REPORT

PALEOGENOMICS

Ancient genomic changes associated with domestication of the horse

Pablo Librado,[1]* Cristina Gamba,[1]* Charleen Gaunitz,[1] Clio Der Sarkissian,[1]
Mélanie Pruvost,[2] Anders Albrechtsen,[3] Antoine Fages,[1,4] Naveed Khan,[1,5]
Mikkel Schubert,[1] Vidhya Jagannathan,[6] Aitor Serres-Armero,[7,8]
Lukas F. K. Kuderna,[7,8] Inna S. Povolotskaya,[7,8] Andaine Seguin-Orlando,[1,9]
Sébastien Lepetz,[10] Markus Neuditschko,[11] Catherine Thèves,[4] Saleh Alquraishi,[12]
Ahmed H. Alfarhan,[12] Khaled Al-Rasheid,[12] Stefan Rieder,[11] Zainolla Samashev,[13]
Henri-Paul Francfort,[14] Norbert Benecke,[15] Michael Hofreiter,[16] Arne Ludwig,[17]
Christine Keyser,[4,18] Tomas Marques-Bonet,[7,8,19] Bertrand Ludes,[4,20]
Eric Crubézy,[4] Tosso Leeb,[6] Eske Willerslev,[1] Ludovic Orlando[1,4]†

The genomic changes underlying both early and late stages of horse domestication remain largely unknown. We examined the genomes of 14 early domestic horses from the Bronze and Iron Ages, dating to between ~4.1 and 2.3 thousand years before present. We find early domestication selection patterns supporting the neural crest hypothesis, which provides a unified developmental origin for common domestic traits. Within the past 2.3 thousand years, horses lost genetic diversity and archaic DNA tracts introgressed from a now-extinct lineage. They accumulated deleterious mutations later than expected under the cost-of-domestication hypothesis, probably because of breeding from limited numbers of stallions. We also reveal that Iron Age Scythian steppe nomads implemented breeding strategies involving no detectable inbreeding and selection for coat-color variation and robust forelimbs.

Horse domestication likely started in the Kazakh steppe with the Botai culture ~5.5 thousand years (ky) ago (1), although earlier (2) and later (3) dates have been proposed. By riding horses, humans could travel well above their own speed, connecting vast territories (4) and revolutionizing warfare with chariotry and cavalry (5). Furthermore, the breeding industry from the 18th century onward was instrumental to modern cities and economies, until horse power was supplanted in the early 20th century.

Horses deeply transformed human civilizations, but humans also reshaped the horse through selection and crossbreeding. Present-day domestic horses show (i) an extreme mitochondrial diversity contrasting with an almost complete homogeneity on their Y chromosome; (ii) higher mutational loads than in wild horses from the Upper Paleolithic; and (iii) selection at genes involved in locomotion, physiology, development, and behavior (6). In the absence of genomes from ancient domestic horses, it remains unclear, however, whether these characteristics were already introduced during early domestication stages.

We therefore explored the potential of 16 domestic horses from three archaeological sites for whole-genome sequencing (Fig. 1A). First, we analyzed a mare radiocarbon dated to ~4.1 ky before present (ky B.P.) [2121 to 2095 calibrated years Before the Common Era (B.C.E.)], belonging to the early Bronze Age Sintashta culture from Chelyabinsk oblast, Russia, which provides the first archaeological evidence of two-wheeled horse chariots (2). Second, we studied two Iron Age stallions from the earliest Scythian royal mound of Arzhan I, Tuva (~2.7 ky B.P.) (7), where >200 harnessed horses were sacrificed during the funerals of an elite member (8). Last, we investigated 13 sacrificed Scythian stallions from the kurgan 11 of Berel', Kazakhstan, dated to ~2.3 ky B.P. (9).

We performed low-depth sequencing to assess DNA preservation levels and molecular signatures typical of ancient DNA (figs. S15 to S18). The exceptional preservation conditions, especially at Berel', where the sepulchral chamber was embedded within permafrost (10), were compatible with whole-genome sequencing of the Sintashta mare and the two Arzhan stallions, as well as 11 Berel' stallions (table S27). Following uracil-specific excision reagent (USER) enzymatic treatment, quality recalibration on the basis of residual damage profiles and read trimming (table S28), we obtained 14 genomes at 1.2 to 10.9 average fold coverage and 0.057 to 0.267% errors per base (fig. S1 and table S1). We also confirmed the mitochondrial sequences previously obtained for Berel' horses (11).

In order to gain insights into the horses' phenotypes and, consequently, past funerary rituals and breeding strategies, we used shotgun and target-enrichment sequencing to investigate 41 Mendelian loci associated with coat color, racing performance, body size, and congenital diseases (Fig. 1, A and B, and fig. S6). Coat coloration genotyping was validated and also expanded for the two Berel' samples not amenable to whole-genome sequencing (BER03_C and BER13_N), with aMPLex-Torrent sequencing (12).

Whereas the Sintashta mare was found to be bay, the Scythian stallions included one cream, two black, two spotted, four bay, and six chestnut individuals (Fig. 1, A and B). This mirrors the coat coloration diversity found at Arzhan II ~2.6 thousand years ago (ka) (13), revealing that Scythians integrated multiple coat-color patterns in elite funerals.

We did not detect mutations causing congenital diseases nor the *DMRT3* allele responsible for ambling (Fig. 1B). We, however, detected alleles associated with racing performance in *ACN9*, *CKM*, *COX4/1*, and *COX4/2*, previously identified in mid-Holocene and Upper Paleolithic horses (14, 15). We found other mutations hitherto unreported in ancient horses, including the *MSTN* mutation associated with muscle hypertrophy and short-distance sprint performance in homozygous Thoroughbred and Quarter horses (the short interspersed nuclear element insertion at the 5′-untranslated region of *MSTN* was, however, absent) (16). The Sintashta mare and four Scythian stallions were heterozygous and were thus probably not as fast as modern sprint racers. The presence of the *MSTN* mutation may indicate that Scythian breeders selected a diversity of endurance and speed potential, provided that further work reveals significantly lower allele frequencies in contemporary wild horses.

[1]Centre for GeoGenetics, Natural History Museum of Denmark, Øster Voldgade 5-7, 1350K Copenhagen, Denmark. [2]Institut Jacques Monod, UMR 7592 CNRS, Université Paris Diderot, 75205 Paris cedex 13, France. [3]Bioinformatics Center, Department of Biology, University of Copenhagen, 2200N Copenhagen, Denmark. [4]Laboratoire d'Anthropobiologie Moléculaire et d'Imagerie de Synthèse, CNRS UMR 5288, Université de Toulouse, Université Paul Sabatier, 31000 Toulouse, France. [5]Department of Biotechnology, Abdul Wali Khan University, Mardan, Pakistan. [6]Institute of Genetics, University of Bern, 3001 Bern, Switzerland. [7]Institute of Evolutionary Biology (CSIC-UPF), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, 08003 Barcelona, Spain. [8]Center for Genomic Regulation (CNAG-CRG), Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 4, 08028 Barcelona, Spain. [9]National High-Throughput DNA Sequencing Center, Copenhagen, Denmark. [10]Centre National de la Recherche Scientifique, Muséum national d'histoire naturelle, Sorbonne Universités, Archéozoologie, Archéobotanique, Sociétés, Pratiques et Environnements (UMR 7209), 55 rue Buffon, 75005 Paris, France. [11]Agroscope, Swiss National Stud Farm, 1580 Avenches, Switzerland. [12]Zoology Department, College of Science, King Saud University, Riyadh 11451, Saudi Arabia. [13]Branch of Institute of Archaeology Margulan, Republic Avenue 24-405, 010000 Astana, Republic of Kazakhstan. [14]CNRS, UMR 7041 Archéologie et Sciences de l'Antiquité, Archéologie de l'Asie Centrale, Maison René Ginouvès, 21 allée de l'Université, 92023 Nanterre, France. [15]German Archaeological Institute, Department of Natural Sciences, Berlin, 14195 Berlin, Germany. [16]University of Potsdam, Faculty of Mathematics and Natural Sciences, Institute for Biochemistry and Biology, Karl-Liebknecht-Strasse 24-25, 14476 Potsdam, Germany. [17]Department of Evolutionary Genetics, Leibniz Institute for Zoo and Wildlife Research, Berlin 10315, Germany. [18]Institut de Médecine Légale, Université de Strasbourg, Strasbourg, France. [19]Catalan Institution of Research and Advanced Studies (ICREA), Passeig de Lluís Companys, 23, 08010 Barcelona, Spain. [20]Institut Médico-Légal, Université Paris Descartes, Paris, France.
*These authors contributed equally to this work. †Corresponding author: Email: lorlando@snm.ku.dk

We next evaluated kinship and inbreeding in the 11 Berel' horses (Fig. 1C), and this took genotyping uncertainty into account (17, 18). Only BER10_K and BER11_L were related (kinship coefficient of 0.451). They showed the same mitochondrial haplotype (figs. S3 and S5) but different Y chromosomes (Fig. 2A). These two horses might represent appreciated members of valuable pedigrees. The lack of kinship found in the majority of horses echoes both Herodotus's depiction of sacrificed horses as gifts from allied tribes spread across vast areas and the diversity of harness ornaments excavated at Berel' (19). Inbreeding coefficients (F) were also close to zero (Fig. 1C), as opposed to all present-day horses tested, which suggests that Scythian reproductive management did not disrupt natural herd structures, in contrast with current practice.

Finally, we used the population branch statistics (PBS) to identify 121 candidate genes selected by Scythian breeders (table S14). As segmental duplications affect patterns of sequence variation, we filtered copy number variations (CNVs) (fig. S13) and found significant functional enrichment for development of the anteroposterior axis and carpal bones (hypergeometric test, adjusted P values ≤ 0.0438) (tables S18 and S23). Genes expressed in the pectoral appendage apical ectodermal ridge, the tibia, the clavicle, and the radius bone, were also overrepresented (tables S21 and S25). This finding is consistent with the increased robustness measured on Berel' metacarpals (and in other

Altay Scythian horses), compared with present-day Mongolian horses (9).

Breast and mammary glands also appeared functionally overrepresented, together with the posterior pituitary, which produces the neurohypophysal hormones oxytocin and vasopressin (hypergeometric test, adjusted P ≤ 0.0493) (tables S18 and S23). Although the former is involved in uterine contraction, lactation, and social bonding between humans and dogs (20), the latter stimulates water retention. Genes transcribed in the urothelium and kidney vasculature were enriched among the tissues where the selection candidates are expressed (table S21). Altogether, these findings suggest that Scythians may have favored gene variants facilitating horse milking, which was practiced since ~5.5 ka (1), and minimizing water loss, an advantage in dry steppe areas where daily water sources are scarce.

The 14 ancient genomes reported here have strong implications for the horse domestication process. First, it has recently been discovered that a now-extinct lineage of wild horses existed in the Arctic until at least ~5.2 ka and significantly contributed to the genetic makeup of present-day domesticates (14, 15). The timing of the underlying admixture event(s) is, however, unknown. Using D statistics, we confirmed that this extinct lineage shared more derived polymorphisms with the Sintashta and especially Scythian horses than with present-day domesticates (Fig. 2B). The domestic horse lineage, thus, experienced a net loss

of archaic introgressed tracts within the past ~2.3 ky.

Furthermore, predomestication horses carry fewer deleterious mutations than present-day domesticates (14). This suggested that the demographic collapse associated with domestication reduced the efficacy of purifying selection in filtering out deleterious alleles. However, using phyloP scores (21) as proxies for the fitness consequences of mutations, we found lower mutational loads in Sintashta and Scythian horses than in both present-day horses and previously sequenced predomestication genomes (14, 15) (Fig. 2C). This pattern was not driven by differences in genome coverage (fig. S29). Therefore, the excess of deleterious mutations in present-day horses—functionally overrepresented in open wounds (hypergeometric test, adjusted P = 0.0402), seizures (0.0390), and dementia (0.0197)—is likely not a consequence of early domestication but of the past ~2.3 ky of breeding.

Our data challenge the evolutionary paradigm that the domestic horse lineage was founded by a limited number of stallions and constantly restocked through mares. The mitochondrial diversity of present-day horses is similar to that of Berel' horses, both in terms of nucleotide diversity (π in Fig. 2D) and haplotypes spread across the entire mitochondrial tree (figs. S3 and S5). For the Y chromosome, Scythian horses showed a 9.45-fold increase in π (Fig. 2D) and many additional haplotypes compared with the few segregating in present-day domesticates (Fig. 2A). This is in agreement with the divergent Scythian haplotype previously reported (22) and reveals a large diversity of domestic male founders taking part in early domestication. The reduction in the stallion population size during the past ~2.3 ky also resulted in a severe decline in the overall effective population size, as reflected by the reduced autosomal π within present-day horses (Fig. 2D). It did not, however, affect the X-to-autosomal diversity (π) ratio, which remains below the 0.75 random mating expectation and is similar in both Berel' and present-day domesticates (~0.55 to 0.60). Present-day domesticates, however, lost diversity for the X chromosome, especially within a ~13-Mb region also showing high fixation index (F_{ST}) values with Berel' horses (fig. S14) and enriched in genes associated with intellectual disability (hypergeometric test, adjusted P = 0.0032); behavior (0.0002 to 0.0032); and long fingers (0.0016) (tables S17 and S20).

We developed a statistical framework to identify the genetic changes selected before the divergence between Berel' horses and present-day domesticates, a period encompassing the beginning and early stages of domestication. The method exploits a population tree and levels of exclusively shared derived (LSD) polymorphisms within predefined groups of individuals (fig. S30). Using forward simulations under eight selective regimes, LSD can detect selection within the timeframe of horse domestication (figs. S31 to S33).

We applied LSD for the seven Berel' horses with highest genome coverage (table S15) and the tree retrieved from patterns of genome-wide variation,
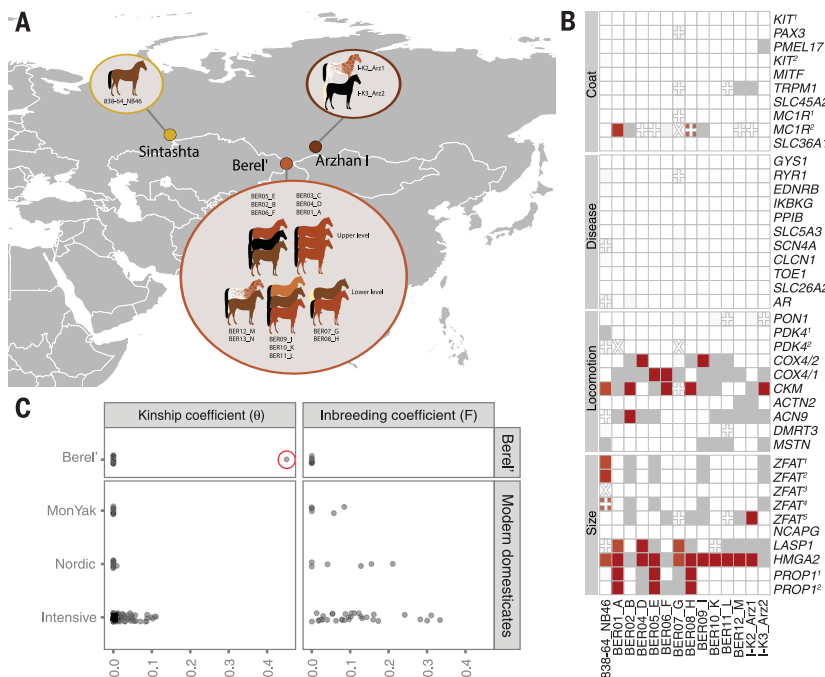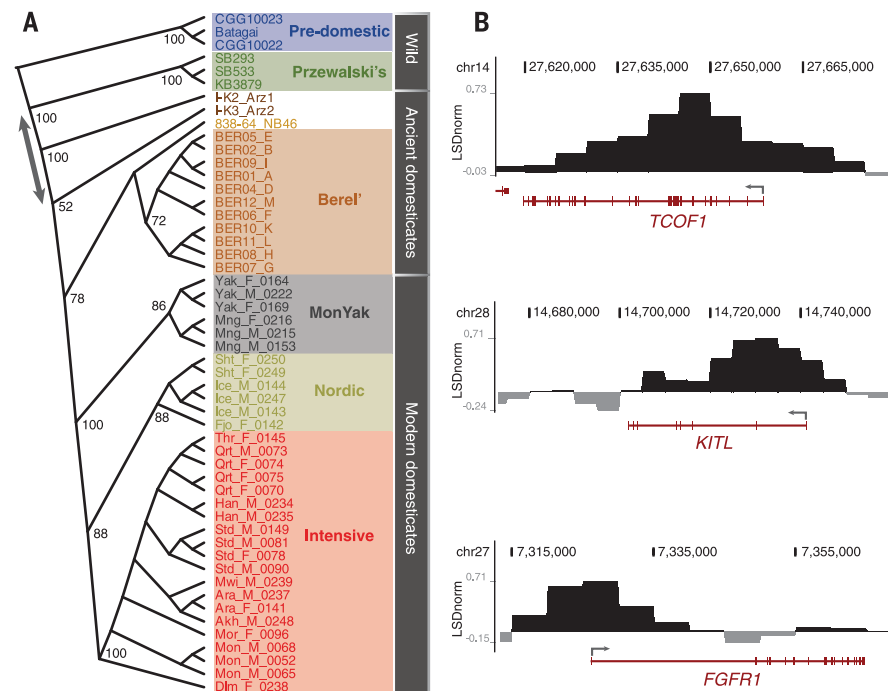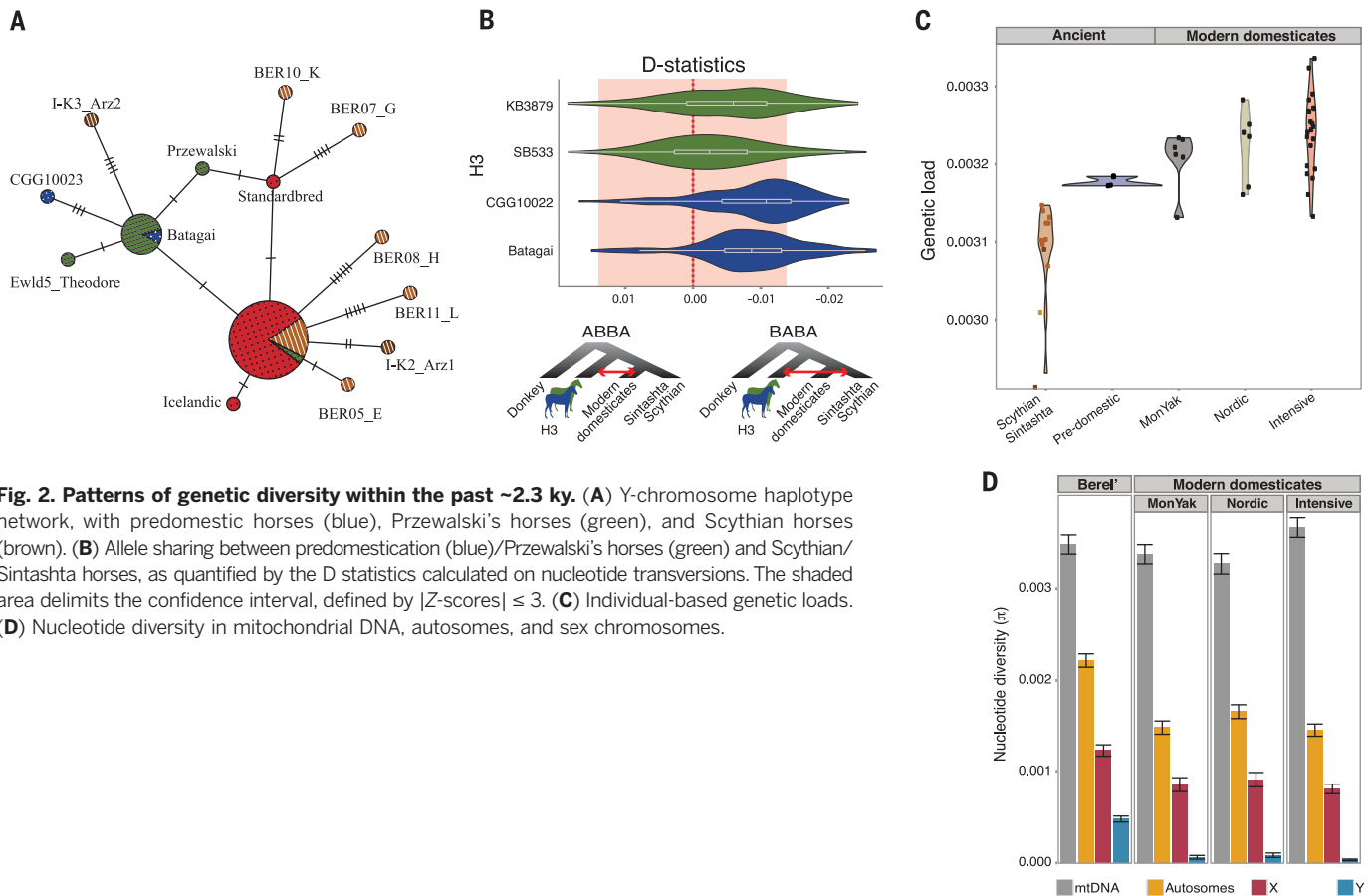

Fig. 1. Ancient horse samples, kinship, inbreeding, and phenotypes. (**A**) Sampling sites. Berel' stallions are displayed as found on site. (**B**) Allele probabilities at 41 predicted Mendelian loci (fig. S6 and table S30). The absence of the causative allele is depicted in white; its presence in heterozygous or homozygous states is shown in dark gray and red, respectively. Plus signs and diagonal crosses depict nongenotyped positions and sites with a sequencing coverage lower than three, respectively. (**C**) Kinship (θ) and inbreeding coefficients (F) for four main groups of domestic horses. The kinship coefficient between BER10_K and BER11_L is highlighted in a red circle.

Fig. 2. Patterns of genetic diversity within the past ~2.3 ky. (**A**) Y-chromosome haplotype network, with predomestic horses (blue), Przewalski's horses (green), and Scythian horses (brown). (**B**) Allele sharing between predomestication (blue)/Przewalski's horses (green) and Scythian/Sintashta horses, as quantified by the D statistics calculated on nucleotide transversions. The shaded area delimits the confidence interval, defined by |Z-scores| ≤ 3. (**C**) Individual-based genetic loads. (**D**) Nucleotide diversity in mitochondrial DNA, autosomes, and sex chromosomes.

Fig. 3. Population affinities and LSD-based selection scan. (**A**) Neighbor-joining population tree with bootstrap node supports and the six horse groups identified. The ancestral branch leading to the divergence between Berel' horses and present-day domesticates is highlighted with a gray arrow. (**B**) Three gene candidates involved in neural crest–development programs and showing LSD support for positive selection during early domestication stages.

showing the Sintashta and Scythian horses as basal to the lineage of domesticates, with Berel' horses forming a monophyletic group (Fig. 3A). This topology was confirmed with outgroup f_3-statistics placing Berel' horses equidistant to all present-day domesticates (fig. S8).

Defining six groups from the tree and filtering CNVs, functional clustering of the 1000 top-ranking 10-kb windows of normalized LSD scores revealed significant enrichment for genes involved in androgen and steroid hormone receptor binding, abnormal synaptic transmissions, and associative learning (hypergeometric test, adjusted P ≤ 0.0311) (tables S24 and S26). This reflects the important cognitive and behavioral changes accompanying animal domestication. Enrichment was also observed for genes related to ear shape (0.0163); neural crest (cell) morphology (0.0293); and genes transcribed in the mesenchyme derived from head neural crest (0.0280) and the substantia nigra (0.0507), a brain region containing neural crest–derived neurons involved in movement, learning, and reward (23) (Fig. 3B). Our findings thus support the neural crest hypothesis of animal domestication, which proposes that developmental changes affecting the tissues and cell types derived from the neural crest underpin traits commonly found in domestic animals, such as coat-color variation and floppy ears (24).

We unveiled important features of the Scythian funerary rituals and revised our views on past

horse breeding and management. Determining exactly which cultures and technologies caused the demise of stallion diversity, the surge in mutational load, and the development of specific equestrian traits will require additional genomes spanning the whole temporal and geographical range of horse domestication.

REFERENCES AND NOTES

1. A. K. Outram et al., Science 323, 1332–1335 (2009).
2. D. W. Anthony, The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World (Princeton Univ. Press, 2007).
3. M. A. Levine, C. Renfrew, K. V. Boyle, Prehistoric Steppe Adaptation and the Horse (McDonald Institute for Archaeological Research, Cambridge, 2003).
4. P. Kelekna, The Horse in Human History (Cambridge Univ. Press, 2009).
5. R. Drews, Early Riders: The Beginnings of Mounted Warfare in Asia and Europe (Routledge, New York, 2004).
6. P. Librado et al., Genetics 204, 423–434 (2016).
7. K. V. Cugunov, H. Parzinger, A. Nagler, Eurasia Antiq. 9, 113–162 (2003).
8. M. Gryaznov, Arzhan: Tsar's Kurgan of the Early Scythian Period [in Russian: Arzhan. Carskij kurgan ranneskifskogo vremeni.] (Nauka, Leningrad, 1980).
9. H.-P. Francfort, S. Lepetz, in Histoire d'équidés: Des textes, des images et des os, N. Boulbes, E. Furet, A. Gardeisen, Eds. (Monographie d'Archéologie Méditerranéenne, spec. ed. 4, CNRS, 2010), pp. 57–104.
10. S. Lepetz, Anthropozoologica 48, 309–321 (2013).
11. C. Keyser-Tracqui et al., Anim. Genet. 36, 203–209 (2005).
12. S. Guimaraes et al., Mol. Ecol. Resour. 10.1111/1755-0998.12565 (2016).
13. A. Ludwig et al., Science 324, 485–485 (2009).
14. M. Schubert et al., Proc. Natl. Acad. Sci. U.S.A. 111, E5661–E5669 (2014).
15. P. Librado et al., Proc. Natl. Acad. Sci. U.S.A. 112, E6889–E6897 (2015).
16. J.-L. L. Rivero, E. W. Hill, Vet. J. 209, 5–13 (2016).
17. F. G. Vieira, M. Fumagalli, A. Albrechtsen, R. Nielsen, Genome Res. 23, 1852–1861 (2013).
18. T. S. Korneliussen, I. Moltke, Bioinformatics 31, 4009–4011 (2015).
19. H.-P. Francfort, G. Ligabue, Z. Samashev, C. R. Acad. Inscr. Belle. 144, 775–806 (2000).
20. M. Nagasawa et al., Science 348, 333–336 (2015).
21. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, Genome Res. 20, 110–121 (2010).
22. S. Lippold et al., Nat. Commun. 2, 450 (2011).
23. A. V. Kulikov et al., Genes Brain Behav. 15, 733–740 (2016).
24. A. S. Wilkins, R. W. Wrangham, W. T. Fitch, Genetics 197, 795–808 (2014).

Editor's Summary

Ancient genomics of horse domestication

The domestication of the horse was a seminal event in human cultural evolution. Librado *et al.* obtained genome sequences from 14 horses from the Bronze and Iron Ages, about 2000 to 4000 years ago, soon after domestication. They identified variants determining coat color and genes selected during the domestication process. They could also see evidence of admixture with archaic horses and the demography of the domestication process, which included the accumulation of deleterious variants. The horse appears to have undergone a different type of domestication process than animals that were domesticated simply for food.

Science, this issue p. 442

Article Tools	Visit the online version of this article to access the personalization and article tools: http://science.sciencemag.org/content/356/6336/442
Permissions	Obtain information about reproducing this article: http://www.sciencemag.org/about/permissions.dtl