Title: Ancient genomes revisit the ancestry of domestic and Przewalski’s horses

Authors: Charleen Gaunitz1,*, Antoine Fages1,2,*, Kristian Hanghøj1,2, Anders Albrechtsen3, Naveed Khan1,4, Mikkel Schubert1, Andaine Seguin-Orlando1,5, Ivy J. Owens6,7, Sabine Felkel8, Olivier Bignon-Lau9, Peter de Barros Damgaard1, Alissa Mittnik10, Azadeh Mohaseb11, Hossein Davoudi12, Saleh Alquraishi13, Ahmed H. Alfarhan13, Khaled A. S. Al-Rasheid13, Eric Crubézy2, Norbert Benecke14, Sandra Olsen15, Dorcas Brown16, David Anthony16, Ken Massy17, Vladimir Pitulko18, Aleksei Kasparov18, Gottfried Brem8, Michael Hofreiter19, Gulmira Mukhtarova20, Nurbol Baimukhanov21, Lembi Lõugas22, Vedat Onar23, Philipp W. Stockhammer10,17, Johannes Krausa10, Bazartseren Boldgiv24, Sainbileg Undrakhbold24, Diimaajav Erdenebaatar25, Sébastien Lepetz11, Marjan Mashkour11, Arne Ludwig26, Barbara Wallner8, Victor Merz27, Ilja Merz27, Viktor Zaibert28, Eske Willerslev1, Pablo Librado1, Alan K. Outram6,*, Ludovic Orlando1,2,+.

Affiliations:

1Centre for GeoGenetics, Natural History Museum of Denmark, Øster Voldgade 5-7, 1350K Copenhagen, Denmark

2Laboratoire d’Anthropobiologie Moléculaire et d’Imagerie de Synthèse UMR 5288, Université de Toulouse, CNRS, Université Paul Sabatier, France

3Bioinformatics Center, Department of Biology, University of Copenhagen, 2200N Copenhagen, Denmark

4Department of Biotechnology, Abdul Wali Khan University, Mardan, Pakistan

5National High-Throughput DNA Sequencing Center, Copenhagen, Denmark

6Department of Archaeology, University of Exeter, Exeter EX4 4QE, UK
The Charles McBurney Laboratory for Geoarchaeology, Division of Archaeology, University of Cambridge, Cambridge CB2 3DZ, UK

Institute of Animal Breeding and Genetics, University of Veterinary Medicine Vienna, 1210 Vienna, Austria

Équipe Ethnologie préhistorique, ArScAn, CNRS UMR 7041, MAE René-Ginouvès, 92023 Nanterre Cedex, France

Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Kahlaische Str. 10, 07745 Jena, Germany

Centre National de la Recherche Scientifique, Muséum National d’Histoire Naturelle, Archéozoologie, Archéobotanique, Sociétés, Pratiques et Environnements (UMR 7209), 55 rue Buffon, 75005 Paris, France

Department of Archaeology, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran

Zoology Department, College of Science, King Saud University, Riyadh 11451, Saudi Arabia

German Archaeological Institute, Department of Natural Sciences, Berlin, 14195 Berlin, Germany

Biodiversity Institute/Museum of Natural History, University of Kansas, Spooner Hall, 1340 Jayhawk Blvd., Lawrence, KS, 66045 USA

Anthropology Department, Hartwick College, Oneonta, NY 13820

Institute for Pre- and Protohistoric Archaeology and Archaeology of the Roman Provinces, Ludwig Maximilians-University Munich, Schellingstraße 12, 80799 München, Germany

Institute for the History of Material Culture, Russian Academy of Sciences, 18 Dvortsoyaya nab., St. Petersburg, 191186, Russia
19 University of Potsdam, Faculty of Mathematics and Natural Sciences, Institute for Biochemistry and Biology, Karl-Liebknecht-Strasse 24-25, 14476 Potsdam, Germany

20 State historical-cultural reserve-museum «Issyk», 040400 Almaty region, Enbekshikazak district, Kazakhstan

21 Shejire DNA project, Abai ave. 150/230, 050046 Almaty, Kazakhstan

22 Archaeological Research Collection, Tallinn University, Rüütli 10, 10130 Tallinn, Estonia

23 Osteoarchaeology Practice and Research Center & Department of Anatomy, Faculty of Veterinary Medicine, Istanbul University, 34320, Avcılar, Istanbul, Turkey

24 Ecology Group, Department of Biology, School of Arts and Sciences, National University of Mongolia, Ulaanbaatar 14201, Mongolia

25 Department of Archaeology, Ulaanbaatar State University, Ulaanbaatar 51, Mongolia

26 Department of Evolutionary Genetics, Leibniz Institute for Zoo and Wildlife Research, Berlin 10315, Germany

27 S.Toraighyrov Pavlodar State University, Joint Research Center for Archeological Studies named after A.Kh. Margulan, Lomov St. 64, cab. 102, Pavlodar, Kazakhstan

28 Scientific Centre of Archaeology and Ethnology, Kostanay State Pedagogical Institute, 110000, Kostanay, Tauelsizdik street, 118, Kazakhstan

*These authors contributed equally to this work. +Corresponding authors: Alan Outram, A.K.Outram@exeter.ac.uk; Ludovic Orlando, ludovic.orlando@univ-tlse3.fr
**Abstract** (116 words) The Eneolithic Botai culture of the Central Asian steppes provides the earliest archaeological evidence for horse husbandry, ~5,500 ya, but the exact nature of early horse domestication remains controversial. We generated 42 ancient horse genomes, including 20 from Botai. Compared to 46 published ancient and modern horse genomes, our data indicate that Przewalski’s horses are the feral descendants of horses herded at Botai and not truly wild horses. All domestic horses dated from ~4,000 ya to present only show ~2.7% of Botai-related ancestry. This indicates that a massive genomic turnover underpins the expansion of the horse stock that gave rise to modern domesticates, which coincides with large-scale human population expansions during the Early Bronze Age.

**One Sentence Summary** (103 characters) Earliest herded horses were the ancestors of feral Przewalski’s horses but not of modern domesticates.

**Main text:** Horses revolutionized human mobility, economy, and warfare (1). They are also associated with the spread of Indo-European languages (2), new forms of metallurgy (3) and provided the fastest land transport until modern times. Together with the lack of diachronic changes in horse morphology (4) and herd structure (5, 6), the scarce archaeological record hampered the study of early domestication. With their preponderance of horse remains, Eneolithic sites (5th and 4th Mill BCE) of the Pontic-Caspian steppe (2, 7) and the northern steppe of Kazakhstan (6, 8) have attracted the most attention.

We reconstructed the phylogenetic origins of the Eneolithic horses associated with the Botai culture of northern Kazakhstan, representing the earliest domestic horses (6, 8). This culture was characterized by a sudden shift from mixed hunting/gathering to an extreme focus on horses, and larger, more sedentary settlements (5). Horse dung on site (6), as well as evidence for poleaxing and against selective body part transportation, suggest controlled slaughter at settlements rather than hunting (9). Tools associated with leather thong production, bit-related dental pathologies (7, 10) and equine milk fats within ceramics support pastoral husbandry, involving milking and harnessing (8).
Geological surveys at the Botai culture site of Krasnyi Yar, Kazakhstan described a polygonal enclosure of ~20 x 15 meters with elevated phosphorus and sodium levels (6), likely corresponding to a horse corral. We revealed a similar enclosure at the eponymous Botai site, ~100 kilometers west of Krasnyi Yar (Fig. 1A), showing close-set post molds, merging to form a palisade trench, and a line of smaller parallel postholes inside (Fig. 1B). Radiocarbon dates on horse bones from these postholes are consistent with the Botai culture (11). The presence of enclosures at Krasnyi Yar and Botai adds on the evidence supporting horse husbandry.

We sequenced the genomes of 20 horses from Botai, and 22 from across Eurasia and spanning the last ~5,000 years (Table S1). With the published genomes of 18 ancient and 28 modern horses, this provided a comparative panel of three wild archaic horses (~42,800-5,100 ya), seven Przewalski’s horses (PH, six modern and one from the 19th century), and 78 domesticates (25 Eneolithic, including five from Borly4, Kazakhstan ~5,000 ya; seven Bronze Age ~4,100-3,000 ya; 18 Iron Age ~2,800-2,200 ya; one Parthian and two Roman ~2,000-1,600 ya; three post-Roman ~1,200-100 ya, and 22 modern from 18 breeds).

The 42 ancient genomes, belonging to 31 horse stallions and 11 mares, were sequenced to an average depth-of-coverage of ~1.1-9.3X (median=3.0X). Damage patterns indicative of ancient DNA were recovered (Figs. S8 and S9). Base quality rescaling and termini trimming resulted in average error rates of 0.07%-0.14% per site (Tables S13 and S14).

Principal Component Analysis (PCA) revealed PH and the archaic horses as two independent clusters (Fig. 2A). Within domesticates, all 25 Botai/Borly4 Eneolithic specimens grouped together to the exclusion of all remaining horses.

Phylogenetic reconstruction confirmed that domestic horses do not form a single monophyletic group as expected if descending from Botai (Fig. 2B). Instead, PH form a highly-drifted, monophyletic group, unambiguously nested within Botai/Borly4 horses. All remaining domesticates cluster within a second,
highly-supported monophyletic group (DOM2). Applying TreeMix (12) to the 60 genomes with minimal 3.0X average depth-of-coverage confirmed this tree topology (Fig. S23).

$f_3$-outgroup and D-statistics (13) support PH as genetically closer to Botai/Borly4 individuals than any DOM2 member (Fig. 2C, Fig. S25 and S26). Finally, ancestry tests (14) confirmed Botai horses as the direct ancestors of Borly4 horses, and the latter as ancestral to the only PH in our dataset pre-dating their massive demographic collapse and introgression of modern domestic genes (15).

$f_3$-outgroup and D-statistics also revealed that Dunaújváros_Duk2 (Duk2), the earliest and most basal specimen within DOM2, was divergent to all other DOM2 members. This is not due to sequencing errors since the internal branch splitting from Duk2 and leading to the ancestor of all remaining DOM2 horses is long (Fig. 2B). This suggests instead shared ancestry between Duk2 and a divergent ghost population. We thus excluded Duk2 in admixture graph reconstructions (16) to avoid bias due to contributions from unsampled lineages (Fig. 3).

In the absence of admixture, the best admixture graph matched the trees reconstructed above. We also reconstructed admixture graphs for five additional scenarios with one or two admixture event(s), including between PH and domesticates (15). Bayes Factors best supported a horse domestication history where a first lineage gave rise to Botai/Borly4 and PH horses, while a second lineage founded DOM2 and provided the source of domestic horses during at least the last ~4,000 years, with minimal contribution from the Botai/Borly4 lineage (95% CI=2.0-3.8%).

The limited Botai/Borly4 ancestry amongst DOM2 members concurs with slightly significant negative D-statistics in the form of (((DOM2_ancient,DOM2_modern),Botai/Borly4),donkey) for some DOM2 members, spanning a large geographical (Western Europe, Turkey, Iran and Central Asia) and temporal range (from ~3,318 ya to ~1,143 ya; Fig. S28). This suggests sporadic introgression of Botai ancestry into multiple DOM2 herds until the last thousand years. This gene flow was mediated through females since 15 Botai/Borly4 individuals carried mitochondrial haplotypes characteristic of DOM2 matrilines (Fig.
S12-S13) but also through males given the persistence of Botai/Borly4-related patrilines within DOM2 (Fig. S15-18).

PH are considered as the last true remaining wild horses, that have never been domesticated (15). Our results reveal that they represent instead the feral descendants of horses first herded at Botai. Instead, it appears that their feralization likely involved multiple biological changes.

Metacarpal measurements in 263 ancient and 112 modern horses indicate that PH have become less robust than their Botai/Borly4 ancestors (Fig. 4A). One Botai individual likely showed limited unpigmented areas and leopard spots as it was heterozygous for four mutations at the TRPM1 locus associated with leopard spotting and carried the ancestral allele at the PATN1 modifier (17, 18) (Fig. 4B). Individuals homozygous for TRPM1 mutations are generally almost completely unpigmented and develop congenital stationary night blindness (17). First maintained at Botai by human management, the haplotype associated with leopard spotting was likely selected against and lost once returning wild, leading to the characteristic PH Dun dilution coloration (19). Genomic regions with signatures of positive selection along the phylogenetic branch separating Borly4 and PH showed functional enrichment for genes associated in humans with cardiomyopathies (p-values<0.0496), melanosis and hyperpigmentation (0.0468), and skeletal abnormalities (0.0594) (Table S18), suggesting that at least some of the morpho-anatomical changes associated with feralization were adaptive.

Additionally, significantly negative D-statistics in the form of (((DOM2,PH),archaic),donkey) previously suggested that the extinct, archaic lineage formed by ~5.1-42.7 ky-old horses from Taymyr and Yakutia, contributed to the genetic ancestry of modern domesticates (20, 21). Although we could confirm such D-statistics (Fig. S29), almost all other D-statistics in the form of (((DOM2,Botai/Borly4),archaic),donkey) were not different from zero (Fig. S30). This indicates selection against the archaic ancestry between ~4,977 and ~118 ya (the time interval separating the youngest Borly4 individual and the earliest PH sequenced). Alternatively, the PH lineage admixed with a divergent population of horses, both unrelated
to the archaic lineage and the ghost population that contributed ancestry to Duk2, since D-statistics revealed Duk2 as closer to Borly4 than to PH (Fig. S31).

Lastly, although the genetic load of PH and Botai/Borly4 genomes was equivalent until ~118 ya, it drastically increased in modern animals (Fig. 4C). This accumulation of deleterious variants was thus not associated with PH feralization but with the recent introgression of deleterious variants from modern domesticates and demographic collapse, which hampered purifying selection.

That none of the domesticates sampled in the last ~4,000 years descend from the horses first herded at Botai entails another major implication. It suggests that during the 3rd Mill BCE at the latest, another unrelated group of horses became the source of all domestic populations that expanded thereafter. This is compatible with two scenarios. First, Botai-type horses experienced massive introgression capture (22) from a population of wild horses until the Botai ancestry was almost completely replaced. Alternatively, horses were successfully domesticated in a second domestication center and incorporated minute amounts of Botai ancestry during their expansion. We cannot identify the locus of this hypothetic center due to a temporal gap in our dataset throughout the 3rd Mill BCE. However, that the DOM2 earliest member was excavated in Hungary adds Eastern Europe to other candidates already suggested, including the Pontic-Caspian steppe (2), Eastern Anatolia (23), Iberia (24), Western Iran and the Levant (25). Notwithstanding the process underlying the genomic turnover observed, the clustering of ~4,023-3,574 year-old specimens from Russia, Romania and Georgia within DOM2 suggests that this clade already expanded throughout the steppes and Europe at the transition between the 3rd and 2nd Mill BCE, in line with the demographic expansion at ~4,500 ya recovered in mitochondrial Bayesian Skylines (Fig. S14).

This study shows that the horses exploited by the Botai people later became the feral PH. Early domestication most likely followed the ‘prey pathway’ whereby a hunting relationship was intensified until reaching concern for future progeny through husbandry, exploitation of milk and harnessing (7). Other horses, however, were the main source of domestic stock over the last ~4,000 years or more.
Ancient human genomics (26) has revealed considerable human migrations ~5,000 ya involving “Yamnaya” culture pastoralists of the Pontic-Caspian steppe. This expansion might be associated with the genomic turnover identified in horses, especially if Botai horses were best suited to localized pastoral activity than to long distance travel and warfare. Future work must focus on identifying the main source of the domestic horse stock and investigating how the multiple human cultures managed the available genetic variation to forge the many horse types known in history.

References and notes


11. See Supplementary Material on Science online.


17. R. R. Bellone et al., Evidence for a retroviral insertion in TRPM1 as the cause of congenital stationary night blindness and leopard complex spotting in the horse. PLoS One. 8, e78280 (2013).


77. V. Eisenmann, Metapodials, System of measurements (2009); http://www.vera-eisenmann.com/metapodials-system-of-measurements


96. Picard tools (2017); https://broadinstitute.github.io/picard/


133. S. Rieder, S. Taourit, D. Mariat, B. Langlois, G. Guérin, Mutations in the agouti (ASIP), the extension (MC1R), and the brown (TYRP1) loci and their association to coat color phenotypes in horses (Equus caballus). *Mamm. Genome.* **12**, 450–455 (2001).


171. Z. Liu et al., Genome-wide analysis reveals signatures of selection for important traits in domestic sheep from different ecoregions. *BMC Genomics*. 17, 863 (2016).


**Acknowledgments**

We thank the British Institute of Persian Studies (BIPS) in Tehran and the National Museum of Iran for providing access to the material from Iran; the Archaeometry Laboratory of the University of Tehran; the staff of the Danish National High-Throughput DNA Sequencing Center; Cristina Gamba and Christian McCrory Constantz for technical support and/or discussions; Laurent Frantz, Dan Bradley and Greger Larson for critical reading of the manuscript. **Funding:** This work was supported by the Danish Council for Independent Research, Natural Sciences (4002-00152B); the Danish National Research Foundation (DNRF94); Initiative d'Excellence Chaires d'attractivité, Université de Toulouse (OURASI); the
Author Contributions: LO conceived the project and designed research; IJO, VZ and AKO designed and carried out field archaeological work; CG, AF, and NK performed ancient DNA laboratory work, with input from LO; PL and LO designed and coordinated computational analyses; KH, PL, MS, AA and LO performed computational analyses; NBe, KM, PWS, VP, AK, GM, NBa, LL, VO, JK, BB, SU, DE, SL, MM, HD, AM, AL, VM, VZ and LO provided/collection samples; OB, SA, AHA, KASAR, EW and LO provided reagents, measurements and material; CG, AF, KH and LO prepared figures and tables, with input from LO; CG, AF, KH, PL, IJO, AKO and LO wrote the supplementary information; AKO and LO wrote the paper, with input from all other co-authors. Competing interests: The authors declare that they have no competing interests. Data and material availability: Individual genome sequence data are available at the European Nucleotide Archive (accession no. PRJEB22390).

Fig. 1. Sample location and corral enclosure at Botai. (A) Archaeological sites. The number of genomes sequenced per site is reported between parentheses if greater than one. Triangles refer to the ancient genomes characterized here while diamonds indicate those previously published. Blue refers to wild ancient individuals, light and dark green refer to the first domestic clade (Botai and Borly4) and yellow to individuals of the second domestic clade (DOM2). The Botai culture site of Krasnyi Yar is indicated although no samples were analyzed from this site. (B). Magnetic gradient survey and excavation at Botai, with interpretation. The enclosure and its excavated boundary are indicated by red and yellow squares, respectively. Round black circles correspond to pit houses.
Fig. 2. Horse genetic affinities. (A) Principal Component Analysis of the genome variation present in 88 ancient and modern genomes. Only the first two principal components are shown. (B) Phylogenetic relationships. The tree was reconstructed on the basis of pairwise distances calculated with ~14.1 million transversion sites. Node support derive from 100 bootstrap pseudo-replicates. The archaeological site and age (ya) of ancient specimens are indicated in the first and last fields of the sample name. (C) $f_3$-outgroup statistics showing the pairwise genetic affinities.

Fig. 3. Admixture graphs. Panels (A-F) present the six scenarios tested. Panel A received decisive Bayes Factor support, as indicated below each corresponding alternative scenario tested. Domestic-Ancient and Domestic-A/B refer to three phylogenetic clusters identified within DOM2 (excluding Duk2): ancient individuals; modern Mongolian, Yakutian (including Tumeski_CGG101397) and Jeju horses, and; all remaining modern breeds. (G) Posterior distributions of admixture proportions along $p_1$ and $p_2$ branches.

Fig. 4. Phenotypic and genomic changes associated with ferality. (A) Indices of the robustness of the 3rd metacarpal bone in various horse populations. Bd = Breadth at the middle of the diaphysis. GL = Maximal/Greatest length. Kent and Kumkeshu/Kozhai represent populations of Kazakhstan from the Iron Age and Eneolithic (Tersek culture), respectively. (B) Genotyping information at the TRPM1 locus (chr1) and the PATN1 modifier (chr3) for Botai/Borly4 horses. The absence, heterozygosis and homozygosis of alleles strongly associated with leopard spotting are depicted in white, dark grey and red, respectively. Crosses indicate insufficient data. The causative LTR insertion at the TRPM1 locus is indicated by the number of reads overlapping both flanks of the insertion site. (C) Individual-based genetic loads. The mauve circle shows the PH specimen from the 19th century.
Supplementary Materials

www.sciencemag.org

Materials and Methods

Figures S1 to S34

Tables S1 to S7 (Tables S8 to S18 as separate Excel files)

References (27-175)
Admixture fraction (%): $p_1 = 0.027$, $p_2 = 0.068$

Blocks:
- Donkey
- Botai
- Borly4
- Archaic
- Przewalski-19th
- Przewalski-21st
- Domestic-A
- Domestic-B
- Domestic-Ancient

Posterior Probability:
- 0.20
- 0.10
- 0.00

Branches:
- 5.56
- 3.58
- 0.95
- 1.93
- 7.55
- 4.03
- 2.22
- 1.94
- 0.13

Numbers:
- 267.42
- 30.67
- 88.46
- 30.62
- 28.59