monitoring of the many living endangered populations of horses, including the CITES-protected Przewalski's horses, for which only non-invasive sampling can be performed, hence, limited DNA material can be made available.

It is remarkable that all previous ancient DNA work has entirely focused on SNP variation. As a consequence, other important sources of genetic variation have been overlooked, including segmental duplication and other structural variation, as well as epigenetic regulatory layers and microbiome composition. Yet, copy number variants represent important contributors of phenotypic variation in dogs (68) and cattle (94). While the characterization of various epigenetic layers of gene regulation is in infancy in modern horses thanks to the significant efforts of the FAANG consortium (13), new statistical methods leveraging systematic bias in post-mortem DNA decay have started to reveal the first ancient epigenetic maps, including cytosine DNA bone and tooth methylomes (see 31 for a review) and nucleosome maps (80). Assuming that current technical difficulties can be overcome, in particular that high-coverage genome data can be gathered for a large number of individuals (32), future ancient DNA work may start exploring how gene expression regulatory networks have been modified during the different phases of the domestication process. Importantly, with the sequencing ancient dental plaque, it has also become possible to characterize ancient oral microbiomes (see 102 for a review) and important differences have been reported in the taxonomic composition of the fecal microbiome of captive and non-captive animals (65). This is also true for horses in which Przewalski's horses the community of fecal bacteria is distinct and more diverse than in Mongolian domestic horses, and superior in animals born in natural reserves than in zoos (66). Future research programmes may start to uncover the impact that domestication and breeding have had on the horse microbiome, which represents a second genome involved in the establishment and maintenance of crucial physiological and cognitive functions for the host, in relation with lifestyle (71, 92).

Finally, even though this review was strictly focused on horses, similar ancient genomic work is ongoing for a number of other domestic animal species, including dogs (7, 26), cattle (95), pigs (27) and goats (17). As seen for horses, these studies have revealed more complex domestication models than previously anticipated (see 25 for a review). There is little doubt that future work will continue to improve our current understanding of the domestication in each of these species. Meta-analyses across species, including the horse, will in turn help address the biological pathways that may have been convergently selected and have provided the basis for the appearance of common traits often shared across domesticated animals (57).

Literature Cited

1. Achilli A, Olivieri A, Soares P, Lancioni H, Hooshiar Kashani B, et al. 2012. Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proc Natl Acad Sci U S A* 109:2449-54
2. Allentoft ME, Sikora M, Sjögren KG, Rasmussen S, Rasmussen M, et al. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522:167-72
3. Andersson LS, Larhammar M, Memic F, Wootz H, Schwochow D, et al. 2012. Mutations in *DMRT3* affect locomotion in horses and spinal circuit function in mice. *Nature* 488:642-6
4. Bellone RR, Holl H, Setaluri V, Devi S, Maddodi N, et al. 2013. Evidence for a retroviral insertion in *TRPM1* as the cause of congenital stationary night blindness and leopard complex spotting in the horse. *PLoS One* 8:e78280
5. Bennett EA, Champlot S, Peters J, Arbuckle BS, Guimaraes S, et al. 2017. Taming the late Quaternary phylogeography of the Eurasiatic wild ass through ancient and modern DNA. *PLoS One* 12:e0174216
6. Boeskorov GG, Potapova OR, Protopopov AV, Plotnikov VV, Maschenko EN, et al. 2018. A study of a frozen mummy of a wild horse from the Holocene of Yakutia, East Siberia, Russia. *Mammal Res* 63:1-8
7. Botigué LR, Song S, Scheu A, Gopalan S, Pendleton AL, et al. 2017. Ancient European dog genomes reveal continuity since the Early Neolithic. *Nat Commun* 8:16082
8. Bower MA, McGivney BA, Campana MG, Gu J, Andersson LS, et al. 2012. The genetic origin and history of speed in the Thoroughbred racehorse. *Nat Commun* 3:643
9. Bowling AT, Zimmermann W, Ryder O, Penado C, Peto S, et al. 2003. Genetic variation in Przewalski's horses, with special focus on the last wild caught mare, 231 Orlitza III. *Cytogenet Genome Res* 102:226-34
10. Boyd L, Houpt KA. 1994. Przewalski's Horse: The History and Biology of an Endangered Species (SUNY Series in Endangered Species). SUNY Series in Endangered Species, pp. 332.
11. Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, et al. 2010. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res* 38:e87
12. Bryc K, Patterson N, Reich D. 2013. A Novel Approach to Estimating Heterozygosity from Low-Coverage Genome Sequence. *Genetics* 195:553-61
13. Burns EN, Bordbari MH, Mienaltowski MJ, Affolter VK, Barro MV, et al. 2018. Generation of an equine biobank to be used for Functional Annotation of Animal Genomes project. *Anim Genet* 49:564-570.

14. Cheng JY, Mailund T, Nielsen R. 2017. Fast admixture analysis and population tree estimation for SNP and NGS data. *Bioinformatics* 33:2148-55
15. Cieslak M, Pruvost M, Benecke N, Hofreiter M, Morales A, et al. 2010. Origin and history of mitochondrial DNA lineages in domestic horses. *PLoS One* 5:e15311
16. Cruz-Dávalos DI, Llamas B, Gaunitz C, Fages A, Gamba C, et al. 2017. Experimental conditions improving in-solution target enrichment for ancient DNA. *Mol Ecol Resour* 17:508-22
17. Daly KG, Maisano Delser P, Mullin VE, Scheu A, et al. 2018. Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science* 361:85-8
18. Der Sarkissian C, Allentoft ME, Ávila-Arcos MC, Barnett R, Campos PF, et al. 2015. Ancient genomics. *Philos Trans R Soc Lond B Biol Sci* 370:20130387
19. Der Sarkissian C, Ermini L, Schubert M, Yang MA, Librado P, et al. 2015. Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. *Curr Biol* 25:2577-83
20. Der Sarkissian C, Vilstrup JT, Schubert M, Seguin-Orlando A, Eme D, et al. 2015. Mitochondrial genomes reveal the extinct Hippiidion as an outgroup to all living equids. *Biol Lett* 11:pii:20141058.
21. Drews R. 2004. *Early Riders. The Beginnings of Mounted Warfare in Asia and Europe*. London/New York: Routledge, pp 218
22. Fages A, Hanghøj K, Khan N, Gaunitz C, Seguin-Orlando A, et al. 2019. Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. *Cell* 177:1419-1435.e31
23. Fages A, Seguin-Orlando A, Germonpré M, Orlando L. 2020. Horse males became over-represented in archaeological assemblages during the Bronze Age. *J Arc Sci*, in press.
24. Felkel S, Vogl C, Rigler D, Jagannathan V, Leeb T, et al. 2018. Asian horses deepen the MSY phylogeny. *Anim Genet* 49:90-3
25. Frantz L, Larson G, Bradley D, Orlando L. 2020. Animal domestication in the era of ancient genomics. *Nat Genet Rev*, in press.
26. Frantz LA, Mullin VE, Pionnier-Capitan M, Lebrasseur O, Ollivier M, et al. 2016. Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science* 352:1228-31
27. Frantz LAF, Haile J, Lin AT, Scheu A, Geörg C, et al. 2019. Ancient pigs reveal a near-complete genomic turnover following their introduction to Europe. *Proc Natl Acad Sci U S A* 116:17231-8
28. Gaunitz C, Fages A, Hanghøj K, Albrechtsen A, Khan N, et al. 2018. Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science* 360:111-4
29. Gu J, MacHugh DE, McGivney BA, Park SD, Katz LM, et al. 2010. Association of sequence variants in CKM (creatine kinase, muscle) and COX4I2 (cytochrome c oxidase, subunit 4, isoform 2) genes with racing performance in Thoroughbred horses: SNP association with elite racing performance. *Equine Vet J* 42:569-575.
30. Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522:207-11
31. Hanghøj K, Orlando L. 2018. Ancient epigenomics. In: Lindqvist C, Rajora OP, eds. *Paleogenomics*. Cham: Springer, pp 1-37.
32. Hanghøj K, Renaud G, Albrechtsen A, Orlando L. 2019. DamMet: ancient methylome mapping accounting for errors, true variants, and post-mortem DNA damage. *Gigascience* 8 pii:giz025
33. Heintzman PD, Zazula GD, MacPhee R, Scott E, Cahill JA, et al. 2017. A new genus of horse from Pleistocene North America. *Elife* 6 pii:e29944
34. Hendricks BL. 2007. *International Encyclopedia of Horse Breeds*. pp 486
35. Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312:282-4
36. Hill EW, Gu J, Eivers SS, Fonseca RG, McGivney BA, et al. 2010. A sequence polymorphism in MSTN predicts sprinting ability and racing stamina in thoroughbred horses. *PLoS One* 5:e8645
37. Hill EW, McGivney BA, Gu J, Whiston R, MacHugh DE. 2010. A genome-wide SNP-association study confirms a sequence variant (g.66493737C>T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses. *BMC Genomics* 11:552
38. Imsland F, McGowan K, Rubin CJ, Henegar C, Sundström E, et al. 2016. Regulatory mutations in TBX3 disrupt asymmetric hair pigmentation that underlies Dun camouflage color in horses. *Nat Genet* 48:152-8
39. Janečka JE, Davis BW, Ghosh S, Paria N, Das PJ, et al. 2018. Horse Y chromosome assembly displays unique evolutionary features and putative stallion fertility genes. *Nat Commun* 9:2945
40. Jansen T, Forster P, Levine MA, Oelke H, Hurler M, et al. 2002. Mitochondrial DNA and the origins of the domestic horse. *Proc Natl Acad Sci U S A* 99:10905-10
41. Jónsson H, Ginolhac A, Schubert M, Johnson PL, Orlando L. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29:1682-4

42. Jónsson H, Schubert M, Seguin-Orlando A, Ginolhac A, Petersen L, et al. 2014. Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proc Natl Acad Sci U S A* 111:18655-60
43. Kelekna P. 2009. *The horse in human history*. Cambridge University Press. pp 476.
44. Larson G, Fuller DQ. 2014. *The Evolution of Animal Domestication*. *Ann Rev Ecol Evol Syst* 45:115-36
45. Lau AN, Peng L, Goto H, Chemnick L, Ryder OA, et al. 2009. Horse domestication and conservation genetics of Przewalski's horse inferred from sex chromosomal and autosomal sequences. *Mol Biol Evol* 26:199-208
46. Leonardi M, Boschin F, Giampoudakis K, Beyer RM, Krapp M, et al. 2018. Late Quaternary horses in Eurasia in the face of climate and vegetation change. *Sci Adv* 4:eaar5589
47. Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature*. 475:493-6
48. Librado P, Der Sarkissian C, Ermini L, Schubert M, Jónsson H, et al. 2015. Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proc Natl Acad Sci U S A* 112:E6889-97
49. Librado P, Fages A, Gaunitz C, Leonardi M, Wagner S, et al. 2016. The Evolutionary Origin and Genetic Makeup of Domestic Horses. *Genetics* 204:23-434
50. Librado P, Gamba C, Gaunitz C, Der Sarkissian C, Pruvost M, et al. 2017. Ancient genomic changes associated with domestication of the horse. *Science* 356:442-5
51. Librado P, Orlando L. 2018. Detecting Signatures of Positive Selection along Defined Branches of a Population Tree Using LSD. *Mol Biol Evol* 35:1520-35
52. Linderholm A, Larson G. 2013. The role of humans in facilitating and sustaining coat colour variation in domestic animals. *Sem Cell Dev Biol* 24:587-93
53. Lindgren G, Backström N, Swinburne J, Hellborg L, Einarsson A, et al. 2004. Limited number of patrilineages in horse domestication. *Nat Genet* 36:335-6
54. Lippold S, Knapp M, Kuznetsova T, Leonard JA, Benecke N, et al. 2011. Discovery of lost diversity of paternal horse lineages using ancient DNA. *Nat Commun* 2:450.
55. Lippold S, Matzke NJ, Reissmann M, Hofreiter M. 2011. Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. *BMC Evol Biol* 11:328.
56. Liu X, Zhang Y, Li Y, Pan J, Wang D, et al. 2019. EPAS1 gain-of-function mutation contributes to high-altitude adaptation in Tibetan horses. *Mol Biol Evol* pii:msz158
57. Lord KA, Larson G, Coppinger RP, Karlsson EK. 2020. The History of Farm Foxes Undermines the Animal Domestication Syndrome. *Trends Ecol Evol* 35:125-36
58. Ludwig A, Pruvost M, Reissmann M, Benecke N, Brockmann GA, et al. 2009. Coat color variation at the beginning of horse domestication. *Science* 324:485.
59. Ludwig A, Reissmann M, Benecke N, Bellone R, Sandoval-Castellanos E, et al. 2015. Twenty-five thousand years of fluctuating selection on leopard complex spotting and congenital night blindness in horses. *Philos Trans R Soc Lond B Biol Sci* 370:20130386
60. Maddison WP, Knowles LL. 2006. Inferring phylogeny despite incomplete lineage sorting. *Syst Biol* 55:21-30
61. Malaspina AS. 2016. Methods to characterize selective sweeps using time serial samples: an ancient DNA perspective. *Mol Ecol* 25:24-41
62. Marciniak S, Perry GH. 2017. Harnessing ancient genomes to study the history of human adaptation. *Nat Rev Genet* 18:659-74
63. McCue ME, Bannasch DL, Petersen JL, Gurr J, Bailey E, et al. 2012. A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. *PLoS Genet* 8:e1002451
64. McGahern A, Bower MA, Edwards CJ, Brophy PO, Sulimova G, et al. 2006. Evidence for biogeographic patterning of mitochondrial DNA sequences in Eastern horse populations. *Anim Genet* 37:494-7
65. McKenzie VJ, Song SJ, Delsuc F, Prest TL, Oliverio AM, et al. 2017. The Effects of Captivity on the Mammalian Gut Microbiome. *Integr Comp Biol* 57:690-704
66. Metcalf JL, Song SJ, Morton JT, Weiss S, Seguin-Orlando A, et al. 2017. Evaluating the impact of domestication and captivity on the horse gut microbiome. *Sci Rep* 7:15497
67. Mittnik A, Massy K, Knipper C, Wittenborn F, Friedrich R, et al. 2019. Kinship-based social inequality in Bronze Age Europe. *Science* 366:731-4
68. Molin AM, Berglund J, Webster MT, Lindblad-Toh K. 2014. Genome-wide copy number variant discovery in dogs using the CanineHD genotyping array. *BMC Genomics* 15:210
69. Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, et al. 2019. The formation of human populations in South and Central Asia. *Science* 365 pii:eaat7487

70. Nistelberger HM, Pálsdóttir AH, Star B, Leifsson R, Gondeka AT et al. 2018. Sexing Viking Age horses from burial and non-burial sites in Iceland using ancient DNA. *J Arc Sci* 101:115-22
71. O'Toole PW, Shiels PG. 2020. The role of the microbiota in sedentary life style disorders and ageing: Lessons from the animal kingdom. *J Intern Med* doi: 10.1111/joim.13021
72. Oakenfull EA, Ryder OA. 1998. Mitochondrial control region and 12S rRNA variation in Przewalski's horse (*Equus przewalskii*). *Anim Genet* 29:456-9
73. Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, et al. 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499:74-8
74. Orlando L, Librado P. 2019. Origin and Evolution of Deleterious Mutations in Horses. *Genes (Basel)* 10 pii:E649
75. Orlando L, Metcalf JL, Alberdi MT, Telles-Antunes M, Bonjean D, et al. 2009. Revising the recent evolutionary history of equids using ancient DNA. *Proc Natl Acad Sci U S A* 106:21754-9
76. Orlando L. 2020. Ancient Genomes Reveal Unexpected Horse Domestication and Management Dynamics. *Bioessays* 42:e1900164
77. Outram AK, Kasparov A, Stear NA, Varfolomeev V, Usmanova E, et al. 2012. Patterns of pastoralism in later Bronze Age Kazakhstan: new evidence from faunal and lipid residue analyses. *J Arc Sci* 39:2424-35
78. Outram AK, Stear NA, Bendrey R, Olsen S, Kasparov A, et al. 2009. The earliest horse harnessing and milking. *Science* 323 :1332-5
79. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, et al. 2012. Ancient admixture in human history. *Genetics* 192:1065-93
80. Pedersen JS, Valen E, Velazquez AM, Parker BJ, Rasmussen M, et al. 2014. Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res* 24:454-66
81. Perry GH, Makarewicz CA. 2019. Horse Paleogenomes and Human-Animal Interactions in Prehistory. *Trends Genet* 35:473-5
82. Petersen J, Mickelson JR, Rendahl AK, Valberg SJ, Andersson LS, et al. 2013. Genome-Wide Analysis Reveals Selection for Important Traits in Domestic Horse Breeds. *PLoS Genet* 9: e1003211
83. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 20:110–21
84. Pruvost M, Bellone R, Benecke N, Sandoval-Castellanos E, Cieslak M, et al. 2011. Genotypes of predomestic horses match phenotypes painted in Paleolithic works of cave art. *Proc Natl Acad Sci U S A* 108 :18626-30
85. Racimo F, Berg JJ, Pickrell JK. 2018. Detecting polygenic adaptation in admixture graphs. *Genetics* 208:1565-1584.
86. Refoyo-Martínez A, da Fonseca RR, Halldórsdóttir K, Árnason E, Mailund T, et al. 2019. Identifying loci under positive selection in complex population histories. *Genome Res* 29:1506-20
87. Rieder S, Taourit S, Mariat D, Langlois B, Gu G, et al. 2001. Mutations in the agouti (ASIP), the extension (MC1R), and the brown (TYRP1) loci and their association to coat color phenotypes in horses (*Equus caballus*). *Mamm Genome* 12:450–5
88. Robb J, Harris OJT. 2018. Becoming gendered in European prehistory: was Neolithic gender fundamentally different? *Am Antiquity* 83:128-47
89. Schraiber JG, Evans SN, Slatkin M. 2016. Bayesian Inference of Natural Selection from Allele Frequency Time Series. *Genetics* 203:493-511
90. Schubert M, Jónsson H, Chang D, Der Sarkissian C, Ermini L, et al. 2014. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc Natl Acad Sci U S A* 111:E5661-9
91. Sommer RS, Benecke N, Løugas L, Nelle O, Schmölcke U. 2011. Holocene survival of the wild in Europe: a matter of open landscape? *J Quat Sci* 26:805-12
92. Sonnenburg JL, Sonnenburg ED. 2019. Vulnerability of the industrialized microbiota. *Science* 366 pii:eaaw9255
93. Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, et al. 1996. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Res* 24:3189-94
94. Upadhyay M, da Silva VH, Megens HJ, Visker MHPW, Ajmone-Marsan P, et al. 2017. Distribution and Functionality of Copy Number Variation across European Cattle Populations. *Front Genet* 8:108
95. Verdugo MP, Mullin VE, Scheu A, Mattiangeli V, Daly KG, et al. 2019. Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. *Science* 365:173-6
96. Vilstrup JT, Seguin-Orlando A, Stiller M, Ginolhac A, Raghavan M, et al. 2013. Mitochondrial phylogenomics of modern and ancient equids. *PLoS One* 8:e55950
97. Vilà C, Leonard JA, Gotherstrom A, Marklund S, Sandberg K, et al. 2001. Widespread origins of domestic horse lineages. *Science* 291:474-7

98. Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, et al. 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 326:865-7
99. Wallner B, Brem G, Müller M, Achmann R. 2003. Fixed nucleotide differences on the Y chromosome indicate clear divergence between *Equus przewalskii* and *Equus caballus*. *Anim Genet* 34:453-6
100. Wallner B, Palmieri N, Vogl C, Rigler D, Bozlak E, et al. 2017. Y Chromosome Uncovers the Recent Oriental Origin of Modern Stallions. *Curr Biol* 27:2029-2035.e5
101. Wallner B, Vogl C, Shukla P, Burgstaller JP, Druml T, et al. 2013. Identification of genetic variation on the horse y chromosome and the tracing of male founder lineages in modern breeds. *PLoS One* 8:e60015
102. Warinner C, Herbig A, Mann A, Fellows Yates JA, Weiß CL, et al. 2017. A Robust Framework for Microbial Archaeology. *Annu Rev Genomics Hum Genet* 18:321-56
103. Warmuth V, Eriksson A, Bower MA, Cañon J, Cothran G, et al. 2011. European domestic horses originated in two holocene refugia. *PLoS One* 6:e18194
104. Weinstock J, Willerslev E, Sher A, Tong W, Ho SY, et al. 2005. Evolution, systematics, and phylogeography of pleistocene horses in the new world: a molecular perspective. *PLoS Biol* 3:e241
105. Wilkins AS, Wrangham RW, Fitch WT. 2014. The “Domestication Syndrome” in Mammals: A Unified Explanation Based on Neural Crest Cell Behavior and Genetics. *Genetics* 197:795–808
106. Wutke S, Andersson L, Benecke N, Sandoval-Castellanos E, Gonzalez J, et al. 2016. The origin of ambling horses. *Curr Biol* 26:R697-9
107. Wutke S, Benecke N, Sandoval-Castellanos E, Döhle HJ, Friederich S, et al. 2016. Spotted phenotypes in horses lost attractiveness in the Middle Ages. *Sci Rep* 6:38548
108. Wutke S, Ludwig A. 2019. Targeted PCR Amplification and Multiplex Sequencing of Ancient DNA for SNP Analysis. *Methods Mol Biol* 1963:141-7
109. Wutke S, Sandoval-Castellanos E, Benecke N, Döhle HJ, Friederich S, et al. 2018. Decline of genetic diversity in ancient domestic stallions in Europe. *Sci Adv* 4:eaap9691
110. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, et al. 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329:75-8
111. Yuan JX, Hou XD, Barlow A, Preick M, Taron UH, et al. 2019. Molecular identification of late and terminal Pleistocene *Equus ovodovi* from northeastern China. *PLoS One* 14:e0216883
112. Zeder M. 2012. Pathways to Animal Domestication. In Gepts P, Famula T, Bettinger R, Brush S, Damania A, McGuire A, et al. (Eds.), *Biodiversity in Agriculture: Domestication, Evolution, and Sustainability*, pp. 227-259. Cambridge: Cambridge University Press.
113. Zimmermann, W. 2013 *International Przewalski's Horse Studbook*.

Terms and Definitions

Post-mortem DNA damage: Biochemical and physical alterations to the DNA backbone that take place after death and result in the fragmentation of DNA into ultrashort molecules and the modification of nucleotides. Some of these modifications, such as cytosine deamination, can be used to sort out genuine ancient DNA templates and modern contamination.

Admixture: Evolutionary process involving the migration of individuals from at least one population source into a recipient population. Significant migration can affect the composition of the resulting genetic pool.

Introgressive capture: Process by which a specific lineage admixes with and progressively incorporates in its genome the genetic variation present in the local populations encountered as it expands from its original distribution range. Ultimately, introgressive capture can result in the total replacement of the original genetic ancestry profile.

Positive selection: Evolutionary process by which allelic variants underlying traits that are advantageous in a given environment or traits of interest to breeders increase in frequency through time.

Negative selection: Evolutionary process by which (slightly) deleterious allelic variants reducing the fitness of carrying individuals are eliminated from the population.

Mutational load: A measure of the overall impact of deleterious mutations present in any individual genome.

Close stud breeding: Rules controlling the reproduction of animals sharing similar traits and belonging to the same breed registry tracing back to the foundation stock of an individual breed.

Site Frequency Spectrum: Distribution of the allele frequencies of a given set of loci in a population.

Cost-of-domestication: Hypothesis positing that the demographic bottleneck experienced by lineages during domestication result in inflating the number of deleterious variants present in the genome, by reducing the efficacy of purifying selection.

Neural crest hypothesis: Hypothesis explaining the emergence of shared traits across various domestic animal species through selection acting on common mechanisms during development, involving the migration and differentiation of neural crest derivatives.

Summary Points list

1. After humans, horses are the animal in which the largest collection of ancient genomes have been characterized.
2. A large fraction of the diversity of lineages existing at the time of horse domestication was lost to extinction.
3. The genetic diversity of horses was considerably reduced following the growing influence of Oriental and Arabian bloodlines during the last thousand years, and the development of modern breeding practices from the 19th century AD.
4. Modern domestic horses do not descend from the earliest domestic horses known in the archaeological record and belonging to the Eneolithic Botai culture.
5. The geographic origin of modern domestic horses is presently unknown.
6. Botai horses represent the direct ancestors of the endangered modern Przewalski's horses.
7. The horse appearance and pace fluctuated significantly during history according to changing breeding preferences.

Future Issues list

1. What was the horse population structure across Eurasia during the time period spanning the early second to the fourth millennium Before Common Era? What is (are) the genetic source(s) of modern domestic horses?
2. Which role did the horse play into the large-scale migrations from steppe herders into Europe and Asia some ~4,000 to ~5,000 years ago?
3. What was the evolutionary trajectory of Botai-related horses after the Eneolithic period and until their discovery in the 19th century AD?
4. What was the exact dynamics of the spread of Oriental and Arabian bloodlines through space and time, and which phenotypic traits favored this expansion?
5. When and in which archaeological context(s) mule breeding first developed?
6. The recolonization process of the Americas by the horse during the Age of Discovery, including the development of many local breeds and feral populations, remains unclear.
7. Which fraction of the past genetic variation is present in modern breeds and/or village and/or feral horses around the world?

Sidebar: Equine ancient DNA beyond the horse

Amongst all members of the equid family, the horse currently represents by far the species most characterized at the ancient DNA level. Yet, besides the horse, there are other equine lineages for which ancient DNA fragments, and even ancient genomes, have been sequenced. In fact, the very first ancient DNA fragment ever sequenced was recovered from the now-extinct quagga zebra (35). It indicated close phylogenetic affinities with the Plains zebra, which were confirmed 30 years later once their genome was sequenced (42). Ancient DNA data have been instrumental to revise the extinction timing of multiple lineages. For example, this revealed the long-term survival into the Late Pleistocene of lineages considered extinct since the Middle Pleistocene (96, 111). They also considerably clarified the equine evolutionary tree (5, 20, 33) and often resulted in lumping together many paleontological remains previously considered as independent diverging lineages (75, 104). It is expected that

similar ancient DNA work similar to what done on horses could advance our understanding of donkey domestication and elucidate the historical contexts into which mule breeding first developed.

Disclosure Statement

The author declares no conflicts of interest.

Acknowledgements

The author would like to thank all members of the AGES team (Archaeology, Genomics, Evolution and Societies) for discussions. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 681605).

Related Resources

The IUCN Red list of threatened species, information for Przewalski's horses

(<https://www.iucnredlist.org/search?query=przewalski%27s%20horse&searchType=species>)

Food and Agriculture Organisation of the United Nations, statistical information for domestic horses

(<http://www.fao.org/livestock-systems/global-distributions/horses/en/>)DerSak