

# The Evolutionary Origin and Genetic Makeup of Domestic Horses

Pablo Librado,\* Antoine Fages,\*<sup>†</sup> Charleen Gaunitz,\* Michela Leonardi,\* Stefanie Wagner,\*<sup>‡,§</sup>  
Naveed Khan,\*<sup>\*\*\*</sup> Kristian Hanghøj,\*<sup>†</sup> Saleh A. Alquraishi,<sup>††</sup> Ahmed H. Alfarhan,<sup>††</sup> Khaled A. Al-Rasheid,<sup>††</sup>

Clio Der Sarkissian,\* Mikkel Schubert,\* and Ludovic Orlando\*<sup>†,1</sup>

\*Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350K, Denmark, <sup>†</sup>Laboratoire d'Anthropobiologie Moléculaire et d'Imagerie de Synthèse, Centre National de la Recherche Scientifique UMR 5288, Université Paul Sabatier (Toulouse III), 31000 Toulouse, France, <sup>‡</sup>UMR1202 Biodiversité Gènes et Communautés, Institut National de la Recherche Agronomique, Université de Bordeaux, F-33610 Cestas, <sup>§</sup>Biodiversité Gènes et Communautés UMR1202, Université de Bordeaux, F-33170 Talence, France, <sup>\*\*</sup>Department of Biotechnology, Abdul Wali Khan University, 23200 Mardan, Pakistan, and <sup>††</sup>Zoology Department, College of Science, King Saud University, Riyadh 11451, Saudi Arabia

**ABSTRACT** The horse was domesticated only 5.5 KYA, thousands of years after dogs, cattle, pigs, sheep, and goats. The horse nonetheless represents the domestic animal that most impacted human history; providing us with rapid transportation, which has considerably changed the speed and magnitude of the circulation of goods and people, as well as their cultures and diseases. By revolutionizing warfare and agriculture, horses also deeply influenced the politico-economic trajectory of human societies. Reciprocally, human activities have circled back on the recent evolution of the horse, by creating hundreds of domestic breeds through selective programs, while leading all wild populations to near extinction. Despite being tightly associated with humans, several aspects in the evolution of the domestic horse remain controversial. Here, we review recent advances in comparative genomics and paleogenomics that helped advance our understanding of the genetic foundation of domestic horses.

**T**HE history of the domestication of the horse remains enigmatic in several aspects due to the absence of clear morphological and osteological differences between wild and early domestic individuals, but also due to the scarcity of paleontological records from some key periods, especially the one preceding the earliest evidence of domestication. This evidence is given by the ~5500-year-old archaeological site of Botai (modern-day Kazakhstan) (Outram *et al.* 2009), at a considerable spatial and temporal distance from the Anatolian domestication centers for sheep and goats (Zeder *et al.* 2006). Unlike other ungulates, horses were not only used as a source of meat and milk, but their stamina and speed also revolutionized warfare and transportation. This also promoted cultural exchange, including the spread of Indo-European languages, religions, science, and art (Kelekna 2009; Anthony 2010). With the introduction of the horse

collar and horseshoes in agriculture, the horse was increasingly used for tilling soils, incrementing farmland productivity in medieval Europe, and remains today an essential asset to the agriculture of the least-developed countries.

Human activities have conversely influenced, directly or indirectly, the evolution of horses, causing a drastic reduction of truly wild populations. After the extinction of the Tarpan horse in 1909, which populated Eastern Europe a few centuries ago, the only surviving wild relative is the endangered Przewalski's horse. The latter was described in the Asian steppes in the 1870s, and overhunted to such an extent that it was, not >90 years later, officially declared extinct in the wild by the International Union for Conservation Nature. The Przewalski's horse survived in captivity due to successful conservation programs, which raised a stock of 12–16 captive founders to >2000 individuals, one-quarter of which are in Mongolian and Chinese reintroduction reserves (King *et al.* 2015). Once “extinct in the wild,” its current conservation status has been upgraded to “endangered.” In parallel to the extinction of wild populations, human-driven management, in particular through selection, has dramatically influenced the recent history of domestic horses, developing

Copyright © 2016 by the Genetics Society of America  
doi: 10.1534/genetics.116.194860

Manuscript received June 14, 2016; accepted for publication August 17, 2016.

<sup>1</sup>Corresponding author: Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5–7, 1350K Copenhagen, Denmark.

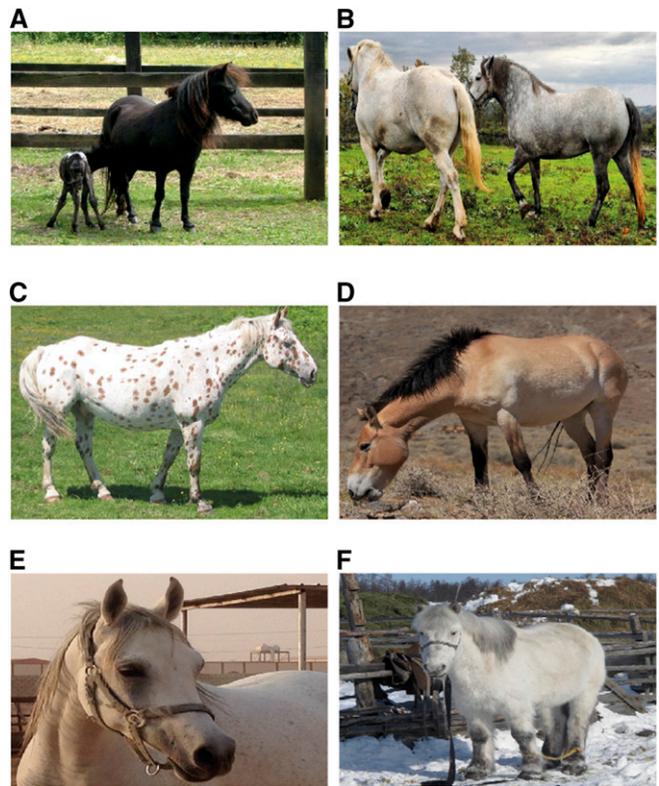
E-mail: Lorlando@snm.ku.dk

multiple breeds with a wide range of phenotypic peculiarities. Although some horse breeds, such as the Thoroughbred racing horses, are still extremely popular, a significant part of this great diversity is currently endangered. According to the Food and Agricultural Organization of the United Nations (FAO 2015), 87 horse breeds are already extinct and among the remaining 905, almost a quarter are categorized as at risk.

The population structure resulting from selective breeding is characterized by high interbreed and low intrabreed genetic diversity (McCue *et al.* 2012), and reflected by a huge array of morphological and behavioral traits (Figure 1). The height at withers, for example, extends from 70 cm in miniature Falabella horses to over 2 m in Shire and Percheron horses; an intraspecific range that is only exceeded by height variation in domestic dogs (Brooks *et al.* 2010). Domestic horses also exhibit striking variation in coat coloration, including the bay or bay-dun wild-type phenotypes, other basic colors like chestnut and black, as well as dilution (*e.g.*, cream and silver), and spotting patterns (*e.g.*, leopard complex, tobiano, and sabino) (Ludwig *et al.* 2009). Horse locomotion has also been recurrently selected, including their ability to perform alternate gaits, such as four-beat, lateral, or diagonal ambling. These alternate gaits come in addition to the three natural gaits (walk, trot, and gallop), and are known to increase the comfort of the rider and positively influence racing performance (Andersson *et al.* 2012; Promerová *et al.* 2014). Due to pleiotropic and/or epistatic effects, some of the traits selected in domestic breeds are, however, indirectly associated with congenital diseases (Bellone *et al.* 2008; McCue *et al.* 2008; Sandmeyer *et al.* 2012). These undesirable associations can be magnified by the extensive level of linkage disequilibrium (LD) that results from the low effective population size ( $N_e$ ) within breeds.

Despite the loss of DNA diversity in wild relatives, many of the phenotypic traits and congenital diseases found in breeds of major economic impact have been successfully mapped into particular genomic regions (Bellone *et al.* 2010; Andersson *et al.* 2012; Makvandi-Nejad *et al.* 2012). This has been facilitated by the inherent structure between modern breeds, but also by recent methodological advances in horse genomics. The three major milestones in horse genomics include the relatively recent sequencing and assembly of a reference genome, which was generated from the Thoroughbred mare Twilight (Wade *et al.* 2009); the development of dedicated DNA-hybridization microarrays (McCue *et al.* 2012; Petersen *et al.* 2013a,b), now targeting up to 670K single nucleotide polymorphisms (SNPs) scattered across the entire genome; and, more recently, the DNA sequencing from fossil remains, which has commenced uncovering the genetic diversity present in extinct populations of wild horses and ancient domestic animals (Box 1).

Here, we review the recent evolution of the horse lineage, with a main focus on the specificities of its domestication process, including the recent demographic history and the genetic basis underlying the domestication and makeup of modern domestic breeds.



**Figure 1** Diversity of breed phenotypes (size, shapes, and coat colors). (A) Falabella (image: E. H. Eckholdt, Wikimedia Commons). (B) Percheron horses (image: Carl Wycoff, Wikimedia Commons). (C) Appaloosa with LP coat (image: Jean-Pol Grandmont and Kersti Nebelsiek, Wikimedia Commons). (D) Przewalski's horse (image: Ludovic Orlando at Seer, one of the Mongolian reintroduction reserves). (E) The Arabian horse (image: Ludovic Orlando at Riyadh, Saudi Arabia). (F) Yakutian horse [image: Morgane Gibert, Centre National de la Recherche Scientifique (CNRS) UMR Anthropobiologie Moléculaire et Imagerie de Synthèse, France. Copyright Morgane Gibert-CNRS-Mountain Areas Farmer Support Organization].

## Predomestication Times

The equid family emerged some 55 MYA in North America, where it further diversified into several genera, including leaf browsers, grazers, and mixed feeders. It later expanded into South America and the Old World, following a complex radiation process (MacFadden 2005). Currently, the equid family is represented merely by the *Equus* genus, which—depending on the taxonomic classification considered—is comprised of seven to nine species, including zebras, asses, donkeys, and horses. Most of the past diversity is thus extinct, especially in the Americas, which experienced a mass extinction of large mammal species (“megafauna”) during the transition from the Late Quaternary to the Holocene, some 11.7 KYA (Burney and Flannery 2005; Faith and Surovell 2009). The underlying circumstances causing this extinction are highly debated, with the proposed drivers including a meteorite impact ~12.9 KYA (Firestone *et al.* 2007), a series of anthropogenic factors (Alroy 2001; Koch and Barnosky 2006), and climate changes (Owen-Smith 1987; Guthrie 2006; Koch and Barnosky 2006).

## Box 1. Ancient DNA: From Small Mitochondrial Fragments to Complete Nuclear Genomes

The history of ancient DNA research is intimately linked to the equid family since the sequencing of short mitochondrial fragments from the extinct quagga zebra, which was achieved from DNA molecules preserved in the tissues of a museum specimen prepared in 1883 (Higuchi *et al.* 1984). After 30 years, the whole nuclear genome of this species has been characterized to an average coverage of depth of approximately eightfold, which confirmed the close genetic affinities with plains zebras from Southern Africa, while unveiling genomic loci underlying their unique genetic makeup (Jónsson *et al.* 2014). The development of innovative molecular methods (for a review, see Orlando *et al.* 2015) has clearly facilitated the transition from the characterization of short sequences to whole-genome sequencing, improving our ability (i) to extract ultrashort and damaged DNA fragments from archaeological material (Dabney *et al.* 2013; Gamba *et al.* 2016), (ii) to construct DNA libraries encompassing the whole extract complexity (Meyer *et al.* 2012; Gansauge and Meyer 2013), and (iii) to even preferentially enrich DNA libraries for authentic ancient templates (Carpenter *et al.* 2013; Fu *et al.* 2013; Gansauge and Meyer 2014). However, the main driver of this transition has been the advent of high-throughput DNA sequencing, which considerably increased the sensitivity and reduced both the cost and time related to the analysis of ancient DNA extracts.

Besides the quagga zebra, other members of the horse family have also attracted the attention of researchers of ancient DNA. These include the horse, but also the donkey (Kimura *et al.* 2011), and a range of extinct equine species such as the hydruntine European ass, which flourished in Central and Western Europe at least until the Iron Age (Orlando *et al.* 2006); the giant Cape zebra (Orlando *et al.* 2009); as well as North and South American specimens (Vilà *et al.* 2001; Weinstock *et al.* 2005; Vilstrup *et al.* 2013; Sarkissian *et al.* 2015). The horse itself currently holds the world record for the oldest genome ever sequenced, which was characterized from bone material preserved in the Yukon permafrost and dated to 560–780 KYA (Orlando *et al.* 2013). More recent horse genomes have been now sequenced, including two radiocarbon dated to ~43 and 16 KYA and spanning the Upper Paleolithic (Schubert *et al.* 2014), one ~5,200 years old from the Holocene (Librado *et al.* 2015), and a handful that lived within the last couple of centuries (Der Sarkissian *et al.* 2015; Librado *et al.* 2015). Many more are underway with the aim to reconstruct the history of genetic changes that have accompanied the emergence of the modern horse.

In addition to whole-genome sequencing, more targeted approaches have illuminated the process of horse domestication. Already 10 years ago, a method pioneered the genotyping of functionally or evolutionarily relevant nuclear markers from ancient specimens. It basically consisted in two-round multiplex PCR amplifications, whereby a range of short loci (~50-bp long) are first coamplified within the same tube, and then separated in different tubes following a second series of individual amplifications (Römpler *et al.* 2006). Coupled with pyrosequencing, this method provided population-wide genotype information from ancient horses along the 5,500 years of domestication, revealing coat-color modifications as an early target of selection (Ludwig *et al.* 2009) and dynamic patterns of preferences in different sociocultural contexts (Ludwig *et al.* 2015). In the near future, we expect that this approach will be superseded by target-enrichment methods coupled with high-throughput sequencing. In horses, such approaches have been applied to only a limited number of loci (Vilstrup *et al.* 2013; Sarkissian *et al.* 2015), but recent procedures enable genotyping millions of loci genome wide (Fu *et al.* 2016), and even entire chromosomes (Fu *et al.* 2013) and genomes (Carpenter *et al.* 2013). This will likely facilitate the identification of the whole series of genetic modifications that have accompanied the recent evolutionary history of this family, domestication and extinction processes included.

By sequencing ancient DNA preserved in sediment cores, Haile and colleagues revealed that humans and megafaunal species, including horses, coexisted in Alaska until at least 10.5 KYA (Haile *et al.* 2009), some 3700 years after the last appearance of horses in the fossil record. These data challenge the ~12.9 KYA meteorite impact as an extinction driver (Firestone *et al.* 2007), and imply a coexistence with humans of at least 2300 years (Rasmussen *et al.* 2014), disproving blitzkrieg overhunting as a credible extinction process for horses (Alroy 2001; Koch and Barnosky 2006). In line with abundance patterns in the horse fossil record (Orlando *et al.* 2013), further genetic analyses confirmed strong correlations between changes in the horse effective population size ( $N_e$ ) and the climate and habitat availability (*i.e.*, open grass-

lands) (Lorenzen *et al.* 2011; Orlando *et al.* 2013). These findings collectively support that the extinction of wild horses in northern America was mainly driven by a combination of climate and vegetation changes. This evidence, however, does not rule out the possibility of anthropogenic effects in Eurasia, where horse remains become increasingly rare in the fossil assemblages after the Last Glacial Maximum until the onset of domestication.

The beginning of the Holocene, some 11.7 KYA, indeed came with a reduction of open landscapes in Western Eurasia (Huntley and Webb 1988), and a modification in the distribution range of wild horses (Boyle 2006; Sommer *et al.* 2011; Bendrey 2012). Simulation-based vegetation reconstructions, integrating paleo-environmental data, showed that

Iberia and the Pontic–Caspian area were grassland steppes during the mid-Holocene (Gallimore *et al.* 2005). Interestingly, these regions still represent hotspots of genetic diversity for horses, as measured by heterozygosity levels and allelic richness at 12 autosomal microsatellites from 24 “traditional” breeds (*e.g.*, native from specific regions). This continuity in the spatial patterns of genetic diversity suggests that local populations of wild horses from Iberia and the Pontic–Caspian steppes could have contributed to some extent to the genetic pool of modern-day European breeds (Warmuth *et al.* 2011).

### Gender-Biased Contributions to Domestication

Being paternally and maternally inherited, the nonrecombining region of the Y chromosome (NRY) and the mitochondrial DNA (mtDNA) directly reflect the male and female demographic histories, respectively. Although the repetitive nature of the NRY region makes its assembly challenging, the NRY and mtDNA markers have provided important insights into the domestication process for a range of livestock and domestic species (Meadows *et al.* 2004; Götherström *et al.* 2005; Natanaelsson *et al.* 2006; Sundqvist *et al.* 2006). In horses, however, the assembly for the whole NRY region is not available yet, as the reference genome was characterized from a single mare individual (Wade *et al.* 2009). Studies targeting specific NRY regions have shown limited variation, with haplotypes differing by one mutational step at best, which indicates that only a handful of paternal lines survived until present-day in domestic horses (Lindgren *et al.* 2004; Brandariz-Fontes *et al.* 2013; Wallner *et al.* 2013; Kreutzmann *et al.* 2014; Han *et al.* 2015). By contrast, analyses initially based on the mtDNA control region (D-loop) (Lister *et al.* 1998; Vilà *et al.* 2001; Jansen *et al.* 2002; Cieslak *et al.* 2010), and more recently on complete mitochondrial genomes (Lippold *et al.* 2011a; Achilli *et al.* 2012), revealed horses as the domestic animal showing one of the largest pools of mitochondrial genetic diversity.

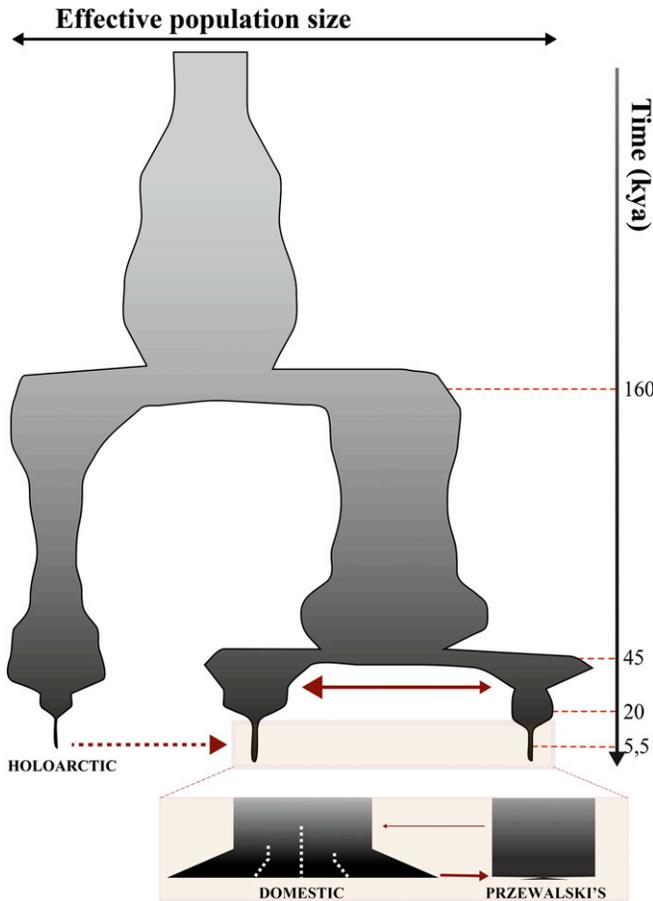
The striking difference in paternally and maternally inherited markers reflects different effective population sizes of mares and stallions. Although this could partly result from the horse polygamous mating system, the sequence diversity found in ancient wild horses suggests that current levels of variability cannot be explained without a sex-biased contribution to the domestic stock (Lippold *et al.* 2011b). More specifically, the decline in NRY diversity appears to be a domestication by-product, perhaps as a result of recent breeding programs aimed at producing valuable stallions for rural regions, as (i) the only Scythian horse sampled, dating to ~2.8 KYA, exhibits an NRY haplotype different from that found in modern individuals (Lippold *et al.* 2011b); and (ii) pedigree-based analyses can trace the most common patriline back to a few, but extremely influential, stallions that lived ~200 years ago (Wallner *et al.* 2013). Only the Yakutian breed is known to display substantial levels of NRY diversity (Librado *et al.* 2015), probably because it originated in

the 13th to 15th century, following the migration of Mongolian tribes to the Sakha Republic (Russia) (Pakendorf *et al.* 2006; Crubézy *et al.* 2010). This area has been traditionally isolated from main trade routes, such as the Silk Road, which acted as gene-flow corridors, shaping patterns of population differentiation in Eurasia (Warmuth *et al.* 2013).

The diversity patterns found in mtDNA sequences point to a completely different demographic history for mares. Present-day mitochondrial haplogroups are almost evenly distributed worldwide and coalesce ~93–160 KYA (Lippold *et al.* 2011a; Achilli *et al.* 2012; Schubert *et al.* 2014; Der Sarkissian *et al.* 2015; Librado *et al.* 2015), a time that not only predates the earliest archaeological evidence of domestication, but also the expansion of anatomically modern humans out of Africa. These studies estimated that a minimum of 17 (Achilli *et al.* 2012) and 46 (Lippold *et al.* 2011a) maternal lines successfully passed and survived into the domestic gene pool, with the latter estimate representing 73% of the mtDNA haplotypes that were already segregating prior to domestication (Lippold *et al.* 2011a). This proportion is higher than that inferred from partial mtDNA hypervariable sequences (~34%) (Cieslak *et al.* 2010).

This implies that horse domestication involved a massive incorporation of maternal lines, possibly through recurrent restocking of wild mares during the spread of horse husbandry. This extensive introgression from the wild opens intriguing questions about the mechanisms maintaining phenotypic traits desirable in domestic horses. It is known, for example, that the Przewalski’s horse can hybridize with domestic horses, and produce a fully viable progeny, despite having an extra chromosome pair ( $2n = 66$  vs.  $2n = 64$  in domestic horses). A recent study sequenced the genomes of 11 recent and 5 historical Przewalski’s horses, and compared them to the hitherto most extensive genome data set for domestic horses (Der Sarkissian *et al.* 2015). This study confirmed three phases of secondary contacts between the ancestral populations of Przewalski’s and domestic horses after their divergence ~45 KYA (Figure 2). The first phase lasted probably until the Last Glacial Maximum and maintained these populations connected by reciprocal gene flow. The second phase involved a drastic reduction of gene flow, which consisted only of an input of the ancestral population of Przewalski’s horses into that of domestic horses. The extent of gene flow was apparently not reduced following domestication. However, an additional signature of reverse gene flow could be detected from domestic horses into Przewalski’s horses, in the beginning of the 20th century, right at the foundation of the captive stock of Przewalski’s horses (Der Sarkissian *et al.* 2015).

In addition to Przewalski’s and domestic horses, complete genome sequencing of ancient horses recently revealed the existence of a third divergent lineage corresponding to a wild population that inhabited the Holarctic region (Schubert *et al.* 2014; Librado *et al.* 2015). This population is currently known from the genomes of only three fossil specimens but survived at least until ~5.2 KYA, a period contemporary with



**Figure 2** Demographic model summarizing the recent evolution of the horse. Red arrows depict secondary contact events, with their size representing the magnitude of gene flow. The area of the shaded surface is proportional to the effective population size over time, as inferred with the pairwise sequentially Markov coalescent program (Li and Durbin 2011). The dashed arrow from the Holoarctic to the domestic population indicates that the date for the gene-flow event represented is currently unknown. The last 18K years are zoomed in to illustrate recent evolutionary processes, such as the bottleneck experienced by the Przewalski's horses, and the population structure and/or expansion observed since the horse domestication (~5.5 KYA).

the early stages of the domestication process. Schubert *et al.* (2014) found that this population significantly contributed to the genetic makeup of domestic horses, with at least 12.9% of modern domestic genomes showing ancestry to this now-extinct lineage (Figure 2). Future work is required to delineate the geographic and temporal limits of this lineage, date the admixture event(s), identify the genomic blocks introgressed in domestic horses, and test which blocks show signatures of positive selection.

### Domestication Centers and Geographic Structure

Phylogeographic studies have first exploited the patterns of mtDNA variation across space, and eventually across time, to successfully date and identify the geographic centers of domestication for a wide range of livestock species (reviewed in Bruford *et al.* 2003; Groeneveld *et al.* 2010). In horses, how-

ever, early surveys of the hypervariable control region showed a minimal correspondence between mtDNA haplogroups, geographic areas, and breed types (Vilà *et al.* 2001; Jansen *et al.* 2002; Cieslak *et al.* 2010). This lack of phylogeographic structure was not an artifact resulting from the limited phylogenetic information present in the hypervariable control region, since it could be further confirmed by independent studies based on complete mitochondrial genomes (Lippold *et al.* 2011a; Achilli *et al.* 2012).

The reasons for an absence of phylogeographic mitochondrial structure are manifold, and include complex interactions between long-distance dispersal rates, extensive restocking from the wild, as well as recent human management. According to mtDNA-based studies leveraging on fossil remains, the ancient patterns of isolation by distance were characterized by an almost panmictic population from the Late Pleistocene until Early Holocene and the Copper Age (Cieslak *et al.* 2010). Only at that time, the horse population started to exhibit a certain degree of substructure within the Eurasian steppes and the Iberian Peninsula (Cieslak *et al.* 2010). This ancient substructure is still faint but detectable today from autosomal microsatellites (Warmuth *et al.* 2011), but recent selective breeding likely contributed to erode the corresponding signature at the mtDNA level. This is indirectly reflected by the Bayesian-skyline-plot (BSP) reconstructions (Drummond *et al.* 2005), which show an almost constant effective population size over time, until the domestication started, and led to a continuous expansion (Lippold *et al.* 2011a; Achilli *et al.* 2012). Although this BSP profile is generally taken as the mark of demographic expansion postdomestication, it can equally well result from violating the random mating assumption made by the BSP model. Selective breeding indeed introduces reproductive islets that impede recent coalescent events, which might mimic the genealogy of an expanding population (Ho and Shapiro 2011; Heller *et al.* 2013).

In contrast to mtDNA, nuclear data display weak but significant patterns of isolation by distance in nonbreed horses (*i.e.*, remote animals maintained outside of stud farms). Fitting explicit stepping-stone scenarios of horse radiation to 26 microsatellite loci genotyped on 322 nonbreed animals, Warmuth *et al.* (2012) found strongest support for domestication starting in the western Eurasian steppes (Warmuth *et al.* 2012), in agreement with the earliest archaeological evidence (Outram *et al.* 2009). The discovery of a second hotspot of microsatellite diversity among Iberian breeds recently revived the question of Iberia as a possible second domestication center (Warmuth *et al.* 2011). It has been proposed, for instance, that the mtDNA haplogroup D1, which is the most-commonly found among Iberian and North-African barb horses, reflects local domestication in the Iberian Peninsula (Jansen *et al.* 2002). However, the molecular analysis of 22 ancient horses failed to detect the haplogroup D1 prior to the medieval period. The limited antiquity of this haplotype within Iberia is consistent with its star-like genealogy, suggestive of an expansion in recent times, which rules out this haplogroup as a marker of a local domestication (Lira

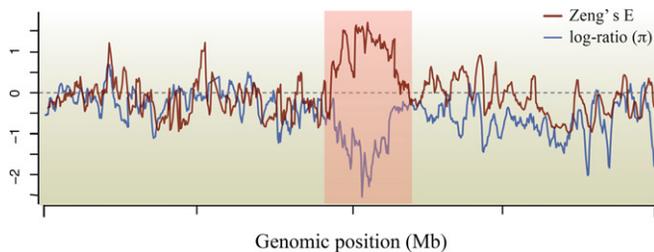
*et al.* 2010). Furthermore, the horse domestication that occurred in the Eurasian steppes was accompanied by an explosion of alleles involved in coat-color variation, which was not observed in Iberia until medieval times (Ludwig *et al.* 2009). Yet, the so-called Lusitano haplogroup C, which is currently restricted to horses native from Portugal, was found to be already present in the Neolithic and Bronze Ages (Lira *et al.* 2010), which leaves the possibility of a secondary domestication in Iberia open and controversial.

## The Genetic Makeup of Domestic Horses

During the domestication of the horse, humans acted as strong selective forces by favoring particular traits of interest, and inadvertently through the use of animals for various tasks far outside the range of their normal behavior. This resulted in a specialization of horses for strength, aesthetics, racing performance, or endurance. Several lines of evidence suggest selection on standing genetic variation as a major component of this process.

### Early Selection Targets

Both prehistoric cave paintings and ancient DNA data have corroborated that some coat-color variants, including black and bay, as well as the leopard complex spotting (LP), were already segregating in wild populations (Ludwig *et al.* 2009; Pruvost *et al.* 2011). The latter is particularly noteworthy since individuals homozygous for the LP allele show congenital stationary night blindness (Bellone *et al.* 2008, 2010; Sandmeyer *et al.* 2012). Despite its detrimental effect, the LP allele appears to have been relatively common in early domestic horses, based on its discovery in 6 out of the 10 samples from the Kirklareli–Kanligecit archaeological site (Turkish Thrace, dating 4.2–4.7 KYA). The LP allele thereafter remained at undetectable frequencies until the early Iron Age, where it was identified again in a 3300–3400-year-old specimen from the Chicha archaeological site (western Siberia). By using the Approximate Bayesian Computation framework (Beaumont *et al.* 2002; Csilléry *et al.* 2010), the authors estimated that this nonmonotonic variation of the LP



**Figure 3** Genome-wide scan of positive selection. The shaded area showcases a genomic region that has potentially undergone a recent episode of positive selection, as the Zeng's  $E$  statistic (Zeng *et al.* 2006) shows an excess of high-frequency derived variants, and the levels of nucleotide diversity ( $\pi$ ) are lower in present-day than in a hypothetical wild ancestor (*i.e.*, the log ratio of their respective  $\pi$  values is below zero and low over a genomic block of significant size).

frequency cannot be solely explained by genetic drift, but must have involved fluctuating selection, first favoring the LP allele during the early Bronze Age, but later counterselecting it during the middle Bronze Age (Ludwig *et al.* 2015). This study highlights dynamic preferences for horse traits and types in past human groups from different social and cultural contexts. In addition to the LP allele, another example of selection targeting standing variation is provided by a 1617-bp deletion on chromosome 8, which prevents the development of a wild-type dun coloration (Imsland *et al.* 2016). Notably, this deletion has been documented in a ~43K-year-old horse from the Taymyr Peninsula in Central Siberia (Imsland *et al.* 2016). Overall, current data thus indicates that a relatively wide range of coat-color variants existed in Eurasian landscapes, long before the earliest evidence of domestication.

Such studies have clearly advanced our understanding of the recent horse evolution, especially with regard to their domestication. They were, however, limited in scope as they targeted only a handful of candidate loci. The whole-genome sequencing of ancient predomestic horses (Box 1) has instead allowed for genome-wide scans of positive selection (Figure 3), pinpointing the broader range of genetic changes associated with horse domestication. Schubert and colleagues implemented this approach for the first time by sequencing the complete genomes of two prehistoric horses dating back to ~43 and 16.5 KYA, and comparing them with a range of genomes spanning a wide panel of domestic breeds (Schubert *et al.* 2014). This provided a set of 125 candidate genes that underwent episodes of positive selection following horse domestication, including some that were previously identified using different approaches, such as *MC1R* (involved in coat-coloration patterns). It also revealed a majority of novel candidates involved in the (i) cardiac and circulatory system, probably reflecting increasing energetic demands related to the extensive horse usage in transportation and locomotion; (ii) bone, limb, and face morphogenesis in relation to the morphological changes documented along the domestication history; and (iii) brain development, neuron growth, and behavior, which likely contributed to the cognitive and social changes associated with domestication.

Schubert and colleagues also estimated that the genome of domestic horses harbors an excess of variants in genomic positions that are highly constrained across mammals, and are thus likely to be deleterious. This accumulation of detrimental variants is a side effect of human-driven selective breeding. The increased variance in reproductive success, especially in stallions, yields a reduction in the effective population size, which compromises the effectiveness of natural selection to purge deleterious mutations (for a review, see Charlesworth 2009). This is in line with evidence observed from other species, such as dogs (Cruz *et al.* 2008; Marsden *et al.* 2016), rice (Lu *et al.* 2006), and tomatoes (Koenig *et al.* 2013); indicating that human preferences have not only favored certain variants during domestication, but also introduced an important genomic cost represented by an

accumulation of deleterious mutations, which increased the propensity of modern domestic breeds to develop genetic disorders.

### Engendering Modern Breeds

With a few notable exceptions, such as the Arabian, Mongolian, and Icelandic horses, breeds are relatively recent human constructs on an evolutionary timescale. The earliest horse studbook, that of the Thoroughbred racing horses, was created in 1791. Therefore, within the last two centuries, humans have imposed strong diverging selection among breeds. This low within-breed variability, and thus relatively high LD patterns, have made the ~54K SNPs targeted by the EquineSNP50 BeadChip extremely useful for horse genetic research (McCue *et al.* 2012). More recently, the Equine Genome Diversity Consortium exploited the complete genome sequences from 20 breeds to develop an Affymetrix array covering ~670K SNPs scattered along the horse genome. In addition to its higher SNP density, this array corrects for known limitations of the EquineSNP50 BeadChip system, such as the ascertainment bias toward Thoroughbred horses (which were overrepresented in the original SNP discovery panel) and the absence of X-chromosome markers. As the Axiom Equine Genotyping microarray (Axiom MNEC670) was made commercially available only since 2015, the EquineSNP50 BeadChip has been the most popular system in horse genomics thus far, and has been applied to address a full range of biological questions. It was, for instance, used to identify genomic blocks that are highly differentiated in particular breeds, as a proxy for understanding the genetic basis of their phenotypic peculiarities (Petersen *et al.* 2013a). Remarkably, the results revealed noncoding regions as breed-specific targets of selection, in accordance with selection acting both on the coding sequence of genes, and on the regulation of their expression. Putatively selected haplotypes encompassed genes previously proposed to be involved in gait (*DMRT2* and *DMRT3*), coat color (*MC1R*), performance (*MSTN*), and size (*IGF1*, *NCAPG*, and *HMG2A2*) (Petersen *et al.* 2013a).

Beyond representing an interesting list of selected candidates, the phenotypic consequences of these mutations were further demonstrated in a range of functional studies. For example, three noncoding polymorphisms around *MSTN* have been associated with racing capabilities in elite and common Thoroughbreds: (i) the g66493797T > C SNP at its first intron (Hill *et al.* 2010a,b; Tozaki *et al.* 2010), (ii) the insertion of a 227-bp-long short interspersed nuclear element (SINE) at its promoter (Hill *et al.* 2010a), and (iii) the BIEC2-417495 SNP located 692 kb upstream of *MSTN* and 30 kb upstream of the glutaminase gene *GLS* (Binns *et al.* 2010). In elite Thoroughbreds, homozygotes for the g66493797 C allele performed better in short (<1300 m) and fast races, heterozygotes in middle distance races (between 1301 and 1900 m), and homozygotes for the g66493797 T allele in long distance races (Hill *et al.* 2010a,b). A correlation

between genotypes at the g66493797 locus and expression was established in untrained Thoroughbreds, with *MSTN* messenger RNA levels decreasing from g66493797 C/C to g66493797 C/T and g66493797 T/T; whereas in trained horses, reduced but comparable *MSTN* expression was observed across all genotypes (McGivney *et al.* 2012). Moreover, and although the precise mechanisms of transcription regulation are presently unknown, the 227-bp SINE insertion was predicted *in silico* to create a CpG island, as well as few novel putative transcription factor binding sites, which may play a role in regulating the *MSTN* expression (van den Hoven *et al.* 2015).

The nonsense mutation in the *DMRT3* transcription factor underpins alternate gaits in horses, such as ambling and tölt (Andersson *et al.* 2012). This mutation, named “gait keeper,” segregates at high frequencies in breeds classified as gaited or bred for harness racing; in contrast to horses lacking this ability which show only marginal allele frequencies (Petersen *et al.* 2013a; Promerová *et al.* 2014). In Finnhorses and Icelandic horses, which can perform alternate gaits, the gait-keeper mutation was found to be beneficial only for homozygous racehorses. When used in other classical riding disciplines, such as jumping or dressage, homozygous horses obtained lower scores than heterozygous or nongait-keeper homozygous animals (Andersson *et al.* 2012; Kristjansson *et al.* 2014; Jäderkvist Fegraeus *et al.* 2015). Different *DMRT3* genotypes thus appear to benefit different types of performance (harness racing vs. classical riding), and both gait-keeper and nongait-keeper alleles segregate at relatively similar frequencies (Jäderkvist Fegraeus *et al.* 2015).

Subtle molecular signals of positive selection, reflecting complex selective regimes, including context-dependent fitness and pleiotropic effects, have started to be detected. One such case is provided by the gain-of-function mutation at the glycogen-synthase *GYS1* gene, which increases the accumulation of glycogen in skeletal muscles and facilitates a better restoration of glycogen postexercise (McCue *et al.* 2008). The *GYS1* mutation is, however, also associated with the postexercise equine polysaccharide storage myopathy type 1, a disease causing a breakdown of type 2A and 2B muscle fibers (rhabdomyolysis) (McCue *et al.* 2008). Such deleterious effects should have purged out the *GYS1* mutation from modern livestock, but it is still found at considerable frequencies in breeds of heavy horses and with long breeding histories, such as the Belgian draft horses (Baird *et al.* 2010; McCue *et al.* 2010). To explain this apparent paradox, the *GYS1* mutation has been proposed to represent a “thrifty” allele, which was once beneficial in the past when conditions of regular hard work and limited nutritional input could favor genotypes leading to a more efficient storage of energy sources in the muscles. This allele would be maladaptive in modern management conditions of unlimited feeding resources and relatively limited exercise. In line with this model, high frequency, LD, and extended haplotype homozygosity patterns found in Belgian heavy draft horses in association with the *GYS1* mutation are unlikely to be explained under

neutrality (McCoy *et al.* 2014). The mutation in *GYS1* might thus have undergone selection in the past, but the relaxed constraints prevailing today, including a limited population size, might ultimately lead to the disappearance of this allele in this breed (McCoy *et al.* 2014). It is noteworthy that in such cases of fluctuating selection, ancient DNA can help recover full time series of allele frequencies (Ludwig *et al.* 2015) and better characterize the dynamics of underlying selective regimes (Malaspinas *et al.* 2012; Schraiber *et al.* 2016).

Controlling for the underlying demographic process is important to accurately pinpoint the targets of natural selection. For example, Yakutian horses, which survive the most extreme winter temperatures in the Northern hemisphere, developed their striking cold adaptations in only ~100 generations (Librado *et al.* 2015). The cross-comparison of the variation present in their genome and that from non-Artic horse breeds revealed that the few kilobases located upstream of translation start sites harbor the highest proportion of adaptive mutations (Librado *et al.* 2015). Such *cis*-regulatory candidates potentially drive the expression of genes participating in adaptive phenotypic features, such as hair density. Numerous genes involved in glucose metabolism indicated that sugar-related antifreezing properties and the ability to regulate seasonal thermogenic requirements might also be essential to survive the Arctic cold. Interestingly, the biological significance of some of these candidates was corroborated in woolly mammoths (Lynch *et al.* 2015), where parallel episodes of positive selection have been described at the *BARX2* and *PHIP* genes, associated with hair development and insulin metabolism, respectively. Similarly, human groups from Siberia (Cardona *et al.* 2014) also show adaptive footprints at the *PRKG1* gene, which participates in shivering by regulating blood vessel constriction. Altogether, such patterns of convergent evolution support regulatory changes as a key mechanism for driving rapid adaptive processes, mainly because these regions offer an important fraction of the segregating variation readily available for natural selection. Considering the equally short evolutionary timescale related to the formation of modern horse breeds, the role of noncoding regions in modern horse phenotypes will thus likely receive a lot of attention in the near future, especially in light of the Functional Annotation of Animal Genomes (FAANG) consortium, which aims at cataloging all the “functional” elements, protein coding or not, present in the horse genome (Andersson *et al.* 2015).

## Conclusions

In this review, we have shown that the horse domestication process was complex and involved continuous genetic restocking from the wild, in a sex-biased manner, mostly from mares (Vilà *et al.* 2001; Jansen *et al.* 2002; Achilli *et al.* 2012; Warmuth *et al.* 2012). The absence of genetic isolation between domestic livestock and wild ancestors observed in the horse does not stand as an exception though, but is increasingly recognized in other domestic animals, such as

cattle (Park *et al.* 2015) and pigs (Frantz *et al.* 2015). Understanding how distinct phenotypic features can be maintained in the presence of homogenizing gene flow thus seems to become one central question common to most, if not all, animal domestication processes. Developing comparative genomics approaches such as those aimed at identifying genomic islands of speciation (Turner *et al.* 2005)—the process underlying species formation and ultimately genetic isolation—will likely help solve this nascent paradox in the near future.

In addition to revealing the true extent of gene flow, in particular from yet-unrecognized and extinct wild lineages (Schubert *et al.* 2014), ancient DNA has also showed that the domestication process was quite dynamic and uneven through space and time, as particular human groups selected different phenotypic traits (Ludwig *et al.* 2015). This suggests that the history of changes that accompanied the early domestication of horses and their further transformation in the course of history will be difficult, if not impossible, to reconstruct from current patterns of genetic diversity. Indeed, the latter mostly reflect the recent history of intensive selection and extensive admixture that followed the development of breeds since the 18th century. As such, it pleads for increased ancient genome surveys of horse remains spread across the whole temporal and geographic range where the human–horse relationship developed. We expect that similar investigations will also largely advance our understanding of domestication processes when applied to other domestic animals. This work has already started in cattle (Edwards *et al.* 2007; Bollongino *et al.* 2008; Orlando 2015; Park *et al.* 2015; Scheu *et al.* 2015), dogs (Ollivier *et al.* 2013; Thalmann *et al.* 2013), swine (Edwards *et al.* 2007; Meiri *et al.* 2013; Ottoni *et al.* 2013; Frantz *et al.* 2015; Ramírez *et al.* 2015), and chickens (Flink *et al.* 2014).

Recent work based on ancient DNA has revealed that, beyond genomes, ancient genome-wide methylation and nucleosome maps can be reconstructed using patterns of DNA degradation along the genome (Gokhman *et al.* 2014, 2016; Orlando and Willerslev 2014; Orlando *et al.* 2015; Pedersen *et al.* 2014). Combined with current efforts of the FAANG consortium (Andersson *et al.* 2015), such approaches promise to reveal the true extent of past individual plasticity and how patterns of gene regulation were modified in response to new domestication/selection targets. Additionally, the recent discovery that dental calculus represents a fantastic source of ancient microbial DNA (Adler *et al.* 2013) and genetic traces of major components of the diet (Warinner *et al.* 2014), encourages the genetic investigation of dietary changes and modification of the oral microbiota that accompanied horse domestication. We anticipate that, together with ancient genome and epigenome reconstruction, these approaches will reveal the true extent of biological transformation that forged the modern horse.

Finally, as informative as the genetic analysis of fossils can be, it remains that a growing number of domestic breeds are currently endangered. It is thus important to undertake

massive genetic efforts aimed at their conservation both as a unique biological heritage of traditional human activities and as an invaluable resource of variation for future selection programs.

## Acknowledgments

This work is dedicated to the memory of Dr. Teri Lear, a fantastic scientist, friend and horse lover, who recently passed away. This work was supported by the Danish Council for Independent Research, Natural Sciences (Grant 4002-00152B); the Danish National Research Foundation (Grant DNR94); Initiative d'Excellence Chaires d'attractivité, Université de Toulouse (OURASI); the International Research Group Program (Grant IRG14-08), Deanship of Scientific Research, King Saud University; a Marie-Curie Individual Fellowship (MSCA-IF-657852); a Villum Fonden Blokstipendier grant, and; a Villum Fonden research project grant (miGENEPI).

## Literature Cited

- Achilli, A., A. Olivieri, P. Soares, H. Lancioni, B. Hooshiar Kashani *et al.*, 2012 Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proc. Natl. Acad. Sci. USA* 109: 2449–2454.
- Adler, C. J., K. Dobney, L. S. Weyrich, J. Kaidonis, A. W. Walker *et al.*, 2013 Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat. Genet.* 45: 450–455.
- Alroy, J., 2001 A Multispecies Overkill Simulation of the End-Pleistocene Megafaunal Mass Extinction. *Science* 292: 1893–1896.
- Andersson, L. S., M. Larhammar, F. Memic, H. Wootz, D. Schwochow *et al.*, 2012 Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. *Nature* 488: 642–646.
- Andersson, L., A. L. Archibald, C. D. Bottema, R. Brauning, S. C. Burgess *et al.*, 2015 Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biol.* 16: 57.
- Anthony, D. W., 2010 *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*. Princeton University Press, Princeton, NJ.
- Baird, J. D., S. J. Valberg, S. M. Anderson, M. E. McCue, and J. R. Mickelson, 2010 Presence of the glycogen synthase 1 (GYS1) mutation causing type 1 polysaccharide storage myopathy in continental European draught horse breeds. *Vet. Rec.* 167: 781–784.
- Beaumont, M. A., W. Zhang, and D. J. Balding, 2002 Approximate Bayesian Computation in Population Genetics. *Genetics* 162: 2025–2035.
- Bellone, R. R., S. A. Brooks, L. Sandmeyer, B. A. Murphy, G. Forsyth *et al.*, 2008 Differential Gene Expression of TRPM1, the Potential Cause of Congenital Stationary Night Blindness and Coat Spotting Patterns (LP) in the Appaloosa Horse (*Equus caballus*). *Genetics* 179: 1861–1870.
- Bellone, R. R., G. Forsyth, T. Leeb, S. Archer, S. Sigurdsson *et al.*, 2010 Fine-mapping and mutation analysis of TRPM1: a candidate gene for leopard complex (LP) spotting and congenital stationary night blindness in horses. *Brief. Funct. Genomics* 9: 193–207.
- Bendrey, R., 2012 From wild horses to domestic horses: a European perspective. *World Archaeol.* 44: 135–157.
- Binns, M. M., D. A. Boehler, and D. H. Lambert, 2010 Identification of the myostatin locus (MSTN) as having a major effect on optimum racing distance in the Thoroughbred horse in the USA. *Anim. Genet.* 41: 154–158.
- Bollongino, R., J. Elsner, J.-D. Vigne, and J. Burger, 2008 Y-SNPs Do Not Indicate Hybridisation between European Aurochs and Domestic Cattle. *PLoS One* 3: e3418.
- Boyle, K. V., 2006 Neolithic wild game animals in Western Europe: the question of hunting, pp. 10–23 in *Animals in the Neolithic of Britain and Europe*, edited by D. Serjeantson and D. Field. Oxbow Books, Oxford.
- Brandariz-Fontes, C., J. A. Leonard, J. L. Vega-Pla, N. Backström, G. Lindgren *et al.*, 2013 Y-Chromosome Analysis in Retuertass Horses. *PLoS One* 8: e64985.
- Brooks, S. A., S. Makvandi-Nejad, E. Chu, J. J. Allen, C. Streeter *et al.*, 2010 Morphological variation in the horse: defining complex traits of body size and shape. *Anim. Genet.* 41: 159–165.
- Bruford, M. W., D. G. Bradley, and G. Luikart, 2003 DNA markers reveal the complexity of livestock domestication. *Nat. Rev. Genet.* 4: 900–910.
- Burney, D. A., and T. F. Flannery, 2005 Fifty millennia of catastrophic extinctions after human contact. *Trends Ecol. Evol.* 20: 395–401.
- Cardona, A., L. Pagani, T. Antao, D. J. Lawson, C. A. Eichstaedt *et al.*, 2014 Genome-Wide Analysis of Cold Adaptation in Indigenous Siberian Populations. *PLoS One* 9: e98076.
- Carpenter, M. L., J. D. Buenrostro, C. Valdiosera, H. Schroeder, M. E. Allentoft *et al.*, 2013 Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. *Am. J. Hum. Genet.* 93: 852–864.
- Charlesworth, B., 2009 Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* 10: 195–205.
- Cieslak, M., M. Pruvost, N. Benecke, M. Hofreiter, A. Morales *et al.*, 2010 Origin and History of Mitochondrial DNA Lineages in Domestic Horses. *PLoS One* 5: e15311.
- Crubézy, E., S. Amory, C. Keyser, C. Bouakaze, M. Bodner *et al.*, 2010 Human evolution in Siberia: from frozen bodies to ancient DNA. *BMC Evol. Biol.* 10: 25.
- Cruz, F., C. Vilà, and M. T. Webster, 2008 The legacy of domestication: accumulation of deleterious mutations in the dog genome. *Mol. Biol. Evol.* 25: 2331–2336.
- Csilléry, K., M. G. B. Blum, O. E. Gaggiotti, and O. François, 2010 Approximate Bayesian Computation (ABC) in practice. *Trends Ecol. Evol.* 25: 410–418.
- Dabney, J., M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann *et al.*, 2013 Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. USA* 110: 15758–15763.
- Der Sarkissian, C., L. Ermini, M. Schubert, M. A. Yang, P. Librado *et al.*, 2015 Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. *Curr. Biol.* 25: 2577–2583.
- Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus, 2005 Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. *Mol. Biol. Evol.* 22: 1185–1192.
- Edwards, C. J., R. Bollongino, A. Scheu, A. Chamberlain, A. Tresset *et al.*, 2007 Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. *Proc. R. Soc. Lond. B Biol. Sci.* 274: 1377–1385.
- Faith, J. T., and T. A. Surovell, 2009 Synchronous extinction of North America's Pleistocene mammals. *Proc. Natl. Acad. Sci. USA* 106: 20641–20645.
- FAO, 2015 *The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture*, edited by B. D. Scherf

- and D. Pilling. FAO Commission on Genetic Resources for Food and Agriculture Assessments, Rome. Available at: <http://www.fao.org/3/a-i4787e/index.html>.
- Firestone, R. B., A. West, J. P. Kennett, L. Becker, T. E. Bunch *et al.*, 2007 Evidence for an extraterrestrial impact 12,900 years ago that contributed to the megafaunal extinctions and the Younger Dryas cooling. *Proc. Natl. Acad. Sci. USA* 104: 16016–16021.
- Flink, L. G., R. Allen, R. Barnett, H. Malmström, J. Peters *et al.*, 2014 Establishing the validity of domestication genes using DNA from ancient chickens. *Proc. Natl. Acad. Sci. USA* 111: 6184–6189.
- Frantz, L. A. F., J. G. Schraiber, O. Madsen, H.-J. Megens, A. Cagan *et al.*, 2015 Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat. Genet.* 47: 1141–1148.
- Fu, Q., M. Meyer, X. Gao, U. Stenzel, H. A. Burbano *et al.*, 2013 DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl. Acad. Sci. USA* 110: 2223–2227.
- Fu, Q., C. Posth, M. Hajdinjak, M. Petr, S. Mallick *et al.*, 2016 The genetic history of Ice Age Europe. *Nature* 534: 200–205.
- Gallimore, R., R. Jacob, and J. Kutzbach, 2005 Coupled atmosphere-ocean-vegetation simulations for modern and mid-Holocene climates: role of extratropical vegetation cover feedbacks. *Clim. Dyn.* 25: 755–776.
- Gamba, C., K. Hanghøj, C. Gaunitz, A. H. Alfarchan, S. A. Alquraishi *et al.*, 2016 Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Mol. Ecol. Resour.* 16: 459–469.
- Gansauge, M.-T., and M. Meyer, 2013 Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* 8: 737–748.
- Gansauge M.-T., and M. Meyer, 2014 Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res.* 24: 1543–1549.
- Gokhman, D., E. Lavi, K. Prüfer, M. F. Fraga, J. A. Riancho *et al.*, 2014 Reconstructing the DNA Methylation Maps of the Neanderthal and the Denisovan. *Science* 344: 523–527.
- Gokhman, D., E. Meshorer, and L. Carmel, 2016 Epigenetics: It's Getting Old. Past Meets Future in Paleoepigenetics. *Trends Ecol. Evol.* 31: 290–300.
- Götherström, A., C. Anderung, L. Hellborg, R. Elburg, C. Smith *et al.*, 2005 Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe. *Proc. R. Soc. Lond. B Biol. Sci.* 272: 2345–2351.
- Groeneveld, L. F., J. A. Lenstra, H. Eding, M. A. Toro, B. Scherf *et al.*, 2010 Genetic diversity in farm animals—a review. *Anim. Genet.* 41: 6–31.
- Guthrie, R. D., 2006 New carbon dates link climatic change with human colonization and Pleistocene extinctions. *Nature* 441: 207–209.
- Haile, J., D. G. Froese, R. D. E. MacPhee, R. G. Roberts, L. J. Arnold *et al.*, 2009 Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proc. Natl. Acad. Sci. USA* 106: 22352–22357.
- Han, H., Q. Zhang, K. Gao, X. Yue, T. Zhang *et al.*, 2015 Y-Single Nucleotide Polymorphisms Diversity in Chinese Indigenous Horse. *Asian-Australas. J. Anim. Sci.* 28: 1066–1074.
- Heller, R., L. Chikhi, and H. R. Siegmund, 2013 The Confounding Effect of Population Structure on Bayesian Skyline Plot Inferences of Demographic History. *PLoS One* 8: e62992.
- Higuchi, R., B. Bowman, M. Freiberger, O. A. Ryder, and A. C. Wilson, 1984 DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312: 282–284.
- Hill, E. W., B. A. McGivney, J. Gu, R. Whiston, and D. E. Machugh, 2010a 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.
- Hill, E. W., J. Gu, S. S. Eivers, R. G. Fonseca, B. A. McGivney *et al.*, 2010b A sequence polymorphism in MSTN predicts sprinting ability and racing stamina in thoroughbred horses. *PLoS One* 5: e8645.
- Ho, S. Y. W., and B. Shapiro, 2011 Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol. Ecol. Resour.* 11: 423–434.
- van den Hoven, R., E. Gür, M. Schlamanig, M. Hofer, A. C. Onmaz *et al.*, 2015 Putative regulation mechanism for the MSTN gene by a CpG island generated by the SINE marker Ins227bp. *BMC Vet. Res.* 11: 138.
- Huntley, B., and T. Webb, III (Editors), 1988 *Vegetation history*, Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Immsland, F., K. McGowan, C.-J. Rubin, C. Henegar, E. Sundström *et al.*, 2016 Regulatory mutations in TBX3 disrupt asymmetric hair pigmentation that underlies Dun camouflage color in horses. *Nat. Genet.* 48: 152–158.
- Jäderkvist Fegraeus, K., L. Johansson, M. Mäenpää, A. Mykkänen, L. S. Andersson *et al.*, 2015 Different DMRT3 Genotypes Are Best Adapted for Harness Racing and Riding in Finnhorses. *J. Hered.* 106: 734–740.
- Jansen, T., P. Forster, M. A. Levine, H. Oelke, M. Hurles *et al.*, 2002 Mitochondrial DNA and the origins of the domestic horse. *Proc. Natl. Acad. Sci. USA* 99: 10905–10910.
- Jónsson, H., M. Schubert, A. Seguin-Orlando, A. Ginolhac, L. Petersen *et al.*, 2014 Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proc Natl Acad Sci USA* 111: 18655–18660.
- Kelekna, P., 2009 *The Horse in Human History*, Cambridge University Press, New York.
- King, S.R.B., L. Boyd, W. Zimmerman, and B.E. Kendall, 2015 *Equus ferus ssp. przewalskii*. The IUCN Red List of Threatened Species 2015: e.T7961A45172099. Available at: <http://dx.doi.org/10.2305/IUCN.UK.2015-2.RLTS.T7961A45172099.en>.
- Kimura, B., F. B. Marshall, S. Chen, S. Rosenbom, P. D. Moehlman *et al.*, 2011 Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication. *Proc. R. Soc. Lond. B Biol. Sci.* 278: 50–57.
- Koch, P. L., and A. D. Barnosky, 2006 Late Quaternary Extinctions: State of the Debate. *Annu. Rev. Ecol. Evol. Syst.* 37: 215–250.
- Koenig, D., J. M. Jiménez-Gómez, S. Kimura, D. Fulop, D. H. Chitwood *et al.*, 2013 Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl. Acad. Sci. USA* 110: E2655–E2662.
- Kreutzmann, N., G. Brem, and B. Wallner, 2014 The domestic horse harbours Y-chromosomal microsatellite polymorphism only on two widely distributed male lineages. *Anim. Genet.* 45: 460.
- Kristjánsson, T., S. Björnsdóttir, A. Sigurdsson, L. S. Andersson, G. Lindgren *et al.*, 2014 The effect of the “Gait keeper” mutation in the DMRT3 gene on gaiting ability in Icelandic horses. *J. Anim. Breed. Genet.* 131: 415–425.
- Li, H., and R. Durbin, 2011 Inference of human population history from individual whole-genome sequences. *Nature* 475: 493–496.
- Librado, P., C. D. Sarkissian, L. Ermini, M. Schubert, H. Jónsson *et al.*, 2015 Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proc. Natl. Acad. Sci. USA* 112: E6889–E6897.
- Lindgren, G., N. Backström, J. Swinburne, L. Hellborg, A. Einarsson *et al.*, 2004 Limited number of patrilineal lines in horse domestication. *Nat. Genet.* 36: 335–336.
- Lippold, S., N. J. Matzke, M. Reissmann, and M. Hofreiter, 2011a Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. *BMC Evol. Biol.* 11: 328.

- Lippold, S., M. Knapp, T. Kuznetsova, J. A. Leonard, N. Benecke *et al.*, 2011b Discovery of lost diversity of paternal horse lineages using ancient DNA. *Nat. Commun.* 2: 450.
- Lira, J., A. Linderholm, C. Olaria, M. Brandström Durling, M. T. P. Gilbert *et al.*, 2010 Ancient DNA reveals traces of Iberian Neolithic and Bronze Age lineages in modern Iberian horses. *Mol. Ecol.* 19: 64–78.
- Lister, A., M. Kadwell, L. M. Kaagen, M. B. Richards, and H. F. Stanley, 1998 Ancient and modern DNA in a study of horse domestication. *Anc. Biomol.* 2: 267–280.
- Lorenzen, E. D., D. Nogués-Bravo, L. Orlando, J. Weinstock, J. Binladen *et al.*, 2011 Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature* 479: 359–364.
- Lu, J., T. Tang, H. Tang, J. Huang, S. Shi *et al.*, 2006 The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends Genet.* TIG 22: 126–131.
- Ludwig, A., M. Pruvost, M. Reissmann, N. Benecke, G. A. Brockmann *et al.*, 2009 Coat Color Variation at the Beginning of Horse Domestication. *Science* 324: 485
- Ludwig, A., M. Reissmann, N. Benecke, R. Bellone, E. Sandoval-Castellanos *et al.*, 2015 Twenty-five thousand years of fluctuating selection on leopard complex spotting and congenital night blindness in horses. *Phil Trans R Soc B* 370: 20130386.
- Lynch, V. J., O. C. Bedoya-Reina, A. Ratan, M. Sulak, D. I. Drautz-Moses *et al.*, 2015 Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic. *Cell Reports* 12: 217–228.
- MacFadden, B. J., 2005 Fossil Horses—Evidence for Evolution. *Science* 307: 1728–1730.
- Makvandi-Nejad, S., G. E. Hoffman, J. J. Allen, E. Chu, E. Gu *et al.*, 2012 Four Loci Explain 83% of Size Variation in the Horse. *PLoS ONE* 7: e39929.
- Malaspinas, A.-S., O. Malaspinas, S. N. Evans, and M. Slatkin, 2012 Estimating allele age and selection coefficient from time-serial data. *Genetics* 192: 599–607.
- Marsden, C. D., D. O.-D. Vecchy, D. P. O'Brien, J. F. Taylor, O. Ramirez *et al.*, 2016 Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc. Natl. Acad. Sci. USA* 113: 152–157.
- McCoy, A. M., R. Schaefer, J. L. Petersen, P. L. Morrell, M. A. Slamka *et al.*, 2014 Evidence of Positive Selection for a Glycogen Synthase (GYS1) Mutation in Domestic Horse Populations. *J. Hered.* 105: 163–172.
- McCue, M. E., S. J. Valberg, M. B. Miller, C. Wade, S. DiMauro *et al.*, 2008 Glycogen synthase (GYS1) mutation causes a novel skeletal muscle glycogenesis. *Genomics* 91: 458–466.
- McCue, M. E., S. M. Anderson, S. J. Valberg, R. J. Piercy, S. Z. Barakzai *et al.*, 2010 Estimated prevalence of the Type 1 Polysaccharide Storage Myopathy mutation in selected North American and European breeds. *Anim. Genet.* 41: 145–149.
- McCue, M. E., D. L. Bannasch, J. L. Petersen, J. Gurr, E. Bailey *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.
- McGivney, B. A., J. A. Browne, R. G. Fonseca, L. M. Katz, D. E. MacHugh *et al.*, 2012 MSTN genotypes in Thoroughbred horses influence skeletal muscle gene expression and racetrack performance. *Anim. Genet.* 43: 810–812.
- Meadows, J. R. S., R. J. Hawken, and J. W. Kijas, 2004 Nucleotide diversity on the ovine Y chromosome. *Anim. Genet.* 35: 379–385.
- Meiri, M., D. Huchon, G. Bar-Oz, E. Boaretto, L. K. Horwitz *et al.*, 2013 Ancient DNA and Population Turnover in Southern Levantine Pigs—Signature of the Sea Peoples Migration? *Sci. Rep.* 3: 3035.
- Meyer, M., M. Kircher, M.-T. Gansauge, H. Li, F. Racimo *et al.*, 2012 A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* 338: 222–226.
- Natanaelsson, C., M. C. Oskarsson, H. Angleby, J. Lundeberg, E. Kirkness *et al.*, 2006 Dog Y chromosomal DNA sequence: identification, sequencing and SNP discovery. *BMC Genet.* 7: 45.
- Ollivier, M., A. Tresset, C. Hitte, C. Petit, S. Hughes *et al.*, 2013 Evidence of Coat Color Variation Sheds New Light on Ancient Canids. *PLoS One* 8: e75110.
- Orlando, L., 2015 The first aurochs genome reveals the breeding history of British and European cattle. *Genome Biol.* 16: 225.
- Orlando, L., and E. Willerslev, 2014 An epigenetic window into the past? *Science* 345: 511–512.
- Orlando, L., M. Mashkour, A. Burke, C. J. Douady, V. Eisenmann *et al.*, 2006 Geographic distribution of an extinct equid (*Equus hydruntinus*: Mammalia, Equidae) revealed by morphological and genetical analyses of fossils. *Mol. Ecol.* 15: 2083–2093.
- Orlando, L., J. L. Metcalf, M. T. Alberdi, M. Telles-Antunes, D. Bonjean *et al.*, 2009 Revising the recent evolutionary history of equids using ancient DNA. *Proc. Natl. Acad. Sci. USA* 106: 21754–21759.
- Orlando, L., A. Ginolhac, G. Zhang, D. Froese, A. Albrechtsen *et al.*, 2013 Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499: 74–78.
- Orlando, L., M. T. P. Gilbert, and E. Willerslev, 2015 Reconstructing ancient genomes and epigenomes. *Nat. Rev. Genet.* 16: 395–408.
- Otoni, C., L. G. Flink, A. Evin, C. Geörg, B. D. Cupere *et al.*, 2013 Pig Domestication and Human-Mediated Dispersal in Western Eurasia Revealed through Ancient DNA and Geometric Morphometrics. *Mol. Biol. Evol.* 30: 824–832.
- Outram, A. K., N. A. Stear, R. Bendrey, S. Olsen, A. Kasparov *et al.*, 2009 The Earliest Horse Harnessing and Milking. *Science* 323: 1332–1335.
- Owen-Smith, N., 1987 Pleistocene Extinctions: The Pivotal Role of Megaherbivores. *Paleobiology* 13: 351–362.
- Pakendorf, B., I. N. Novgorodov, V. L. Osakovskij, A. P. Danilova, A. P. Protod'jakonov *et al.*, 2006 Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. *Hum. Genet.* 120: 334–353.
- Park, S. D. E., D. A. Magee, P. A. McGettigan, M. D. Teasdale, C. J. Edwards *et al.*, 2015 Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the phylogeography and evolution of cattle. *Genome Biol.* 16: 234.
- Pedersen, J. S., E. Valen, A. M. V. Velazquez, B. J. Parker, M. Rasmussen *et al.*, 2014 Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res.* 24: 454–466.
- Petersen, J. L., J. R. Mickelson, A. K. Rendahl, S. J. Valberg, L. S. Andersson *et al.*, 2013a Genome-wide analysis reveals selection for important traits in domestic horse breeds. *PLoS Genet.* 9: e1003211.
- Petersen, J. L., J. R. Mickelson, E. G. Cothran, L. S. Andersson, J. Axelsson *et al.*, 2013b Genetic Diversity in the Modern Horse Illustrated from Genome-Wide SNP Data. *PLoS One* 8: e54997.
- Promerová, M., L. S. Andersson, R. Juras, M. C. Penedo, and M. Reissmann *et al.*, 2014 Worldwide frequency distribution of the “Gait keeper” mutation in the DMRT3 gene. *Anim. Genet.* 45: 274–282.
- Pruvost, M., R. Bellone, N. Benecke, E. Sandoval-Castellanos, M. Cieslak *et al.*, 2011 Genotypes of predomestic horses match phenotypes painted in Paleolithic works of cave art. *Proc. Natl. Acad. Sci. USA* 108: 18626–18630.
- Ramírez, O., W. Burgos-Paz, E. Casas, M. Ballester, E. Bianco *et al.*, 2015 Genome data from a sixteenth century pig illuminate modern breed relationships. *Heredity* 114: 175–184.
- Rasmussen, M., S. L. Anzick, M. R. Waters, P. Skoglund, M. DeGiorgio *et al.*, 2014 The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506: 225–229.
- Römpler, H., P. H. Dear, J. Krause, M. Meyer, N. Rohland *et al.*, 2006 Multiplex amplification of ancient DNA. *Nat. Protoc.* 1: 720–728.

- Sandmeyer, L. S., R. R. Bellone, S. Archer, B. S. Bauer, J. Nelson *et al.*, 2012 Congenital stationary night blindness is associated with the leopard complex in the miniature horse. *Vet. Ophthalmol.* 15: 18–22.
- Sarkissian, C. D., J. T. Vilstrup, M. Schubert, A. Seguin-Orlando, D. Eme *et al.*, 2015 Mitochondrial genomes reveal the extinct Hippidion as an outgroup to all living equids. *Biol. Lett.* 11: 20141058.
- Scheu, A., A. Powell, R. Bollongino, J.-D. Vigne, A. Tresset *et al.*, 2015 The genetic prehistory of domesticated cattle from their origin to the spread across Europe. *BMC Genet.* 16: 54.
- Schraiber, J. G., S. N. Evans, and M. Slatkin, 2016 Bayesian inference of natural selection from allele frequency time series. *GENETICS* 23: 493–511.
- Schubert, M., H. Jönsson, D. Chang, C. Der Sarkissian, L. Ermini *et al.*, 2014 Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc Natl Acad Sci USA* 111: E5661–E5669.
- Sommer, R. S., N. Benecke, L. Lõugas, O. Nelle, and U. Schmölcke, 2011 Holocene survival of the wild horse in Europe: a matter of open landscape? *J. Quat. Sci.* 26: 805–812.
- Sundqvist, A.-K., S. Björnerfeldt, J. A. Leonard, F. Hailer, Å. Hedhammar *et al.*, 2006 Unequal Contribution of Sexes in the Origin of Dog Breeds. *Genetics* 172: 1121–1128.
- Thalmann, O., B. Shapiro, P. Cui, V. J. Schuenemann, S. K. Sawyer *et al.*, 2013 Complete Mitochondrial Genomes of Ancient Canids Suggest a European Origin of Domestic Dogs. *Science* 342: 871–874.
- Tozaki, T., T. Miyake, H. Kakoi, H. Gawahara, S. Sugita *et al.*, 2010 A genome-wide association study for racing performances in Thoroughbreds clarifies a candidate region near the MSTN gene: A genome-wide scan for racing performances. *Anim. Genet.* 41: 28–35.
- Turner, T. L., M. W. Hahn, and S. V. Nuzhdin, 2005 Genomic Islands of Speciation in *Anopheles gambiae*. *PLoS Biol.* 3: e285.
- Vilà, C., J. A. Leonard, A. Götherström, S. Marklund, K. Sandberg *et al.*, 2001 Widespread Origins of Domestic Horse Lineages. *Science* 291: 474–477.
- Vilstrup, J. T., A. Seguin-Orlando, M. Stiller, A. Ginolhac, M. Raghavan *et al.*, 2013 Mitochondrial Phylogenomics of Modern and Ancient Equids. *PLoS One* 8: e55950.
- Wade, C. M., E. Giulotto, S. Sigurdsson, M. Zoli, and S. Gnerre *et al.*, 2009 Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 326: 865–867.
- Wallner, B., C. Vogl, P. Shukla, J. P. Burgstaller, T. Druml *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.
- Warinner, C., J. Hendy, C. Speller, E. Cappellini, R. Fischer *et al.*, 2014 Direct evidence of milk consumption from ancient human dental calculus. *Sci. Rep.* 4: 7104.
- Warmuth, V., A. Eriksson, M. A. Bower, J. Cañon, and G. Cothran *et al.*, 2011 European Domestic Horses Originated in Two Holocene Refugia. *PLoS One* 6: e18194.
- Warmuth, V., A. Eriksson, M. A. Bower, G. Barker, E. Barrett *et al.*, 2012 Reconstructing the origin and spread of horse domestication in the Eurasian steppe. *Proc. Natl. Acad. Sci. USA* 109: 8202–8206.
- Warmuth, V. M., M. G. Campana, A. Eriksson, M. Bower, G. Barker *et al.*, 2013 Ancient trade routes shaped the genetic structure of horses in eastern Eurasia. *Mol. Ecol.* 22: 5340–5351.
- Weinstock, J., E. Willerslev, A. Sher, W. Tong, S. Y. W. Ho *et al.*, 2005 Evolution, Systematics, and Phylogeography of Pleistocene Horses in the New World: A Molecular Perspective. *PLoS Biol.* 3: e241.
- Zeder, M. A., E. Emshwiller, B. D. Smith, and D. G. Bradley, 2006 Documenting domestication: the intersection of genetics and archaeology. *Trends Genet.* 22: 139–155.
- Zeng, K., Y.-X. Fu, S. Shi, and C.-I. Wu, 2006 Statistical Tests for Detecting Positive Selection by Utilizing High-Frequency Variants. *Genetics* 174: 1431–1439.

Communicating editor: J. Rine