

ARTICLE

Received 14 Apr 2011 | Accepted 20 Jul 2011 | Published 23 Aug 2011

DOI: 10.1038/ncomms1447

Discovery of lost diversity of paternal horse lineages using ancient DNA

Sebastian Lippold¹, Michael Knapp², Tatyana Kuznetsova³, Jennifer A. Leonard^{4,5}, Norbert Benecke⁶, Arne Ludwig⁷, Morten Rasmussen⁸, Alan Cooper⁹, Jaco Weinstock¹⁰, Eske Willerslev⁸, Beth Shapiro¹¹ & Michael Hofreiter^{1,12}

Modern domestic horses display abundant genetic diversity within female-inherited mitochondrial DNA, but practically no sequence diversity on the male-inherited Y chromosome. Several hypotheses have been proposed to explain this discrepancy, but can only be tested through knowledge of the diversity in both the ancestral (pre-domestication) maternal and paternal lineages. As wild horses are practically extinct, ancient DNA studies offer the only means to assess this ancestral diversity. Here we show considerable ancestral diversity in ancient male horses by sequencing 4 kb of Y chromosomal DNA from eight ancient wild horses and one 2,800-year-old domesticated horse. Both ancient and modern domestic horses form a separate branch from the ancient wild horses, with the Przewalski horse at its base. Our methodology establishes the feasibility of re-sequencing long ancient nuclear DNA fragments and demonstrates the power of ancient Y chromosome DNA sequence data to provide insights into the evolutionary history of populations.

¹ Research Group Molecular Ecology, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany. ² Allan Wilson Centre for Molecular Ecology and Evolution, Department of Anatomy and Structural Biology, University of Otago, Dunedin 9054, New Zealand. ³ Department of Palaeontology, Faculty of Geology, Moscow State University, 119991, Russia. ⁴ Conservation and Evolutionary Genetics Group, Estación Biológica de Doñana (EBD-CSIC), 41092 Seville, Spain. ⁵ Department of Evolutionary Biology, Uppsala University, 75236, Sweden. ⁶ Department of Eurasia, German Archaeological Institute, D-14195 Berlin, Germany. ⁷ Evolutionary Genetics, Leibniz Institute for Zoo and Wildlife Research, D-10252 Berlin, Germany. ⁸ Centre for GeoGenetics, Copenhagen University, 1350, Denmark. ⁹ Australian Centre for Ancient DNA, University of Adelaide, South Australia 5005, Australia. ¹⁰ School of Humanities (Archaeology), University of Southampton SO17 1BJ, UK. ¹¹ Department of Biology, The Pennsylvania State University, University Park, 16802, USA. ¹² Department of Biology, University of York, YO10 5DD, UK. Correspondence and requests for materials should be addressed to S.L. (email: sebastian_lippold@eva.mpg.de).

The Y chromosome is a valuable tool in population genetics, as it provides a means to directly assess evolutionary processes that only affect the paternal lineage. The use of Y chromosome data in population genetic analyses became widely established in reconstructions of human evolutionary relationships and demographic processes (for a review see ref. 1). However, very few Y chromosome studies have focused on non-model taxa. This may in part be due to challenges associated with developing Y chromosomal markers which include a proliferation of repeat elements and the low genetic diversity characteristic of the Y chromosome². In a recent comparative analysis of five mammalian species, two (wolf and field vole) show low levels of nucleotide diversity on the Y chromosome (π_Y 0.4×10^{-4} and 1.7×10^{-4} , respectively), while the other three (lynx, reindeer and cattle) had no diversity at all³. Low Y chromosomal genetic diversity was also observed in sheep⁴, cattle^{3,5} and dogs^{6,7}. These results suggest a general pattern of low male effective population size in domestic mammals, which may be attributable to breeding practices associated with domestication, where few males are selected to mate with a wider variety of females.

One of the most extreme examples of contrasting levels of genetic diversity between maternal and paternal markers is the domestic horse. Using microsatellite data⁸ and up to 14.3 kb of sequence data from 52 individuals representing 15 breeds, all but one investigation has failed to detect any diversity on the Y chromosome of modern horses^{8–11}. In the study that did detect diversity, a single polymorphic microsatellite was reported from a sample of domestic Chinese horses¹². In contrast, the maternally inherited mitochondrial genome shows abundant diversity both among and within horse breeds with limited, to no, correlation between breeds and mitochondrial DNA haplotypes^{13–17}.

Several hypotheses have been proposed to explain the contrasting mitochondrial and Y chromosome diversity in domestic horses. First, the number of domestic founders may have differed between the sexes, with a small number of males (low male effective population size)^{8–11} and a larger number of females, the latter possibly originating from multiple geographical regions^{14,15}. Second, reproductive success among males may be strongly skewed because of the naturally polygamous mating system of horses^{18,19} or resulting from a breeding scheme imposed by humans during or after domestication, where a few select studs were preferentially mated with many mares^{20,21}. Third, as generally suggested for uniparentally inherited sex chromosomes³, selection may have eliminated genetic diversity on horse Y chromosomes due to either purifying background selection or selective sweeps caused by positive selection. These hypotheses are not mutually exclusive, and multiple forces may have operated together to eliminate variability on the domestic horse Y chromosome.

Unfortunately, almost no wild horses remain; the only surviving wild horse population is a small captive stock of Przewalski horses, which represent the closest living relatives of domesticated horses. Notably, Przewalski horses experienced an extreme population

bottleneck during the last century; the captive stock was founded by only eight females and five males, and hybridization with domesticated horses cannot be excluded²². Therefore, DNA amplified from ancient remains provides the only means to investigate the extent and nature of the genetic diversity of wild horses. Thus far, this approach has been used only for mtDNA^{13,14,16,23–25} the results of which suggest that domestication was not a significant bottleneck for horse mtDNA diversity²⁶. Consequently, high mtDNA diversity in domestic horses has been explained by high diversity of the founding population, multiple origins of domestication, further domestication events during the Iron Age, and backcrossing with wild mares from different populations.

In contrast to mtDNA, the technical challenges associated with large-scale targeted re-sequencing of ancient nuclear DNA have so far prevented studies of the Y chromosomal diversity of past horse populations. For example, the high copy-number of mtDNA per cell has facilitated its use in ancient DNA analyses, as the probability that fragments survive over time is greater simply because of the larger number of starting molecules. It has recently been shown that the ratio of autosomal DNA to mtDNA increases from ~1:152 in modern tissues to 1:245–17,480 in ancient tissues, most likely owing to differential preservation²⁷. For the Y chromosome, this ratio decreases by another factor of two. In addition to fewer starting molecules, the expected low diversity in the Y chromosome means that longer regions of the Y chromosome need to be sequenced to observe sufficient polymorphisms for analysis. Further, Y chromosomes can only be found in the remains of male individuals, and the sex of the remains is generally unknown.

In this study, we successfully amplified 4 kb of Y chromosome DNA from nine ancient horse specimens, including one 2,800-year-old domesticated horse. This represents the first ancient Y chromosome re-sequencing dataset to date. Using these data, we investigated Y chromosome diversity in pre-domestication ancient wild-horse populations, and compared the results with what is known about Y-chromosome diversity in modern domestic and Przewalski's horses. We found that the ancient horses harboured considerable Y-chromosome diversity.

Results

Sequencing of ancient DNA. We sequenced a total of 4,062 bp of Y-chromosomal DNA from each of eight wild horses from permafrost sites in Siberia and North America, and one 2,800-year-old domestic stallion (see Table 1 and Fig. 1). Including the six sites that differ between modern Przewalski's horse and modern domestic horses, we identified 28 segregating sites among all sequenced horses (Supplementary Table S1). Each of our ancient horses carried a unique haplotype with pairwise sequence differences among individuals ranging from 1 to 16 substitutions (Table 2). All sequence positions were replicated at least twice, excluding ancient DNA damage as a possible cause for the polymorphic positions observed. Nucleotide diversity (π_Y) was estimated to be 1.89×10^{-3}

Table 1 | Information about the ancient samples amplified for the full 4 kb Y chromosomal DNA sequence.

Sample	Figure abbreviation	Location	C ₁₄ date	Latitude	Longitude
ARZ-1-3	1	Arzan, South Siberia	2,800 bp	52.1	93.6
JAL-292	2	Lower Goldstream, Alaska	NA	64.5	-147.4
JAL-310	3	Chatom, Alaska	NA	57.5	-134.9
YG 109.6	4	Quartz Creek, Yukon	> 47,000 bp	63.7	-139.1
MGV01_niche3.3	5	Bluefish Cave, Yukon	NA	67.3	-139.4
BL-O ₄₈₅	6	Bol'shoy Lyakhovsky Island, Siberia	26,500 ± 600 bp	73.3	141.3
BL-O ₂₅₀	7	Bol'shoy Lyakhovsky Island, Siberia	16,800 ± 70 bp	73.3	141.3
BL-O ₇₂₈	8	Bol'shoy Lyakhovsky Island, Siberia	> 44,000 bp	73.3	141.3
ML-O ₁₁₂	9	Malyi Lyakhovsky Island, Siberia	39,460 ± 400 bp	74.1	141.3

(standard deviation (s.d.) 3.00×10^{-4}) in the wild horses including the Przewalski horse. Since nucleotide diversity was found to be zero in modern domestic horse Y chromosome sequences¹⁰, and because demographic processes between the timing of domestication and today remain unknown, it is not possible to accurately estimate π_Y for domestic horses. As an approximation, we calculated π_Y for modern domestic horses plus the single ancient domestic haplotype sequenced as part of this study to be 4.00×10^{-5} (s.d. 4.00×10^{-5}).

Phylogenetic relationship of Y chromosome haplotypes. The phylogenetic relationship among all 62 available horse Y chromosome haplotypes (nine from our study plus 52 modern horses and the Przewalski horse haplotype from ref. 10) is depicted in a median joining network (Fig. 2). The haplotype of the 2,800-year-old domestic horse is most similar to that of modern horses, differing by four substitutions. The ancient horses cluster into three branches in the network: one consists exclusively of North American samples, one consists of a single Siberian sample, and the third one shares haplotypes from both North America and Siberia but is dominated by Siberian haplotypes. The Przewalski horse is basal to the domestic lineage, and shares a 4-bp deletion with domesticated horses that is not found in any ancient wild horse.

Incorporating temporally sampled data may artificially increase observed diversity, if the mutation rate is fast relative to the temporal span of the sequences. Although the ancient horses investigated lived during different time periods (ranging from >47ky–2.8ky years, Table 1), the temporal distribution of our samples does not seem to inflate our diversity estimates, as no correlation appears

between the number of pairwise substitutions and the age of the samples (Spearman correlation coefficient, P -values based on exact matrix permutation: $r = -0.033$, $P = 0.942$). This pattern is maintained after swapping the dates of the two infinitely dated samples ($r = -0.152$, $P = 0.658$). Further, we performed a molecular-clock based phylogenetic analysis both to estimate the age of the most recent common ancestor of all of the Y chromosome haplotypes and to determine when the various lineages diverged (Fig. 3). The time to the most recent common ancestor of all Y chromosomal horse haplotypes is 92–380 ka bp, with a mean of 208 ka bp. The shape of the MCMC genealogy indicates that most of the Y chromosome lineages emerged before the age of the oldest sample (53,800 years BP). As only two Y chromosome lineages persist today (the modern domestic lineage and the Przewalski lineage) this suggests a significantly higher diversity in the past.

Discussion

The relationship of Przewalski's horse to modern domestic horses remains controversial^{9,11,17,28–30}. Przewalski horses are generally viewed as either the last surviving wild-horse population, or a feral-horse population derived from a primitive domestic lineage. The issue is confounded by a recent population bottleneck²² that is likely to have reduced the genetic diversity within Przewalski horses significantly.

Today, two mtDNA haplotypes are found in Przewalski horses, and neither of these is present in modern horses. It has been proposed therefore that Przewalski horses are not ancestral to modern domestic horses^{23,30}. However, the Przewalski haplotypes do fall within the large diversity of modern horse mtDNA^{14,15,17,25}. A similar pattern was shown for autosomal DNA^{11,31} and X chromosomal sequences¹¹, where it was not possible to separate the Przewalski horses phylogenetically from domestic horses, although differences in the autosomal and X chromosomal nucleotide diversity in both taxa indicate a different evolutionary history¹¹. Our results indicate that the single Przewalski's horse Y chromosome haplotype^{9,10} falls within the greater Y chromosomal diversity of domestic and ancient wild horses. Interestingly, the Przewalski Y chromosome haplotype is more closely related to the two domestic horse haplotypes in our data set than any of the ancient wild horses. Thus, in agreement with the other genetic markers, the Y chromosome data presented here supports historic isolation, but, at the same time, a close evolutionary relationship between domestic horse and Przewalski's horse. All 52 domestic horses that have been sequenced to date, representing 15 modern horse breeds, have identical Y chromosome haplotypes¹⁰. One hypothesis to explain this suggests that modern horses have little Y chromosome diversity because the wild horses from which they were domesticated were also not diverse, due in part to the harem mating system in horses, implying skewed reproductive success of males¹⁹.

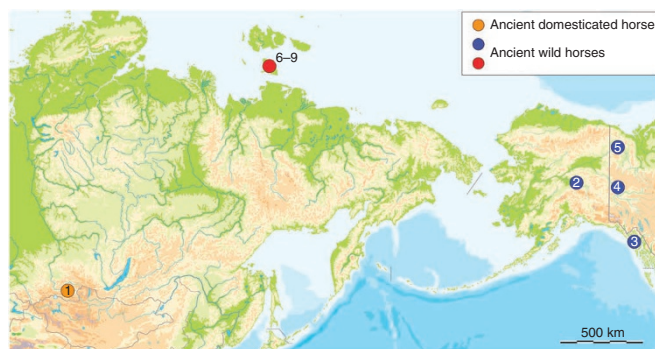


Figure 1 | Geographic origin of the ancient horse samples. The continental-scale origin of the ancient wild horses is further emphasized in red (Eurasia) and blue (North America). Sample Abbreviations: 1 = ARZ-1-3, 2 = JAL-292, 3 = JAL-310, 4 = YG 109.6, 5 = MGVo1_niche3.3, 6 = BL-O₄₈₅, 7 = BL-O₂₅₀, 8 = BL-O₇₂₈, 9 = ML-O₁₁₂.

Table 2 | Number of pairwise nucleotide differences among all horse sequences.

	<i>E. caballus</i>	1	2	3	4	5	6	7	8	9
<i>E. caballus</i>										
1 ARZ-1-3	4									
2 JAL-292	15	11								
3 JAL-310	13	9	2							
4 YG109.6	12	8	3	1						
5 MGVo1_niche3.3	11	7	10	8	9					
6 BL-O ₄₈₅	11	7	12	10	9	8				
7 BL-O ₂₅₀	13	9	16	14	13	8	12			
8 BL-O ₇₂₈	9	5	10	8	7	2	6	8		
9 ML-O ₁₁₂	11	7	12	10	9	2	8	8	2	
<i>E. przewalskii</i>	6	3	8	6	5	4	4	8	2	4

E. caballus denotes the single haplotype found in 52 modern horses.

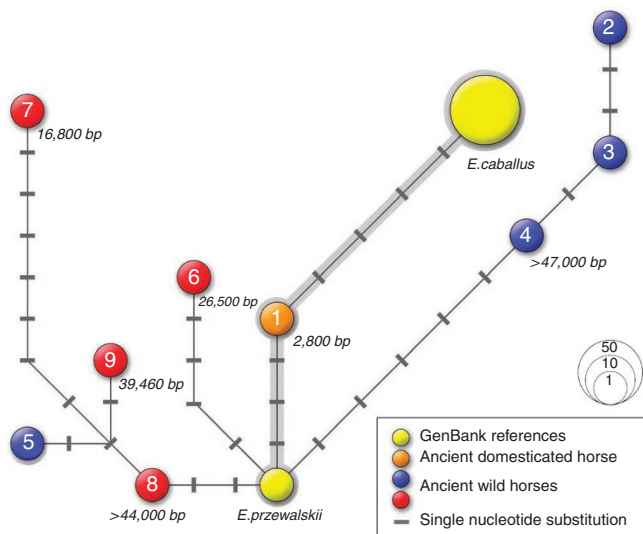


Figure 2 | Median-Joining network based on 4,062 bp Y chromosomal sequence. The continental-scale origin of the ancient wild horses is shown by different colours (red: Eurasia; blue: North America). The 2,800-year-old domestic horse haplotype is shown in orange and sequences retrieved from NCBI GenBank in yellow. Haplotypes sharing the 4bp deletion are shaded in grey. Sample Abbreviations: 1 = ARZ-1-3, 2 = JAL-292, 3 = JAL-310, 4 = YG 109.6, 5 = MGVo1_niche3.3, 6 = BL-O₄₈₅, 7 = BL-O₂₅₀, 8 = BL-O₇₂₈, 9 = ML-O₁₁₂.

Our results reject this hypothesis, suggesting instead that the Y chromosome diversity estimated from ancient wild horses (π_Y 1.89×10^{-3}) is high, and particularly high in comparison to that estimated previously for other wild mammals (for example, European rabbit π_Y 1.34×10^{-3} (ref. 32), wild boar π_Y 0.98×10^{-3} (ref. 33), felidae π_Y $0-0.995 \times 10^{-3}$ (ref. 34) and wolf π_Y 0.04×10^{-3} (ref. 3)). Although it is difficult to directly compare absolute values of diversity among different species, these numbers show that ancient wild horses harboured substantial genetic diversity on the Y chromosome. Because we sample over a window of time rather than within a single time-frame, the diversity measurements may be artificially inflated if new mutations arise during the sampling period. However, the age range of the samples from which our data are derived is small relative to the mutation rate of the Y chromosome. We therefore expect few if any novel mutations to arise during this period, and little influence on the diversity estimate.

The abundant Y chromosomal diversity found in wild horses is in stark contrast to the complete lack of variability in modern horses. This result argues against the absence of Y chromosomal diversity in modern horses being based on properties intrinsic to wild horses, such as continuous strong selection on the Y chromosome or a strong reproductive skew among males.

Our results therefore support the hypothesis that the lack of genetic diversity in extant horses may be a consequence of the domestication process. This loss of diversity at domestication may have been achieved either through the incorporation of very few wild male horses in the domestic stocks⁸⁻¹¹, a global selective sweep of the Y chromosome⁸, or breeding practices developed after domestication that reduced the effective number of males in the domestic species^{20,21}. The first hypothesis predicts that low levels of Y chromosome diversity will be found in all historic and prehistoric domestic horses. The second and third hypotheses both predict high Y chromosome genetic diversity in early domestic horses followed by a decrease to modern/near the modern very low level of diversity.

The single, domesticated horse sequence in our data set originates from a Scythian tomb and dates to 2,800 years BP. Artefacts

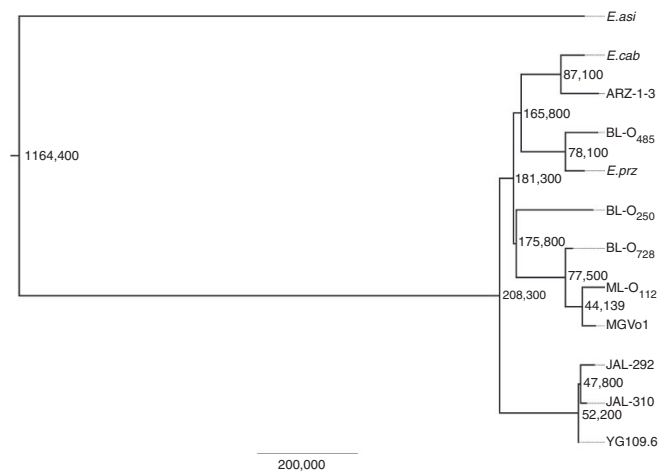


Figure 3 | Chronogram of horse Y chromosome evolution. Topologies and divergence time estimates (node labels) for the respective branches were inferred using BEAST v1.6.0 (refs 51,52) under the settings described in the main text. One donkey (*E. asinus*) sequence was used as an outgroup.

recovered from the same site from which the specimen originates have been associated with riding, and show direct evidence of domestication³⁵. This sample shows a haplotype that is closely related to, but distinct from the modern haplotype, from which it differs by four substitutions. Given the relatively young age of the sample and the estimated substitution rate of 0.85% per million years, it is unlikely that the haplotype found in the Scythian horse is a direct ancestor of the haplotype that characterizes all sequenced modern horses. Although data from a single ancient domesticated horse is not conclusive, it does show that more genetic variation existed within domestic horses 2,800 years ago than which exists today. However, the single sample cannot distinguish between breeding practices or a global selective sweep as the cause of the eventual complete loss of genetic diversity in domestic horses. To characterize both the initial level of Y chromosomal diversity in domestic horses and the processes by which this was lost, it will be necessary to obtain data from both early domestic horses, such as those from Botai^{36,37}, as well as from later periods such as the Iron age or Medieval times, ideally in combination with mitochondrial and autosomal sequence data.

So far, ancient DNA studies comparing homologous, replicated sections of DNA from multiple individuals have been mostly limited to mitochondrial DNA. Although nuclear DNA sequences from three Neanderthal specimens have been published recently³⁸, these were obtained by low coverage shotgun sequencing, an approach that is not generally scalable to address population genetic questions. However, our results show that by using a regular two-step multiplex PCR, it is possible to obtain nuclear and even Y chromosomal DNA data sets suitable for population studies.

We found substantial genetic diversity among ancient horse Y chromosomal sequences, demonstrating that wild horses exhibited Y chromosomal diversity before domestication. The single 2,800-year-old domestic horse suggests that some level of Y chromosomal diversity still existed in domestic horses several thousands of years after domestication, although the lineage identified was closely related to the modern domestic lineage. These results clearly demonstrate both the feasibility and power of ancient Y chromosomal DNA sequence data to reveal past population processes and provide a more complete picture based on the history of both sexes.

Methods

DNA extraction. We extracted DNA from 90 ancient horse samples from Eurasia and North America (Supplementary Table S2). To prepare the bone samples for extraction, we first cleaned the exterior surface of the bone using a Dremel tool to

remove any potential surface contaminants. We then removed a 100–250 mg bone sample, which we pulverized with a mortar and pestle. We extracted DNA from the powder using the silica-based method described in ref. 39.

Primer design. To investigate Y chromosome diversity in ancient and modern horses, we selected nine fragments within the noncoding regions of the *Equus caballus* Y chromosome reference sequence (Supplementary Table S3). Five of these were first described in⁹. In addition to these five regions, we selected introns 1–3 of the amelogenin gene and the 3' untranslated region of the *SRY* gene. All of those nine fragments were sequenced in 52 horses and the Przewalski horse¹⁰. Primer3 software (<http://frodo.wi.mit.edu/>) was used to design 88 primer pairs spanning the target region of ~4 kb (Supplementary Table S4). Each of these 88 fragments was then compared with the Horse Genome (<http://www.broadinstitute.org/mammals/horse31>) using a BLAT Search (<http://www.genome.ucsc.edu/cgi-bin/hgBlat41>). We excluded fragments that match with at least 50 bp and more than 80% identity to non Y chromosomal regions to reduce the probability of amplifying non Y chromosomal fragments of similar length and sequence. To test the multiplexing suitability and male-specific amplification of our primer set, we performed an initial PCR test using DNA from one modern male and female.

PCR amplification and sequencing. Immediately following extraction, we amplified one mitochondrial, one X-specific and one Y-specific fragment (Supplementary Table S5) to test the extracts for DNA preservation and to identify male horses (Supplementary material sex test). On the basis of amplification success and on a selection strategy to optimize the geographic distribution of samples, we selected twelve samples for further analyses (Supplementary Table S2).

We used a two-step multiplex PCR⁴² to amplify the 88 fragments (Supplementary Table S4) from these 12 ancient samples. Extraction blanks were amplified for all of the 88 fragments to check for contamination. Further, all 88 fragments of each sample were amplified twice, starting from independent first step reactions, to detect potential DNA damage patterns.

First-step PCR amplifications were performed in 25 µl reactions with 5 µl of DNA extract, 2 U AmpliTaq Gold DNA Polymerase and 1×buffer (Invitrogen), 1 mg ml⁻¹ BSA (Sigma-Aldrich), 4 mM MgCl₂, 250 µM of each dNTP, and 0.15 µM of each primer set (odd and even; Supplementary Table S3). Second-step PCR amplifications were performed in 25 µl reactions with 5 µl of 1:50 diluted first step PCR product, 0.1 U AmpliTaq Gold DNA Polymerase and 1×buffer (Invitrogen), 1 mg ml⁻¹ BSA (Sigma-Aldrich), 4 mM MgCl₂, 250 µM of each dNTP, and 1.5 µM of specific primer. The PCR thermal cycling conditions in the first-step multiplex PCR consisted of 95 °C for 12 min, followed by 35 cycles of denaturation at 94 °C for 20 s, annealing at 56 °C for 30 s and extension at 72 °C for 30 s, followed by a final extension at 72 °C for 4 min. For second-step PCR annealing temperatures, see Supplementary Table S4. After second amplification, all extraction blank fragments and 16 (out of 88) randomly chosen fragments of each sample were loaded onto 2% agarose gels to test for clean controls and amplification success, respectively. Because the three non-permafrost horses chosen showed a low amplification success rate, for these samples, all 88 fragments were checked on a gel. The pattern of low success rate persisted, and these three samples were excluded from further processing.

PCR products of the nine remaining samples (Table 1) were purified using the Agencourt Ampure PCR purification kit, following the manufacturer's protocol, with some modifications that result in a fragment-length-specific cutoff during purification⁴³. The purified products were quantified using a PicoGreen plate read on a Stratagene MX 3005P QPCR System. On the basis of this quantification, all 88 fragments per sample were normalized and pooled. The sample-specific fragment pools were barcoded for 454 high throughput sequencing using the methods described in^{43,44}. After qPCR based quantification⁴⁵, up to six barcoded sample-specific fragment pools (6×88 fragments) were sequenced on 1/16 of a 454 GS FLX run.

454 FLX Data processing and analysis. Sequence reads were sorted based on their specific bar code using the program untag (<http://bioinf.eva.mpg.de/pts/43>). Individual fragments were identified using demultiplex (<http://bioinf.eva.mpg.de/pts/43>). Demultiplex searches for target primer sequences within untagged sequences, thereby identifying the reads. All reads containing the target specific 3' and 5' priming site and having a minimum of 85% identity to the reference were aligned to the target reference sequence. The consensus sequence for each fragment was called according to a 66% majority rule for all fragments, for which at least three reads per replicate were observed. The sequence data for both replicates from each sample were aligned, a consensus sequence was called and single fragments were merged, resulting in a total of 4,062 bp of sequence for each individual. For fragments with positions that differ in the two consensus replicates, we performed a third PCR, and a consensus sequence was called according to majority rule. Finally, an independent replication for all fragments containing polymorphic positions was performed in Copenhagen for the sample MGVo1_niche3.3. As all positions were replicated at least twice from independent PCR amplifications, we can rule out ancient DNA damage as a cause for any sequence variation observed among the obtained haplotypes.

We then identified polymorphic positions based on a comparison of the complete 4,062 bp Y chromosome alignment of our nine ancient horses, the modern

E. caballus and the *E. przewalskii* haplotypes (Supplementary Table S3). We calculated a distance matrix showing the number of pairwise nucleotide differences among individuals using MEGA v4 (<http://www.megasoftware.net/>)⁴⁶. We used DnaSP v5.10 to calculate nucleotide diversity (π) (<http://www.ub.es/dnasp/>)⁴⁷. A median joining network was constructed using the software package Network 4.5 (<http://www.fluxus-engineering.com/>)⁴⁸.

As the ancient samples are from different time periods (Table 1), we then tested for a correlation between the number of pairwise substitutions and the temporal differences between the ¹⁴C dated samples to determine whether our diversity estimates were biased by age differences among the samples. We conducted a Spearman's rank correlation with *P*-values based on exact matrix permutation in R (version 2.10.0⁴⁹). As the two samples associated with infinite radiocarbon ages (YG109.6, BL-O₇₂₈) could be incorrectly ranked, their minimum ages were switched and the test performed again.

Using the Akaike information criterion implemented in MODELTEST 3.7⁵⁰, we identified GTR + I as the best fitting nucleotide substitution model for our alignment of the nine ancient horses and the three previously published Y chromosome sequences (Supplementary Table S3). Bayesian phylogenetic and molecular clock analyses were then performed using BEAST v1.6.0 (refs 51,52) under the GTR + I model and assuming a strict molecular clock. To determine the best fitting coalescent model, marginal likelihoods were compared using Bayes Factors⁵³ between constant-size coalescent, an exponential growth, an expansion growth and a Bayesian skyline plot model, the latter allowing a flexible model of past population dynamics⁵⁴ (Supplementary Table S6). For each analysis, we ran three MCMC chains of 10,000,000 iterations with trees and model parameter values sampled from the posterior distribution every 1,000th iteration. For each analysis, the first 10% were discarded from each run as burn-in, and the remainder combined. Convergence of the chains and effective sample sizes were verified using the program TRACER v1.5.0. The constant size model fit the data better than the more complex exponential growth and Bayesian skyline plot models and only marginally worse than the expansion growth model (log₁₀ BF: -0.053). As this is no decisive difference (decisive = log₁₀ BF > 2 (ref. 55)), the constant population size model was assumed to provide the best fit for the data.

To estimate divergence times of the different haplotypes a final BEAST analysis was performed, in which evolutionary and coalescent model parameters were as for the best-fitting model above, but samples for which no radiocarbon date (JAL-292, JAL-310, MGVo1_niche3.3) or only a lower bound (infinite radiocarbon dates; BL-O₇₂₈, YG 109.6) was available were also included by sampling their ages from a predefined distribution⁵⁶. For the undated sequences, we sample from a lognormal distribution with 95% CIs between 600 and 80,000 years, and the weight of the sample density around 22,000 years. For the infinitely dated samples, the 95% CIs include the range 30,000–80,000 years, and the weight of the sample density is concentrated around 52,000 years. A further calibration was incorporated at the time of divergence between *E. asinus*, and the remaining lineages: We used a lognormal prior sampling between 1.0 and 5.5 myrs; these confidence intervals incorporate both the fossil record age estimates^{57–59} and previous divergence estimates based on molecular data⁶⁰. The results of the tip-dating analysis are shown in Supplementary Table S7.

References

1. Jobling, M. A. & Tyler-Smith, C. The human Y chromosome: an evolutionary marker comes of age. *Nat. Rev. Genet.* **4**, 598–612 (2003).
2. Greminger, M. P., Krutzen, M., Schelling, C., Pienkowska-Schelling, A. & Wandeler, P. The quest for Y-chromosomal markers - methodological strategies for mammalian non-model organisms. *Mol. Ecol. Resour.* **10**, 409–420 (2010).
3. Hellborg, L. & Ellegren, H. Low Levels of Nucleotide Diversity in Mammalian Y Chromosomes. *Mol. Biol. Evol.* **21**, 158–163 (2004).
4. Meadows, J. R. S., Hawken, R. J. & Kijas, J. W. Nucleotide diversity on the ovine Y chromosome. *Anim. Genet.* **35**, 379–385 (2004).
5. Götherström, A. *et al.* Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe. *Proc. R. Soc. Lond. Ser. B-Biol. Sci.* **272**, 2345–2350 (2005).
6. Natanaelsson, C. *et al.* Dog Y chromosomal DNA sequence: identification, sequencing and SNP discovery. *BMC Genet.* **7**, 45 (2006).
7. Sundqvist, A. K. *et al.* Unequal contribution of sexes in the origin of dog breeds. *Genetics* **172**, 1121–1128 (2006).
8. Wallner, B., Piumi, F., Brem, G., Müller, M. & Achmann, R. Isolation of Y Chromosome-specific Microsatellites in the Horse and Cross-species Amplification in the Genus *Equus*. *J. Hered.* **95**, 158–164 (2004).
9. Wallner, B., Brem, G., Müller, M. & Achmann, R. Fixed nucleotide differences on the Y chromosome indicate clear divergence between *Equus przewalskii* and *Equus caballus*. *Anim. Genet.* **34**, 453–456 (2003).
10. Lindgren, G. *et al.* Limited number of patrilineal lineages in horse domestication. *Nature Genet.* **36**, 335–336 (2004).
11. Lau, A. N. *et al.* Horse Domestication and Conservation Genetics of Przewalski's Horse Inferred from Sex Chromosomal and Autosomal Sequences. *Mol. Biol. Evol.* **26**, 199–208 (2009).
12. Ling, Y. *et al.* Identification of Y Chromosome Genetic Variations in Chinese Indigenous Horse Breeds. *J. Hered.* **101**, 639–643 (2010).

13. Lister, A. M. *et al.* Ancient and modern DNA in a study of horse domestication. *Ancient Biomolecules* **2**, 267 (1998).
14. Vilà, C. *et al.* Widespread Origins of Domestic Horse Lineages. *Science* **291**, 474–477 (2001).
15. Jansen, T. *et al.* Mitochondrial DNA and the origins of the domestic horse. *Proc. Natl. Acad. Sci. USA* **99**, 10905–10910 (2002).
16. McGahern, A. M. *et al.* Mitochondrial DNA sequence diversity in extant Irish horse populations and in ancient horses. *Anim. Genet.* **37**, 498–502 (2006).
17. Kavar, T. & Dovc, P. Domestication of the horse: Genetic relationships between domestic and wild horses. *Livestock Science* **116**, 1–14 (2008).
18. Emlen, S. T. & Oring, L. W. Ecology sexual selection, and evolution of mating systems. *Science* **197**, 215–223 (1977).
19. Asa, C. S. Male reproductive success in free-ranging feral horses. *Behav. Ecol. Sociobiol.* **47**, 89–93 (1999).
20. Levine, M. A. Botai and the origins of horse domestication. *J. Anthropol. Archaeol.* **18**, 29–78 (1999).
21. Cunningham, E. P., Dooley, J. J., Splan, R. K. & Bradley, D. G. Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses. *Anim. Genet.* **32**, 360–364 (2001).
22. Volf, J. K. E. & Prokopová, L. *General studbook of the Przewalski horse* (Zoological Garden Prague, 1991).
23. Cai, D. W. *et al.* Ancient DNA provides new insights into the origin of the Chinese domestic horse. *J. Archaeol. Sci.* **36**, 835–842 (2009).
24. Lei, C. Z. *et al.* Multiple maternal origins of native modern and ancient horse populations in China. *Anim. Genet.* **40**, 933–944 (2009).
25. Cieslak, M. *et al.* Origin and history of mitochondrial DNA lineages in domestic horses. *PLoS One* **5**, e15311 (2010).
26. Lira, J. *et al.* Ancient DNA reveals traces of Iberian Neolithic and Bronze Age lineages in modern Iberian horses. *Mol. Ecol.* **19**, 64–78 (2010).
27. Schwarz, C. *et al.* New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. *Nucleic Acids Res.* **37**, 3215–3229 (2009).
28. Bowling, A. T. *et al.* Genetic variation in Przewalski's horses, with special focus on the last wild caught mare, 231 Orlitza III. *Cytogenet. Genome Res.* **102**, 226–234 (2003).
29. Ishida, N., Oyunsuren, T., Mashima, S., Mukoyama, H. & Saitou, N. Mitochondrial-DNA sequences of various species of the genus equus with special reference to the phylogenetic relationship between Przewalski's wild horse and domestic horse. *J. Mol. Evol.* **41**, 180–188 (1995).
30. Oakenfull, E. A. & Ryder, O. A. Mitochondrial control region and 12S rRNA variation in Przewalski's horse (*Equus przewalskii*). *Anim. Genet.* **29**, 456–459 (1998).
31. Wade, C. M. *et al.* Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse. *Science* **326**, 865–867 (2009).
32. Geraud, A., Rogel-Gaillard, C. & Ferrand, N. High levels of nucleotide diversity in the European rabbit (*Oryzolagus cuniculus*) SRY gene. *Anim. Genet.* **36**, 349–351 (2005).
33. Ramirez, O. *et al.* Integrating Y-chromosome, mitochondrial, and autosomal data to analyze the origin of pig breeds. *Mol. Biol. Evol.* **26**, 2061–2072 (2009).
34. Luo, S. J. *et al.* Development of Y chromosome intraspecific polymorphic markers in the Felidae. *J. Hered.* **98**, 400–413 (2007).
35. Keyser-Tracqui, C. *et al.* Mitochondrial DNA analysis of horses recovered from a frozen tomb (Berel site, Kazakhstan, 3rd Century BC). *Anim. Genet.* **36**, 203–209 (2005).
36. Outram, A. K. *et al.* The earliest horse harnessing and milking. *Science* **323**, 1332–1335 (2009).
37. Ludwig, A. *et al.* Coat color variation at the beginning of horse domestication. *Science* **324**, 485–485 (2009).
38. Green, R. E. *et al.* A draft sequence of the neandertal genome. *Science* **328**, 710–722 (2010).
39. Rohland, N., Siedel, H. & Hofreiter, M. A rapid column-based ancient DNA extraction method for increased sample throughput. *Mol. Ecol. Resour.* **10**, 677–683 (2010).
40. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).
41. Kent, W. J. BLAT - The BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
42. Römler, H. *et al.* Multiplex amplification of ancient DNA. *Nat. Protoc.* **1**, 720–728 (2006).
43. Meyer, M., Stenzel, U. & Hofreiter, M. Parallel tagged sequencing on the 454 platform. *Nat. Protoc.* **3**, 267–278 (2008).
44. Stiller, M., Knapp, M., Stenzel, U., Hofreiter, M. & Meyer, M. Direct multiplex sequencing (DMPS) - a novel method for targeted high-throughput sequencing of ancient and highly degraded DNA. *Genome Res.* **19**, 1843–1848 (2009).
45. Meyer, M. *et al.* From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Res.* **36**, e5 (2008).
46. Kumar, S., Nei, M., Dudley, J. & Tamura, K. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform.* **9**, 299–306 (2008).
47. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
48. Bandelt, H. J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).
49. R: *A Language and Environment for Statistical Computing v. version 2.10.0* (R Foundation for Statistical Computing: Vienna, 2010).
50. Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818 (1998).
51. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, 699–710 (2006).
52. Drummond, A. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
53. Suchard, M. A., Weiss, R. E. & Sinsheimer, J. S. Bayesian selection of continuous-time Markov chain evolutionary models. *Mol. Biol. Evol.* **18**, 1001–1013 (2001).
54. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).
55. Kass, R. E. & Raftery, A. E. BAYES FACTORS. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
56. Shapiro, B. *et al.* A Bayesian phylogenetic method to estimate unknown sequence ages. *Mol. Biol. Evol.* **28**, 879–887 (2010).
57. Eisenmann, V. & Baylac, M. Extant and fossil Equus (Mammalia, Perissodactyla) skulls: a morphometric definition of the subgenus Equus. *Zoologica Scripta* **29**, 89–100 (2000).
58. Prado, J. L. & Alberdi, M. T. A cladistic analysis of the horses of the tribe Equini. *Palaeontology* **39**, 663–680 (1996).
59. Forstén, A. Mitochondrial-DNA time-table and the evolution of Equus: comparison of molecular and paleontological evidence. *Annales Zoologici Fennici* **28**, 301–309 (1992).
60. Orlando, L. *et al.* Revising the recent evolutionary history of equids using ancient DNA. *Proc. Natl. Acad. Sci. USA* **106**, 21754–21759 (2009).

Acknowledgements

We thank Matthias Meyer, Nadin Rohland, Cesare de Filippo, Monika Reißmann and Kay Prüfer for helpful discussions; the MPI EVA Sequencing Group for operating the 454 sequencer; Udo Stenzel for assisting with the 454 data analysis; Roger Mundry for assisting with the statistical analysis in R; and Christine Green for comments on the manuscript. The American Museum of Natural History (New York), Department of Tourism and Culture (Whitehorse), the Natural History Museum University of Kansas (Lawrence), the Canadian Museum of Civilization, the Römisch-Germanisches Zentralmuseum (Neuwied), the Landesamt für Denkmalpflege Baden-Württemberg, the Thüringisches Landesamt für Denkmalpflege und Archäologie, the Institute of Archaeology (Sankt Petersburg), and Klaas Post provided samples. This project was supported by the Max Planck Society (S.L. and M.H.), the Deutsche Forschungsgemeinschaft (LU 852/6-2 & AL 287/6-2, N.B. and A.L.), the Swedish Research Council (J.A.L.), The Natural Environment Research Council and Wellcome Trust (A.C.) and the Danish National Research Foundation (M.R., J.W., E.W.).

Author contributions

M.H. and S.L. conceived and designed the experiments. S.L. performed the experiments. M.R. did independent replications. S.L., M.K. and B.S. analysed the data. N.B., J.A.L., T.K., A.L., J.W., A.C. and E.W. provided samples. The first draft was written by S.L. and M.H. All authors contributed to the final version of the paper.

Additional information

Accession codes: All sequences have been deposited in nucleotide core GenBank database under the accession codes GQ495709 to GQ495789.

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Lippold, S. *et al.* Discovery of lost diversity of paternal horse lineages using ancient DNA. *Nat. Commun.* **2**:450 doi: 10.1038/ncomms1447 (2011).