

## GENETICS

# Decline of genetic diversity in ancient domestic stallions in Europe

Saskia Wutke,<sup>1,2\*</sup> Edson Sandoval-Castellanos,<sup>3\*</sup> Norbert Benecke,<sup>4</sup> Hans-Jürgen Döhle,<sup>5</sup> Susanne Friederich,<sup>5</sup> Javier Gonzalez,<sup>6</sup> Michael Hofreiter,<sup>6</sup> Lembi Lõugas,<sup>7</sup> Ola Magnell,<sup>8</sup> Anna-Sapfo Malaspinas,<sup>9†</sup> Arturo Morales-Muñiz,<sup>10</sup> Ludovic Orlando,<sup>11,12</sup> Monika Reissmann,<sup>13</sup> Alexandra Trinks,<sup>6</sup> Arne Ludwig<sup>1‡</sup>

Present-day domestic horses are immensely diverse in their maternally inherited mitochondrial DNA, yet they show very little variation on their paternally inherited Y chromosome. Although it has recently been shown that Y chromosomal diversity in domestic horses was higher at least until the Iron Age, when and why this diversity disappeared remain controversial questions. We genotyped 16 recently discovered Y chromosomal single-nucleotide polymorphisms in 96 ancient Eurasian stallions spanning the early domestication stages (Copper and Bronze Age) to the Middle Ages. Using this Y chromosomal time series, which covers nearly the entire history of horse domestication, we reveal how Y chromosomal diversity changed over time. Our results also show that the lack of multiple stallion lineages in the extant domestic population is caused by neither a founder effect nor random demographic effects but instead is the result of artificial selection—initially during the Iron Age by nomadic people from the Eurasian steppes and later during the Roman period. Moreover, the modern domestic haplotype probably derived from another, already advantageous, haplotype, most likely after the beginning of the domestication. In line with recent findings indicating that the Przewalski and domestic horse lineages remained connected by gene flow after they diverged about 45,000 years ago, we present evidence for Y chromosomal introgression of Przewalski horses into the gene pool of European domestic horses at least until medieval times.

**INTRODUCTION**

Present-day domestic horses are unique among livestock animals because of their almost complete lack of variability on the paternally inherited Y chromosome (1–5). In stark contrast to that, the maternally inherited mitochondrial DNA of horses shows huge variation (6–10). So far, only very few studies detected polymorphic Y chromosomal markers among modern horses, but these additional lineages seem to be derived from the ancestral domestic haplotype relatively recently (5) or were limited to a restricted geographical area (11).

The pattern of high maternal diversity combined with extremely reduced paternal variation is generally explained by unequal sex ratios at the beginning of domestication. Likely, a large number of mares but only few stallions contributed to founding the domestic horse population (2). However, further support for these discrepancies arose from continued sex-specific breeding enhancing the reproductive success of

certain male lineages, particularly after the development of modern studs during the last few centuries (5). Recent ancient DNA (aDNA) analyses of paternal variation have detected the coexistence of multiple Y chromosomal haplotypes in pre-Iron Age and Iron Age domestic stallions, which suggests that Y chromosomal diversity was much larger in past than in present-day horse populations (12, 13). However, because of limited sampling efforts, these earlier studies could not capture the diachronical changes in paternal diversity.

To clarify how the domestication process of horses shaped the diversity of the Y chromosome, we genotyped 16 loci on the Y chromosome in 96 (out of a total of 350) archaeological remains (table S1), spread across Eurasia and spanning the last ~5000 years. This data set provides the first full time series of Y chromosomal variation from early domestication stages until the Middle Ages.

**RESULTS AND DISCUSSION**

To clarify how the domestication process of horses and the potential input of the Przewalski horse shaped the diversity of the horse Y chromosome, we successfully genotyped 16 single-nucleotide polymorphism (SNP) loci on the Y chromosome in 96 (out of a total of 350) archaeological horse remains (table S1), spread across Eurasia and spanning the last ~5000 years. In agreement with previous studies, the Y chromosome showed low genetic diversity, with only 4 of 16 SNPs being polymorphic, defining four haplotypes (table S2). In addition to the haplotypes exhibited by present-day domestic horses (Y-HT-1), we detected two new haplotypes (Y-HT-3 and Y-HT-4) characterized by two differences and one difference from the domestic haplotype, respectively (Fig. 1). Somewhat surprisingly, we also detected in several of the successfully genotyped aDNA samples the Przewalski horse haplotype (Y-HT-2), which differs at all four polymorphic positions from the present-day domestic haplotype.

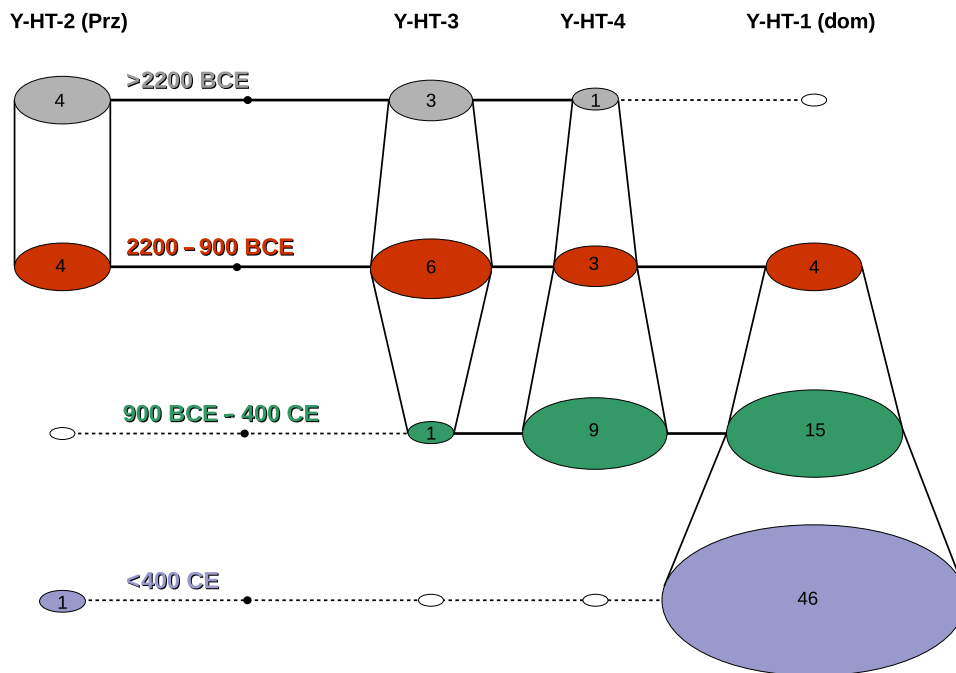
To detect patterns of temporal variation, we divided the samples into four different time bins based on archaeological periods (before

<sup>1</sup>Department of Evolutionary Genetics, Leibniz Institute for Zoo and Wildlife Research, 10315 Berlin, Germany. <sup>2</sup>Department of Environmental and Biological Sciences, University of Eastern Finland, 80101 Joensuu, Finland. <sup>3</sup>Institute of Animal Breeding and Husbandry, University of Kiel, Kiel 24098, Germany. <sup>4</sup>Department of Natural Sciences, German Archaeological Institute, 14195 Berlin, Germany. <sup>5</sup>Landesamt für Denkmalpflege und Archäologie Sachsen-Anhalt—Landesmuseum für Vorgeschichte, 06114 Halle (Saale), Germany. <sup>6</sup>Faculty of Mathematics and Natural Sciences, Institute for Biochemistry and Biology, University of Potsdam, 14476 Potsdam, Germany. <sup>7</sup>Archaeological Research Collection, Tallinn University, Rüütli 10, 10130 Tallinn, Estonia. <sup>8</sup>National Historical Museums, Contract Archaeology, 226 60 Lund, Sweden. <sup>9</sup>Institute of Ecology and Evolution, University of Bern, 3012 Bern, Switzerland. <sup>10</sup>Laboratory of Archaeozoology, Universidad Autónoma de Madrid, Madrid, Spain. <sup>11</sup>Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350K Copenhagen, Denmark. <sup>12</sup>Université de Toulouse, Université Paul Sabatier, Laboratoire Anthropologie Moléculaire et Imagerie de Synthèse, CNRS UMR 5288, Toulouse, France. <sup>13</sup>Faculty of Life Sciences, Albrecht Daniel Thaer-Institute, Humboldt University Berlin, 10115 Berlin, Germany.

\*These authors contributed equally to this work.

†Present address: Department of Computational Biology, University of Lausanne and SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland.

‡Corresponding author. Email: ludwig@izw-berlin.de



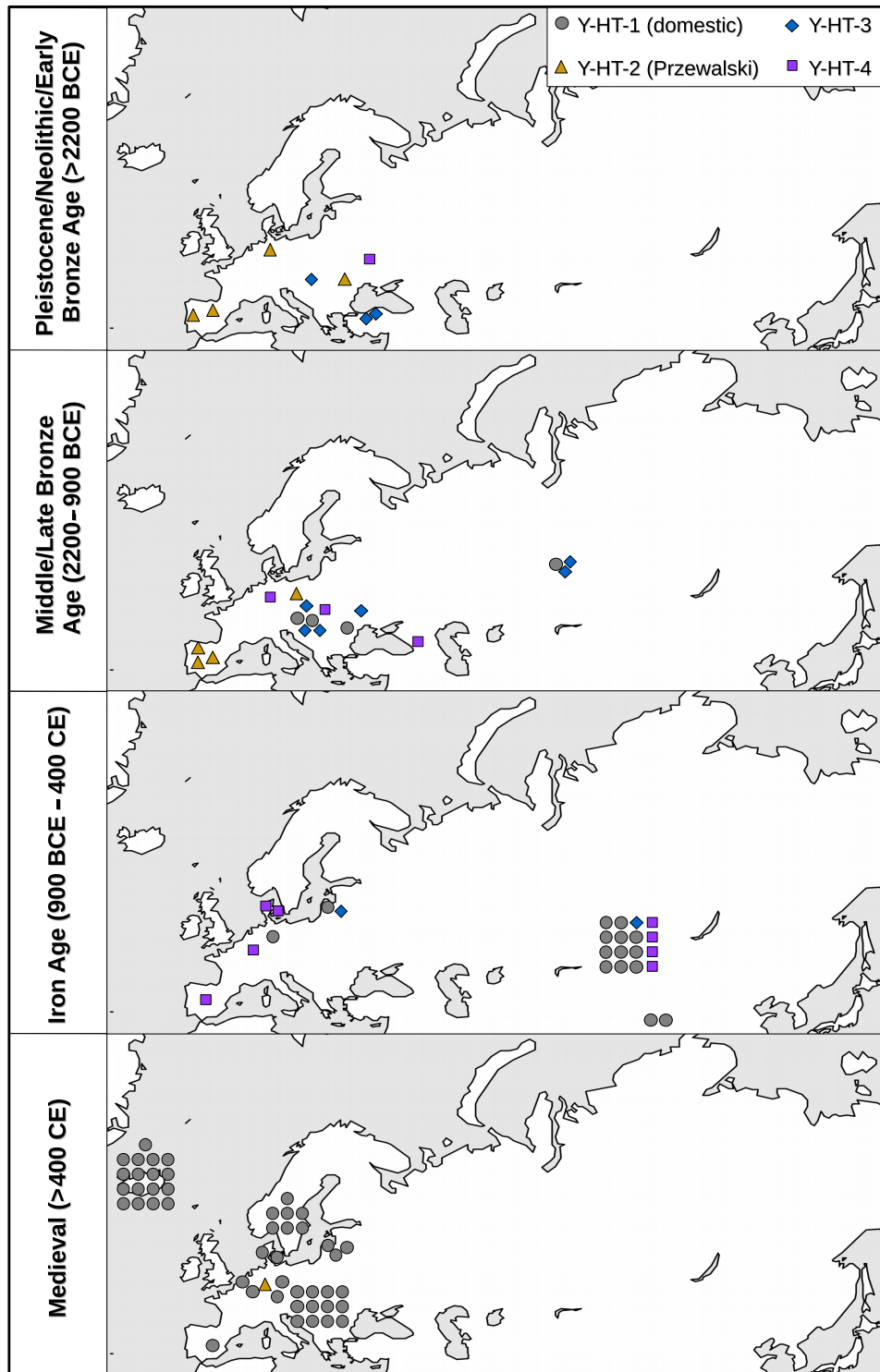
**Fig. 1. Stallion lineages through time.** Temporal haplotype network of the four detected Y chromosome haplotypes. Age of the samples indicated by multiple layers separated by color; vertical lines connecting the haplotypes of consecutive layers/ages represent which haplotype was transferred into a later/younger period. Numbers constitute the respective number of individuals showing this particular haplotype for that period. Prz, Przewalski; Dom, domestic.

2200 BCE, 2200–900 BCE, 900 BCE–400 CE, and younger than 400 CE; Fig. 2). On the basis of these, we constructed a temporal haplotype network (Fig. 1) and simulated the haplotype frequencies over time (Fig. 3). All analyses show the reduction of Y chromosome variation through time, which is especially pronounced for haplotypes Y-HT-1 and Y-HT-4. Wild and early domestic horses from the Neolithic and Bronze Age (>900 BCE) display comparatively high variation, despite their smaller sample size ( $n = 9$  and  $17$  in time bins 1 and 2 versus  $n = 24$  and  $46$  in time bins 3 and 4; Fig. 2), with all four detected haplotypes occurring at relatively even frequency in the first two defined periods (~0.10 to 0.36; Fig. 3 and table S3). Moreover, among European horses, the haplotype representing modern Przewalski horses (Y-HT-2) was very abundant during the two older periods and even constitutes the most frequent one before 2200 BCE (oldest time bin). Haplotype Y-HT-1, on the other hand, which dominates in present-day stallions, was only detected later than 2200 BCE. Although its estimated age roughly correlates with the onset of domestication (3500 BCE; Fig. 4), this haplotype only started to become more frequent during later periods. All analyses indicate low Y-HT-1 frequency before 2200 BCE (fig. S1). However, the first time bin (>2200 BCE) only includes samples from Europe, with the easternmost samples falling on the longitude of the Black Sea (Fig. 2). Accordingly, Y-HT-1 could have been present with higher frequencies at that time in populations from further East. In the Iron Age (time bin 3, 900 BCE–400 CE), Y-HT-1 already represented the most frequent haplotype.

The first record of the modern domestic Y chromosome haplotype stems from two Bronze Age samples of similar age. Notably, both samples were found in two distantly located regions: present-day Slovakia (2000–1600 BCE, dated by archaeological context) and western Siberia ( $^{14}\text{C}$ -dated: 1609–1436 cal. BCE). Although a very recent study proposes an oriental origin of this haplotype (14), we cannot determine the geographical origin of Y-HT-1 with certainty, because this haplotype

has not been found thus far in predomestic or wild stallions. There are two possible scenarios: (i) Y-HT-1 emerged within the domestic population by mutation and (ii) Y-HT-1 was already present in wild horses and entered the domestic population either at the beginning of domestication (but initially restricted to Asian horses) or later by introgression (from wild Y-HT-1 carrying studs during the Iron Age). Crosses between domestic animals and their wild counterparts have been observed in several domestic species (15–18); thus, the simplest explanation would be that we missed Y-HT-1 in older samples because of limited geographical sampling. However, the estimated haplotype age is contemporary (Fig. 4) with the assumed starting point of horse domestication ~4000–3500 BCE (19), rendering it likely that Y-HT-1 originated within the domestic horse gene pool. Still, we cannot rule out definitively that it appeared before domestication.

Independent of its geographical origin, Y-HT-1 progressively replaced all other haplotypes—except for one additional lineage that is restricted to Yakutian horses (11). Considering our data, this trend in paternal diversity toward dominance of the modern lineage appears to start in the Bronze Age and becomes even more pronounced during the Iron Age. The Bronze Age was a time of large-scale human migrations across Eurasia (20–22), movements that were undoubtedly facilitated by the spread of horses as a means of transport and warfare. At that time, the western Eurasian steppes were inhabited by highly mobile cultures that largely relied on horses (20, 21, 23, 24). The genetic admixture of northern and central European humans with Caucasians/eastern Europeans did correlate with the spread of the Yamnaya culture from the Pontic-Caspian steppe (25), an area that has repeatedly been suggested as the center of horse domestication (19, 26, 27). Given the importance of domestic horses, it appears that deliberate selection/rejection of certain stallions by these people might have contributed to the loss of paternal diversity. The spread of humans out of this region might also have resulted in the spread of Y-HT-1 from Asia to Europe. This scenario

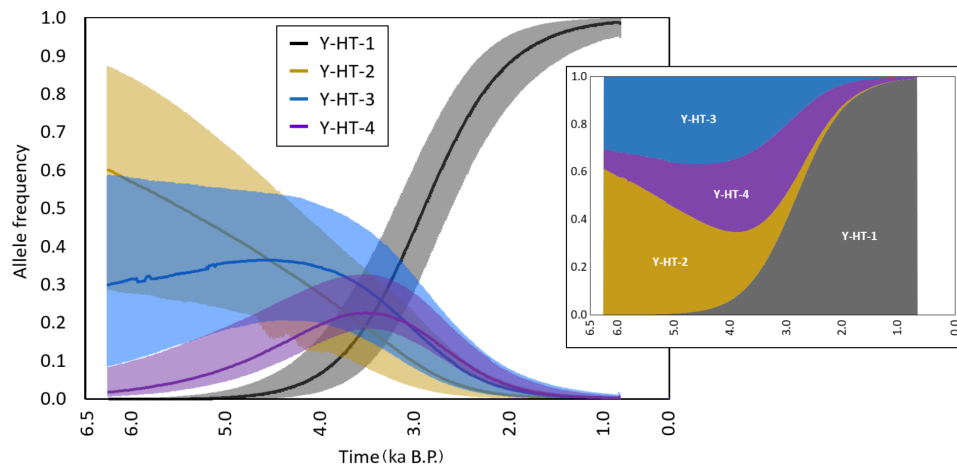


**Fig. 2. Decline of paternal diversity began in Asia.** Maps displaying age, locality, and haplotype (different colors) of each successfully genotyped sample.

also agrees with recent findings that the low male diversity of extant horses is not caused by recruiting only a limited number of stallions during early domestication (13).

To evaluate whether the observed changes in Y haplotype frequencies can be explained by drift or whether selection needs to be invoked, we

applied several analyses. The different methods assumed different demographic scenarios. All methods consistently provided substantial statistical support for positive selection on Y-HT-1 (100% positive coefficients; Fig. 4 and fig. S3). Thus, it appears that deliberate breeding contributed to the loss of paternal diversity. This inference was also



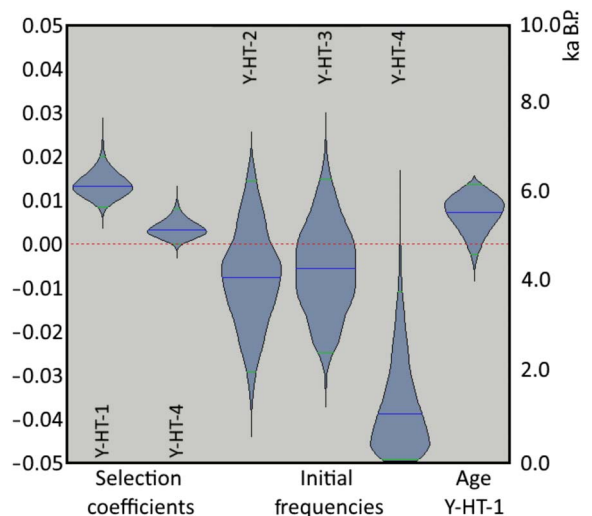
**Fig. 3. The frequencies of Y chromosome haplotypes started to change during the Late Bronze Age (1600–900 BCE).** Inferred temporal trajectories of haplotype frequencies. Each haplotype is displayed by a different color. The shaded area represents the 95% highest-density region. The trajectories were constructed taking the median values across frequencies from the simulations of the Bayesian posterior sample. The small chart represents the stacked frequencies; the amplitude of each colored area is proportional to the median haplotype frequencies (normalized) at a given time. The x and y axes of the small chart match those in the large one. Ka, thousands of years.

supported by our model choice analysis, which provided overwhelming support for the selection model (Bayes factor =  $9.4 \times 10^{11}$ ) when compared to a pure drift model.

When we tested whether a temporally changing selection coefficient provided a better fit for the data, by assigning different coefficients to different periods, we found that one single coefficient for all periods obtained the best combined support of the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) (fig. S4), suggesting that the selective forces acting upon haplotype Y-HT-1 did not change substantially over time.

In addition, we detected strong selection acting also on other Y chromosome haplotypes. Although values were highly positive for Y-HT-4 (98.2% of selection coefficients were positive), selection for Y-HT-2 and Y-HT-3 was significantly negative (100% of coefficients negative). Our analyses indicate that Y-HT-2 declined throughout the sampled time frame, Y-HT-3 initially remained constant and then declined, Y-HT-4 rose before 1500 BCE and subsequently decreased in frequency, and Y-HT-1 almost aggressively invaded the population until fixation (Fig. 3). The whole pattern, including the turnover of Y-HT-4, is best explained by positive selection on Y-HT-4 and even stronger positive selection on Y-HT-1 (Fig. 4).

Y-HT-1 rapidly replaced the already advantageous haplotype Y-HT-4, its most closely related lineage. With Y-HT-4 probably being older than Y-HT-1, it seems possible that the selective advantage of Y-HT-4 was retained in Y-HT-1, which then acquired its own additional advantageous mutations and, consequently, an even higher selective value. Such a situation, where younger haplotypes are under increasingly higher selection, could be a case of selective escalation. In this scenario, the Y-HT-1-specific mutations further improved the advantageous phenotype of Y-HT-4, allowing Y-HT-1 to invade the gene pool of domestic horses. The speed with which Y-HT-1 reached fixation supports this escalation hypothesis. Future research could focus on identifying genes that display nonsynonymous mutations shared only by Y-HT-1 and Y-HT-4 and investigating their phenotypic expression. Possible traits could well be phenotypically visible features because increasing evidence indicates that, in addition to sper-



**Fig. 4. Positive selection for modern domestic haplotype leads to the decline of diversity.** Violin plots of selection coefficients and initial haplotype frequencies, representing the posterior density functions of the selection coefficients of haplotypes Y-HT-1 and Y-HT-4, the age of Y-HT-1, and the initial haplotype frequencies of all haplotypes but Y-HT-1. The scale on the left is for the selection coefficients, whereas the scale on the right is for the age of Y-HT-1. The initial allele frequencies are on a scale from 0.0 (bottom) to 1.0 (top).

matogenesis and male fertility, several diseases and physiological and behavioral traits in males are regulated by Y chromosomal genes (28); such Y-chromosome-associated traits could explain how early breeders chose the stallions they selected for breeding.

During the Iron Age, when Y-HT-1 started to obtain its dominating status, the Roman Empire was ruling Europe and its bordering regions. The Romans changed the mode of animal breeding in their territories by shifting the focus from female-based to male-based selective breeding (29). This change in animal husbandry should have raised the reproductive



success of the few stallions selected for reproduction, increasing the frequency of their Y chromosomal lineages. Although we cannot exclude the possibility that animal husbandry and breeding lost their organization with the end of the Roman Empire, the fact that horses were basically fixed for Y-HT-1 already at the beginning of the Middle Ages would have prevented other haplotypes from rising again during medieval times.

The presence of the Y chromosome haplotype carried by present-day Przewalski horses (Y-HT-2) in early domestic stallions and a European wild horse (Pie05; table S2) could be the result of introgression of Przewalski stallions. Although the original distribution of the Przewalski horse is unknown, it was probably much larger than that of the relict population in Mongolia that produced modern Przewalski horses and might even have extended into Central Europe. However, it is also possible that either Przewalski horses were among the initially domesticated horses or that Y-HT-2 occurred both in Przewalski horses and in those wild horses that are the ancestors of domestic horses, based on autosomal DNA data (30). Regardless of how Y-HT-2 entered the domestic gene pool, it was eventually lost, as were all haplotypes except Y-HT-1. In our sample set, Y-HT-2 was undetectable as early as the third time bin. However, it is possible that Y-HT-2 may have been present during this time period, but with a frequency below 0.11 (with 95% probability). The inferred time trajectories for Y-HT-2 frequencies suggest that it could nevertheless have persisted at very low frequencies until the Middle Ages (Fig. 3). On the basis of these simulations, this finding could be interpreted as a relic of this haplotype's formerly higher frequency in the domestic horse gene pool. It is also possible that the presence of this haplotype could be the result of mating a wild stallion with a domestic mare, a frequently reported breeding practice when wild horses were still widely distributed. However, a significant contribution of the Przewalski horse to the gene pool of modern domestic horses has been almost ruled out by recent genomic studies (13, 31, 32).

## MATERIALS AND METHODS

### Sample preparation, aDNA extraction, amplification, and sequencing

We analyzed 350 samples from China; Mongolia; Siberia; eastern, middle, and northern Europe; and the Iberian Peninsula (table S1). The ages of the samples range from the Neolithic/Copper Age to medieval times and were determined either by  $^{14}\text{C}$ -dating of the samples or by their archaeological context. DNA was extracted from 100- to 200-mg samples of bone or tooth in laboratories dedicated to aDNA analyses at the Leibniz Institute for Zoo and Wildlife Research (Ludwig Laboratory) in Berlin, Germany, following the standard procedures to avoid contamination. Independent replications were carried out in a different laboratory at the University of Potsdam (Hofreiter Laboratory), Germany. The tissue surfaces of all samples were removed by abrasion to minimize contamination. Samples were ground with a cryomill and incubated in 1 ml of extraction buffer [0.45 M EDTA, proteinase K (0.25 mg/ml) (pH 8.0)] overnight under rotation at 37°C. An initial scan for the sex of the samples was done by polymerase chain reaction (PCR) using sex-specific primers [see the study of Lippold *et al.* (12) for details] and subsequent visualization on an agarose gel. The following steps were conducted according to an optimized extraction protocol for short DNA fragments (33). All samples were analyzed for 16 previously published SNPs (12) amplified with 12 primer pairs (table S4). Multiplex PCRs were performed in 20- $\mu\text{l}$  reactions using 4  $\mu\text{l}$  of DNA extract and further containing 1 $\times$  AmpliTaq Gold PCR buffer II (Applied

Biosystems); 4 mM  $\text{MgCl}_2$ ; bovine serum albumin (1 mg/ml); 250  $\mu\text{M}$  dATP (2'-deoxyadenosine 5'-triphosphate), dCTP (2'-deoxycytidine 5'-triphosphate), and dGTP (2'-deoxyguanosine 5'-triphosphate); 500  $\mu\text{M}$  dUTP (deoxyuridine 5'-triphosphate); 150 nM of each primer; and 2 U of AmpliTaq Gold (Applied Biosystems). For each sample, one PCR was performed by adding 1 U of heat-labile uracil-DNA glycosylase and an initial incubation step of 15 min at 37°C to control for carryover contamination. Target regions were amplified with target-specific primer pairs that have been tagged with common sequence tags (CS1 and CS2, Fluidigm). After multiplex amplification, PCR products were sequenced using a combination of multiplexed Illumina MiSeq sequencing and traditional Sanger capillary electrophoresis. For Illumina sequencing, the amplicons were indexed using a maximum of 384 barcodes (containing Illumina adaptor sequences) in an additional amplification step and then pooled. To remove primer and adaptor dimers and other amplification products, we purified the multiplex PCR products using an AMPure PCR purification system (Agencourt) with a 1.8-fold ratio of SPRI (solid phase reversible immobilization) beads to reaction volume. The pooled and purified amplicon libraries were then sequenced on an Illumina MiSeq sequencer using CS1 and CS2 and their reverse complements as sequencing primers according to the manufacturer's instructions. For a few samples, the PCR products were generated with the same target-specific primer pairs and then sequenced on the ABI 3130xl Genetic Analyzer by applying the BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific).

### Processing and alignment of sequencing data

Traditional Sanger sequences were edited and aligned in Geneious v8.1 (Biomatters). Illumina reads were automatically demultiplexed by the Illumina MiSeq software and further processed by applying the following steps. Illumina sequencing adaptors were cut from the reads using cutadapt (34), and reads were then filtered for quality with Trimmomatic 0.33 (35), discarding all reads that contain more than 8 base pairs (bp) with a Phred score below 20 and allowing only reads with a minimum length of 50 bp. The trimmed and filtered paired-end reads were merged to a single read with FLASH (36), improving the overall quality. Using Bowtie 2 (37) and SAMtools 0.1.19 (38), these merged reads were mapped to the reference sequences (G72335.1, G72336.1, G72338.1, G72339.1, and AB091794.1) for the respective target regions. For each sample and locus, a consensus sequence was generated in Geneious v8.1 (Biomatters) from which the genotype of the Y chromosomal SNPs could be deduced. With these, SNP haplotype frequencies were calculated using the R package "pegas" v0.8-1 (39). To visualize the detected haplotypes during the different time periods, we constructed a heterochronous parsimony network using the R-script TempNet (40). We also tested for temporal changes in haplotype frequencies among four time bins (before 2200 BCE, 2200–900 BCE, 900 BCE–400 CE, and younger than 400 CE) by means of a temporal test of allele frequencies (41).

### Statistical analysis

We applied the sequential Markov Chain Monte Carlo (S-MCMC) method reported by Wutke *et al.* (42, 43) to estimate coefficients of selection that historically acted upon the Y chromosome haplotypes in horses. This method performs statistical inference by using computer simulations in a Bayesian MCMC framework. The use of simulations allowed taking into account the uncertainties associated to radiocarbon or stratigraphic ages of the ancient samples, sampling error, generation time, population sizes, and the change of generation, which was modeled

by a Wright-Fisher model (42, 43). Following previous work (8, 10, 19), the simulations also incorporated a growth in the horse population that started with the onset of domestication as in the studies of Wutke *et al.* (42, 43), but effective population sizes ( $N_e$ 's) and generation time were considered noise variables because of limited information in the data for estimating them. We tested alternative demographic histories with no population growth and an ample range of  $N_e$  values. Therefore, the estimated parameters were the selection coefficients, the age of the Y-HT-1 haplotype, and, alternatively, its initial haplotype frequency (see below). The simulations incorporated two alternative models addressing the initial state of Y-HT-1: (i) It was forced to be present since the beginning of the simulated time frame at a frequency that was a parameter or (ii) it was allowed to occur, emulating its appearance by mutation, at some point between the beginning of the simulated time frame and its first detection (the age became the parameter here). Both models were compared by means of Bayes factors. The MCMC was run under a Metropolis-Hasting algorithm incorporating a Gibbs sampler. The priors and details of the S-MCMC procedure followed those by Wutke *et al.* (42, 43). We ran four chains, 500,000 steps each, with a burn-in of 50%. We also replicated the estimation of selection coefficients by means of approximate Bayesian computation (44) and the path augmentation method (45). The statistical significance of the selection coefficients was assessed by obtaining the proportion of positive or negative values in the Bayesian posterior sample, for example, >95% of coefficients being positive could be interpreted as “significant” in a one-tailed test, and the same for negative values. In addition, we obtained Bayes factors for comparing a model with no selection (changes provided only by gene drift) and the model with selection upon Y-HT-1. Because selective forces might have changed over time (46), we performed a model choice analysis for selecting the number of time bins (with different selection coefficient each) that best fitted our data in the S-MCMC analysis. We tested from one to six time bins by means of both the AIC and the BIC. Time bins were defined considering main historical ages and sample sizes. In addition, considering that other haplotypes could be under selection, we replicated the inference by S-MCMC in a multiallelic mode, in which Wright-Fisher simulations were performed by multinomial, rather than binomial, samplings and several haplotypes were allowed to show selection. We assessed the number of haplotypes with signs of selection by means of AIC and BIC. Finally, we reported the temporal trajectories of haplotype frequencies of the Bayesian posterior sample to assess the changes in frequencies over time. The simulations were programmed in the language Fortran by means of the platform Microsoft Visual Studio and Intel Parallel Studio, using the code reported by Wutke *et al.* (42, 43), which is available at [www.dropbox.com/sh/uns1gsprj7xbz0/AAAEr4hfya16P5nejpXVV-nUfa?dl=0](http://www.dropbox.com/sh/uns1gsprj7xbz0/AAAEr4hfya16P5nejpXVV-nUfa?dl=0).

## SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/4/4/eaap9691/DC1>

- table S1. Additional information for all samples analyzed in this study.
- table S2. Genotyping results of Y chromosomal SNPs for successfully genotyped samples.
- table S3. Test of temporal changes in allele frequencies between the four time bins.
- table S4. Settings and operators of the MCMC procedure used for the final inference of selection coefficients and other parameters.
- fig. S1. Y-HT-1 showed very low allele frequency in early domestic horses.
- fig. S2. The three simulation approaches concur regarding a low initial Y-HT-1 frequency.
- fig. S3. All analyses yield strong support for positive selection of Y-HT-1.
- fig. S4. One selection coefficient for Y-HT-1 across all time periods was mostly supported.
- fig. S5. Several Y chromosome haplotypes were under selection.

## REFERENCES AND NOTES

1. B. Wallner, G. Brem, M. Müller, R. Achmann, Fixed nucleotide differences on the Y chromosome indicate clear divergence between *Equus przewalskii* and *Equus caballus*. *Anim. Genet.* **34**, 453–456 (2003).
2. G. Lindgren, N. Backström, J. Swinburne, L. Hellborg, A. Einarsson, K. Sandberg, G. Cothran, C. Vilà, M. Binns, H. Ellegren, Limited number of patrilineages in horse domestication. *Nat. Genet.* **36**, 335–336 (2004).
3. B. Wallner, F. Piumi, G. Brem, M. Müller, R. Achmann, Isolation of Y chromosome-specific microsatellites in the horse and cross-species amplification in the genus *Equus*. *J. Hered.* **95**, 158–164 (2004).
4. A. N. Lau, L. Peng, H. Goto, L. Chemnick, O. A. Ryder, K. D. Makova, Horse domestication and conservation genetics of Przewalski's horse inferred from sex chromosomal and autosomal sequences. *Mol. Biol. Evol.* **26**, 199–208 (2009).
5. B. Wallner, C. Vogl, P. Shukla, J. P. Burgstaller, T. Druml, G. Brem, Identification of genetic variation on the horse Y chromosome and the tracing of male founder lineages in modern breeds. *PLOS ONE* **8**, e60015 (2013).
6. C. Vilà, J. A. Leonard, A. Götherström, S. Marklund, K. Sandberg, K. Lidén, R. K. Wayne, H. Ellegren, Widespread origins of domestic horse lineages. *Science* **291**, 474–477 (2001).
7. T. Jansen, P. Forster, M. A. Levine, H. Delke, M. Hurler, C. Renfrew, J. Weber, K. Olek, Mitochondrial DNA and the origins of the domestic horse. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10905–10910 (2002).
8. M. Cieslak, M. Pruvost, N. Benecke, M. Hofreiter, A. Morales, M. Reissmann, A. Ludwig, Origin and history of mitochondrial DNA lineages in domestic horses. *PLOS ONE* **5**, e15311 (2010).
9. S. Lippold, N. J. Matzke, M. Reissmann, M. Hofreiter, Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. *BMC Evol. Biol.* **11**, 328 (2011).
10. A. Achilli, A. Olivieri, P. Soares, H. Lancioni, B. Hooshari Kashani, U. A. Perego, S. G. Nergadze, V. Carossa, M. Santagostino, S. Capomaccio, M. Felicetti, W. Al-Achkar, M. C. T. Penedo, A. Verini-Supplizi, M. Houshmand, S. R. Woodward, O. Semino, M. Silvestrelli, E. Giulotto, L. Pereira, H.-J. Bandelt, A. Torroni, Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 2449–2454 (2012).
11. P. Librado, C. Der Sarkissian, L. Ermini, M. Schubert, H. Jónsson, A. Albrechtsen, M. Fumagalli, M. A. Yang, C. Gamba, A. Seguin-Orlando, C. D. Mortensen, B. Petersen, C. A. Hoover, B. Lorente-Galdos, A. Nedoluzhko, E. Boulygina, S. Tsygankova, M. Neuditschko, V. Jagannathan, C. Thèves, A. H. Alfharhan, S. A. Alquraishi, K. A. S. Al-Rasheid, T. Sicheritz-Ponten, R. Popov, S. Grigoriev, A. N. Alekseev, E. M. Rubin, M. McCue, S. Rieder, T. Leeb, A. Tikhonov, E. Crubézy, M. Slatkin, T. Marques-Bonet, R. Nielsen, E. Willerslev, J. Kantanen, E. Prokhortchouk, L. Orlando, Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6889–E6897 (2015).
12. S. Lippold, M. Knapp, T. Kuznetsova, J. A. Leonard, N. Benecke, A. Ludwig, M. Rasmussen, A. Cooper, J. Weinstock, E. Willerslev, B. Shapiro, M. Hofreiter, Discovery of lost diversity of paternal horse lineages using ancient DNA. *Nat. Commun.* **2**, 450 (2011).
13. P. Librado, C. Gamba, C. Gaunitz, C. Der Sarkissian, M. Pruvost, A. Albrechtsen, A. Fages, N. Khan, M. Schubert, V. Jagannathan, A. Serres-Armero, L. F. K. Kuderna, I. S. Povolotskaya, A. Seguin-Orlando, S. Lepetz, M. Neuditschko, C. Thèves, S. Alquraishi, A. H. Alfharhan, K. Al-Rasheid, S. Rieder, Z. Samashev, H.-P. Francfort, N. Benecke, M. Hofreiter, A. Ludwig, C. Keyser, T. Marques-Bonet, B. Ludes, E. Crubézy, T. Leeb, E. Willerslev, L. Orlando, Ancient genomic changes associated with domestication of the horse. *Science* **356**, 442–445 (2017).
14. S. Mona, G. Catalano, M. Lari, G. Larson, P. Boscato, A. Casoli, L. Sineo, C. Di Patti, E. Pecchioli, D. Caramelli, G. Bertorelle, Population dynamic of the extinct European aurochs: Genetic evidence of a north-south differentiation pattern and no evidence of post-glacial expansion. *BMC Evol. Biol.* **10**, 83 (2010).
15. J. Eriksson, G. Larson, U. Gunnarsson, B. Bed'hom, M. Tixier-Boichard, L. Strömstedt, D. Wright, A. Jungerius, A. Vereijken, E. Randi, P. Jensen, L. Andersson, Identification of the *yellow skin* gene reveals a hybrid origin of the domestic chicken. *PLOS Genet.* **4**, e1000010 (2008).
16. O. Hanotte, D. G. Bradley, J. W. Ochieng, Y. Verjee, E. W. Hill, J. E. O. Rege, African pastoralism: Genetic imprints of origins and migrations. *Science* **296**, 336–339 (2002).
17. E. L. C. Verkaar, I. J. Nijman, M. Beeke, E. Hanekamp, J. A. Lenstra, Maternal and paternal lineages in cross-breeding bovine species. Has wisent a hybrid origin? *Mol. Biol. Evol.* **21**, 1165–1170 (2004).
18. A. K. Outram, N. A. Stear, R. Bendrey, S. Olsen, A. Kasparov, V. Zaibert, N. Thorpe, R. P. Evershed, The earliest horse harnessing and milking. *Science* **323**, 1332–1335 (2009).
19. B. Wallner, N. Palmieri, C. Vogl, D. Rigler, E. Bozlak, T. Druml, V. Jagannathan, T. Leeb, R. Fries, J. Tetens, G. Thaller, J. Metzger, O. Distl, G. Lindgren, C.-J. Rubin, L. Andersson, R. Schaefer, M. McCue, M. Neuditschko, S. Rieder, C. Schlotterer, G. Brem, Y chromosome uncovers the recent oriental origin of modern stallions. *Curr. Biol.* **27**, 2029–2035.e5 (2017).

20. 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).
21. K. Kristiansen, T. B. Larsson, *The Rise of Bronze Age Society: Travels, Transmissions and Transformations* (Cambridge Univ. Press, 2005).
22. C. Gamba, E. R. Jones, M. D. Teasdale, R. L. McLaughlin, G. Gonzalez-Fortes, V. Mattiangeli, L. Domboróczki, I. Kővári, I. Pap, A. Anders, A. Whittle, J. Dani, P. Raczky, T. F. G. Higham, M. Hofreiter, D. G. Bradley, R. Pinhasi, Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257 (2014).
23. P. Kelekna, *The Horse in Human History* (Cambridge Univ. Press, 2009).
24. P. F. Kuznetsov, The emergence of Bronze Age chariots in eastern Europe. *Antiquity* **80**, 638–645 (2006).
25. M. E. Allentoft, M. Sikora, K.-G. Sjogren, S. Rasmussen, M. Rasmussen, J. Stenderup, P. B. Damgaard, H. Schroeder, T. Ahlström, L. Vinner, A.-S. Malaspinas, A. Margaryan, T. Higham, D. Chival, L. Lynnerup, L. Harvig, J. Baron, P. Della Casa, P. Dąbrowski, P. R. Duffy, A. V. Ebel, A. Epimakhov, K. Frei, M. Furmanek, T. Gralak, A. Gromov, S. Gronkiewicz, G. Grupe, T. Hajdu, R. Jarysz, V. Khartanovich, A. Khokhlov, V. Kiss, J. Kolář, A. Kriška, I. Lasak, C. Longhi, G. McGlynn, A. Merkevicius, I. Merkyte, M. Metspalu, R. Mkrtychyan, V. Moiseyev, L. Paja, G. Pálfi, D. Pokutta, Ł. Pospieszny, T. D. Price, L. Saag, M. Sablin, N. Shishlina, V. Smrčka, V. I. Sionov, V. Szeverényi, G. Tóth, S. V. Trifanova, L. Varul, M. Vízce, L. Yepiskoposyan, V. Zhitenev, L. Orlando, T. Sicheritz-Pontén, S. Brunak, R. Nielsen, K. Kristiansen, E. Willerslev, Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
26. V. Warmuth, A. Eriksson, M. A. Bower, G. Barker, E. Barrett, B. K. Hanks, S. Li, D. Lomitashvili, M. Ochir-Goryaeva, G. V. Sizonov, V. Soyonov, A. Manica, Reconstructing the origin and spread of horse domestication in the Eurasian steppe. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 8202–8206 (2012).
27. S. L. Olsen, Early horse domestication on the Eurasian steppe, in *Documenting Domestication, New Genetic and Archaeological Paradigms*, M. A. Zeder, D. G. Bradley, E. Emshwiller, B. D. Smith, Eds. (University Presses of California, 2006), pp. 245–272.
28. L. K. Case, C. Teuscher, Y genetic variation and phenotypic diversity in health and disease. *Biol. Sex Differ.* **6**, 6 (2015).
29. J. Peters, *Römische Tierhaltung und Tierzucht: eine Synthese aus archäozoologischer Untersuchung und schriftlich-bildlicher Überlieferung* (Leidorf, 1998), vol. 5.
30. C. Gaunitz, A. Fages, K. Hanghøj, A. Albrechtsen, N. Khan, M. Schubert, A. Seguin-Orlando, I. J. Owens, S. Felkel, O. Bignon-Lau, P. de Barros Damgaard, A. Mittnik, A. F. Mohaseb, H. Davoudi, S. Alquraishi, A. H. Alfarhan, K. A. S. Al-Rasheid, E. Crubézy, N. Benecke, S. Olsen, D. Brown, D. Anthony, K. Massy, V. Pitulko, A. Kasparov, G. Brem, M. Hofreiter, G. Mukhtarova, N. Baimukhanov, L. Lóugas, V. Onar, P. W. Stockhammer, J. Krause, B. Boldgiv, S. Undrakhbold, D. Erdenebaatar, S. Lepetz, M. Mashkour, A. Ludwig, B. Wallner, V. Merz, I. Merz, V. Zaubert, E. Willerslev, P. Librado, A. K. Outram, L. Orlando, Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science* **360**, 111–114 (2018).
31. C. Der Sarkissian, L. Ermini, M. Schubert, M. A. Yang, P. Librado, M. Fumagalli, H. Jónsson, G. K. Bar-Gal, A. Albrechtsen, F. G. Vieira, B. Petersen, A. Ginolhac, A. Seguin-Orlando, K. Magnussen, A. Fages, C. Gamba, B. Lorente-Galdos, S. Polani, C. Steiner, M. Neuditschko, V. Jagannathan, C. Feh, C. L. Greenblatt, A. Ludwig, N. I. Abramson, W. Zimmermann, R. Schafberg, A. Tikhonov, T. Sicheritz-Pontén, E. Willerslev, T. Marques-Bonet, O. A. Ryder, M. McCue, S. Rieder, T. Leeb, M. Slatkin, L. Orlando, Evolutionary genomics and conservation of the endangered Przewalski's horse. *Curr. Biol.* **25**, 2577–2583 (2015).
32. H. Goto, O. A. Ryder, A. R. Fisher, B. Schultz, S. L. Kosakovsky Pond, A. Nekrutenko, K. D. Makova, A massively parallel sequencing approach uncovers ancient origins and high genetic variability of endangered Przewalski's horses. *Genome Biol. Evol.* **3**, 1096–1106 (2011).
33. J. Dabney, M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. Garcia, S. Pääbo, J.-L. Arsuaga, M. Meyer, Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758–15763 (2013).
34. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17**, 10–12 (2011).
35. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
36. T. Magoč, S. L. Salzberg, FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
37. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
38. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
39. E. Paradis, pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics* **26**, 419–420 (2010).
40. S. Prost, C. N. K. Anderson, TempNet: A method to display statistical parsimony networks for heterochronous DNA sequence data. *Methods Ecol. Evol.* **2**, 663–667 (2011).
41. E. Sandoval-Castellanos, Testing temporal changes in allele frequencies: A simulation approach. *Genet. Res.* **92**, 309–320 (2010).
42. S. Wutke, L. Andersson, N. Benecke, E. Sandoval-Castellanos, J. Gonzalez, J. H. Hallsson, L. Lóugas, O. Magnell, A. Morales-Muniz, L. Orlando, A. H. Páldóttir, M. Reissmann, M. B. Muñoz-Rodríguez, M. Ruttikay, A. Trinks, M. Hofreiter, A. Ludwig, The origin of ambling horses. *Curr. Biol.* **26**, R697–R699 (2016).
43. S. Wutke, N. Benecke, E. Sandoval-Castellanos, H.-J. Döhle, S. Friederich, J. Gonzalez, J. H. Hallsson, M. Hofreiter, L. Lóugas, O. Magnell, A. Morales-Muniz, L. Orlando, A. H. Páldóttir, M. Reissmann, M. Ruttikay, A. Trinks, A. Ludwig, Spotted phenotypes in horses lost attractiveness in the Middle Ages. *Sci. Rep.* **6**, 38548 (2016).
44. M. A. Beaumont, Approximate Bayesian computation in evolution and ecology. *Annu. Rev. Ecol. Syst.* **41**, 379–406 (2010).
45. J. G. Schraiber, S. N. Evans, M. Slatkin, Bayesian inference of natural selection from allele frequency time series. *Genetics* **203**, 493–511 (2016).
46. A. Ludwig, M. Reissmann, N. Benecke, R. Bellone, E. Sandoval-Castellanos, M. Cieslak, G. G. Fortes, A. Morales-Muñiz, M. Hofreiter, M. Pruvost, Twenty-five thousand years of fluctuating selection on leopard complex spotting and congenital night blindness in horses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130386 (2015).

**Acknowledgments:** We thank J. H. Hallsson and A. H. Páldóttir for providing Icelandic samples, D. Lieckfeldt and M. Pruvost for laboratory assistance, and D. Lenz for helping with next-generation sequencing data analyses. **Funding:** This work was funded by the Deutsche Forschungsgemeinschaft (LU 852/7-4). A.M.-M.'s research benefited from funding by projects HAR2014-55722-P and HAR2017-88325-P of the Spanish Ministry of Economy, Industry, and Competitiveness. A.-S.M. was financed by European Research Council (starting grant) and the Swiss National Science Foundation (Ambizione grant). **Author contributions:** A.L. designed the study. A.L., M.H., N.B., and M.R. planned the project. A.L., M.R., and S.W. designed the experiments. S.W. conducted the experiments. J.G. and A.T. performed replications. H.-J.D., S.F., L.L., O.M., A.M.-M., and L.O. provided samples, data, and information about the archeological/historical context. E.S.-C. and A.-S.M. ran statistical analyses. A.L., E.S.-C., S.W., and M.H. wrote the draft. All authors contributed to writing the final manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The code used for simulation calculations is available at [www.dropbox.com/sh/uns1gsprj7xbz0/AAAEr4hfya16P5nejpXVnUfa?dl=0](http://www.dropbox.com/sh/uns1gsprj7xbz0/AAAEr4hfya16P5nejpXVnUfa?dl=0). All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 15 September 2017

Accepted 1 March 2018

Published 18 April 2018

10.1126/sciadv.aap9691

**Citation:** S. Wutke, E. Sandoval-Castellanos, N. Benecke, H.-J. Döhle, S. Friederich, J. Gonzalez, M. Hofreiter, L. Lóugas, O. Magnell, A.-S. Malaspinas, A. Morales-Muñiz, L. Orlando, M. Reissmann, A. Trinks, A. Ludwig, Decline of genetic diversity in ancient domestic stallions in Europe. *Sci. Adv.* **4**, eaap9691 (2018).

## Decline of genetic diversity in ancient domestic stallions in Europe

Saskia Wutke, Edson Sandoval-Castellanos, Norbert Benecke, Hans-Jürgen Döhle, Susanne Friederich, Javier Gonzalez, Michael Hofreiter, Lembi Lõugas, Ola Magnell, Anna-Sapfo Malaspinas, Arturo Morales-Muñiz, Ludovic Orlando, Monika Reissmann, Alexandra Trinks and Arne Ludwig

*Sci Adv* 4 (4), eaap9691.  
DOI: 10.1126/sciadv.aap9691

### ARTICLE TOOLS

<http://advances.sciencemag.org/content/4/4/eaap9691>

### SUPPLEMENTARY MATERIALS

<http://advances.sciencemag.org/content/suppl/2018/04/16/4.4.eaap9691.DC1>

### REFERENCES

This article cites 41 articles, 12 of which you can access for free  
<http://advances.sciencemag.org/content/4/4/eaap9691#BIBL>

### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

---

*Science Advances* (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science Advances* is a registered trademark of AAAS.