

# The origins and spread of domestic horses from the Western Eurasian steppes

Librado P., Khan N., Fages A., Kusliy M.A., Suchan T., Tonasso-Calvière L., Schiavinato S., Alioglu D., Fromentier A., Perdereau A., Aury J.M., Gaunitz C., Chauvey L., Seguin-Orlando A., Der Sarkissian C., Southon J., Shapiro B., Tishkin A.A., Kovalev A.A., Alquraishi S., Alfarhan A.H., Al-Rasheid K.A.S., Seregély T., Klassen L., Iversen R., Bignon-Lau O., Bodu P., Olive M., Castel J.C., Boudadi-Maligne M., Alvarez N., Germonpré M., Moskal-del Hoyo M., Wilczyński J., Pospuła S., Lasota-Kuś A., Tunia K., Nowak M., Rannamäe E., Saarma U., Boeskorov G., Lõugas L., Kyselý R., Peške L., Bălăşescu A., Dumitraşcu V., Dobrescu R., Gerber D., Kiss V., Szécsényi-Nagy A., Mende B.G., Gallina Z., Somogyi K., Kulcsár G., Gál E., Bendrey R., Allentoft M.E., Sirbu G., Dergachev V., Shephard H., Tomadini N., Grouard S., Kasparov A., Basilyan A.E., Anisimov M.A., Nikolskiy P.A., Pavlova E.Y., Pitulko V., Brem G., Wallner B., Schwall C., Keller M., Kitagawa K., Bessudnov A.N., Bessudnov A., Taylor W., Magail J., Gantulga J.O., Bayarsaikhan J., Erdenebaatar D., Tabaldiev K., Mijiddorj E., Boldgiv B., Tsagaan T., Pruvost M., Olsen S., Makarewicz C.A., Valenzuela Lamas S., Albizuri Canadell S., Nieto Espinet A., Iborra M.P., Lira Garrido J., Rodríguez González E., Celestino S., Olària C., Arsuaga J.L., Kotova N., Pryor A., Crabtree P., Zhumatayev R. *Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia*

---

## Abstract

Domestication of horses fundamentally transformed long-range mobility and warfare<sup>1</sup>. However, modern domesticated breeds do not descend from the earliest domestic horse lineage associated with archaeological evidence of bridling, milking and corralling<sup>2–4</sup> at Botai, Central Asia around 3500 bc<sup>3</sup>. Other longstanding candidate regions for horse domestication, such as Iberia<sup>5</sup> and Anatolia<sup>6</sup>, have also recently been challenged. Thus, the genetic, geographic and temporal origins of modern domestic horses have remained unknown. Here we pinpoint the Western Eurasian steppes, especially the lower Volga-Don region, as the homeland of modern domestic horses. Furthermore, we map the population changes accompanying domestication from 273 ancient horse genomes. This reveals that modern domestic horses ultimately replaced almost all other local populations as they expanded rapidly across Eurasia from about 2000 bc, synchronously with equestrian material culture, including Sintashta spoke-wheeled chariots. We find that equestrianism involved strong selection for critical locomotor and behavioural adaptations at the GSDMC and ZFPM1 genes. Our results reject the commonly held association<sup>7</sup> between horseback riding and the massive expansion of Yamnaya steppe pastoralists into Europe around 3000 bc<sup>8,9</sup> driving the spread of Indo-European languages<sup>10</sup>. This contrasts with the scenario in Asia where Indo-Iranian languages, chariots and horses spread together, following the early second millennium bc Sintashta culture<sup>11,12</sup>.

<http://dx.doi.org/10.1038/s41586-021-04018-9>

---

## References

- [1] Kelekna, P. *The Horse in Human History* (Cambridge Univ. Press, 2009).
- [2] Outram, A. K. et al. The earliest horse harnessing and milking. *Science* 323, 1332–1335 (2009) DOI: 10.1126/science.1168594
- [3] Gaunitz, C. et al. Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science* 360, 111–114 (2018). DOI: 10.1126/science.aao3297
- [4] Olsen, S. L. in *Horses and Humans: The Evolution of Human Equine Relationships* (eds Olsen S. L. et al.) 81–113 (Archaeopress, 2006).
- [5] Fages, A. et al. Tracking five millennia of horse management with extensive ancient genome time series. *Cell* 177, 1419–1435.e31 (2019). DOI: 10.1016/j.cell.2019.03.049
- [6] Guimaraes, S. et al. Ancient DNA shows domestic horses were introduced in the southern Caucasus and Anatolia during the Bronze Age. *Sci. Adv.* 6, eabb0030 (2020).
- [7] Anthony, D. W. *The Horse, the Wheel and Language* (Princeton Univ. Press, 2007).
- [8] Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522, 207–211 (2015). DOI: 10.1038/nature14317
- [9] Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. *Nature* 522, 167–172 (2015). DOI: 10.1038/nature14507
- [10] Demoule, J. P. *Mais où sont passés les Indo-Européens ? Le mythe d'origine de l'Occident* (Le Seuil, 2014).
- [11] de Barros Damgaard, P. et al. 137 ancient human genomes from across the Eurasian steppes. *Nature* 557, 369–374 (2018). DOI: 10.1038/s41586-018-0094-2
- [12] Narasimhan, V. M. et al. The formation of human populations in South and Central Asia. *Science* 365, eaat7487 (2019). DOI: 10.1126/science.aat7487
- [13] Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B* 370, 20130624 (2015). DOI: 10.1098/rstb.2013.0624
- [14] Schubert, M. et al. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl Acad. Sci. USA* 111, E5661–E5669 (2014). DOI: 10.1073/pnas.1416991111
- [15] Librado, P. et al. Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proc. Natl Acad. Sci. USA* 112, E6889–E6897 (2015). DOI: 10.1073/pnas.1513696112
- [16] Petkova, D., Novembre, J. & Stephens, M. Visualizing spatial population structure with estimated effective migration surfaces. *Nat. Genet.* 48, 94–100 (2016). DOI: 10.1038/ng.3464
- [17] Patterson, N. et al. Ancient admixture in human history. *Genetics* 192, 1065–1093 (2012). DOI: 10.1534/genetics.112.145037
- [18] Harney, É., Patterson, N., Reich, D. & Wakeley, J. Assessing the performance of qpAdm: a statistical tool for studying population admixture. *Genetics* 217, iyaa045 (2021). DOI: 10.1093/genetics/iyaa045
- [19] Molloy, E. K., Durvasula, A. & Sankararaman, S. Advancing admixture graph estimation via maximum likelihood network orientation. *Bioinformatics* 37, i142–i150 (2021).
- [20] Battey, C., Ralph, P. L. & Kern, A. D. Predicting geographic location from genetic variation with deep neural networks. *eLife* 9, e54507 (2020). DOI: 10.7554/eLife.54507
- [21] de Barros Damgaard, P. et al. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* 360, eaar7711 (2018). DOI: 10.1126/science.aar7711
- [22] Reinhold, S. et al. in *Appropriating Innovations: Entangled Knowledge in Eurasia, 5000–1500 bce* (eds Stockhammer, P. W. & Maran, J.) 78–97 (Oxbow Books, 2017).
- [23] Kristiansen, K. in *Trade and Civilization. Economic Networks and Cultural Ties, from Prehistory to the Early Modern Period* (eds Kristiansen, K. et al.) (Cambridge Univ. Press, 2018).
- [24] Chechushkov I. V., & Epimakhov, A. V. in *The Puzzle of Indo-European Origins and Dispersals: Archeology, Linguistics and Genetics* (eds Kristiansen, K. et al.) (Cambridge Univ. Press, in the press).
- [25] Littauer, M. A., & Crouwel, J. H. The origin of the true chariot. *Antiquity* 70, 934–939 (1996). DOI: 10.1017/S0003598X00084192
- [26] Lindner, S. Chariots in the Eurasian Steppe: a Bayesian approach to the emergence of horse-drawn transport in the early second millennium BC. *Antiquity* 94, 361–380 (2020). DOI: 10.15184/aqy.2020.37
- [27] Moorey, P. R. S. Pictorial evidence for the history of horse-riding in Iraq before the Kassite period. *Iraq* 32, 36–50 (1970). DOI: 10.2307/4199890
- [28] Kanne, K. Riding, ruling, and resistance equestrianism and political authority in the Hungarian Bronze Age. *Curr. Anthropol.* (in the press).
- [29] Suri, P. et al. Genome-wide meta-analysis of 158,000 individuals of European ancestry identifies three loci associated with chronic back pain. *PLoS Genet.* 14, e1007601 (2018). DOI: 10.1371/journal.pgen.1007601

- [30] Jiang, H. et al. Two GWAS-identified variants are associated with lumbar spinal stenosis and Gasdermin-C expression in Chinese population. *Sci. Rep.* 10, 21069 (2020). DOI: 10.1038/s41598-020-78249-7
- [31] Tikker, L. et al. Inactivation of the GATA cofactor ZFPM1 results in abnormal development of dorsal raphe serotonergic neuron subtypes and increased anxiety-like behavior. *J. Neurosci.* 40, 8669–8682 (2020). DOI: 10.1523/JNEUROSCI.2252-19.2020
- [32] Takahashi, A. & Miczek, K. A. Neurogenetics of aggressive behavior: studies in rodents. *Curr. Top. Behav. Neurosci.* 17, 3–44 (2014). DOI: 10.1007/7854\_2013\_263
- [33] Schmitt, T. & Varga, Z. Extra-Mediterranean refugia: the rule and not the exception? *Frontiers Zool.* 9, 22 (2012). DOI: 10.1186/1742-9994-9-22
- [34] Spasskaya, N. N., & Pavlinov, I. in *Zoological Research* (Arch. Zoological Museum, Moscow State Univ., 2016).
- [35] Colledge, S., Conolly, J., Crema, E., & Shennan, S. Neolithic population crash in northwest Europe associated with agricultural crisis. *Quat. Res.* 92, 686–707 (2019). DOI: 10.1017/qua.2019.42
- [36] Outram, A. K. & Bogaard, A. *Subsistence and Society in Prehistory: New Directions in Economic Archaeology* (Cambridge Univ. Press, 2019).
- [37] Anthony, D. W. in *Social Complexity in Prehistoric Eurasia: Monuments, Metals and Mobility* (eds Hanks, B. K. & Lindruff, K. M.) Ch. 4 (2009).
- [38] Maran, J., Bajenaru, R., Ailincăi, S.-C., Popescu, A.-D. & Hansen, S. I. Objects, ideas and travelers. Contacts between the Balkans, the Aegean and Western Anatolia during the Bronze and Early Iron Age. In: *Proc. of the Conference in Tulcea 10-13 November, 2017* (Rudolf Habelt, 2020).
- [39] Glob, P. V. *Denmark: An Archaeological History from the Stone Age to the Vikings* (Cornell Univ. Press, 1971).
- [40] Gimbutas, M. The first wave of Eurasian Steppe pastoralists into Copper Age Europe. *J. Indo. Eur. Stud.* 5, 277–338 (1977).
- [41] Anthony, D. W. The “Kurgan Culture,” Indo-European origins, and the domestication of the horse: a reconsideration. *Curr. Anthropol.* 27, 291–313 (1986). DOI: 10.1086/203441
- [42] Renfrew, C. They ride horses, don't they?: Mallory on the Indo-Europeans. *Antiquity* 63, 843–847 (1989). DOI: 10.1017/S0003598X00077012
- [43] Vandkilde, H. *Culture and Change in Central European Prehistory* (Aarhus Univ. Press, 2007).
- [44] Häusler, A. in *Indogermanen und das Pferd* (eds Hänsel, B. & Zimmer, S.) 217–257 (*Archaeolingua* Alapitvány, 1994).
- [45] Kroonen, G., Barjamovic, G. & Peyrot, M. Linguistic supplement to de Barros Damgaard et al. 2018: Early Indo-European languages, Anatolian, Tocharian and Indo-Iranian <https://zenodo.org/record/1240524#.YFtLgGjTVMQ> (2018).
- [46] South, A. rworldmap: a new R package for mapping global data. *R J.* 3, 35–43 (2011). DOI: 10.32614/RJ-201-006
- [47] Brownrigg, R. maps: draw geographical maps. R package version 3.3.0 <https://CRAN.R-project.org/package=maps> (2018).
- [48] Reimer, P. et al. The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* 62, 725 (2020). DOI: 10.1017/RDC.2020.41
- [49] Ramsey, C. B. Bayesian analysis of radiocarbon dates. *Radiocarbon* 51, 337–360 (2009). DOI: 10.1017/S0033822200033865
- [50] Seguin-Orlando, A. et al. Heterogeneous hunter-gatherer and steppe-related ancestries in Late Neolithic and Bell Beaker genomes from present-day France. *Curr. Biol.* 31, 1072–1083.e10 (2021). DOI: 10.1016/j.cub.2020.12.015
- [51] Gamba, C. et al. Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Mol. Ecol. Resour.* 16, 459–469 (2016). DOI: 10.1111/1755-0998.12470
- [52] Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* 9, 88 (2016). DOI: 10.1186/s13104-016-1900-2
- [53] Kalbfleisch, T. S. et al. Improved reference genome for the domestic horse increases assembly contiguity and composition. *Commun. Biol.* 1, 197 (2018). DOI: 10.1038/s42003-018-0199-z
- [54] Xu, X. & Arnason, U. The complete mitochondrial DNA sequence of the horse, *Equus caballus*: extensive heteroplasmy of the control region. *Gene* 148, 357–362 (1994). DOI: 10.1016/0378-1119(94)90713-7
- [55] Felkel, S. et al. The horse Y chromosome as an informative marker for tracing sire lines. *Sci. Rep.* 9, 6095 (2019) DOI: 10.1038/s41598-019-42640-w
- [56] Poulet, M. & Orlando, L. Assessing DNA sequence alignment methods for characterizing ancient genomes and methylomes. *Front. Ecol. Evol.* 8, 105 (2020). DOI: 10.3389/fevo.2020.00105
- [57] Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29, 1682–1684 (2013). DOI: 10.1093/bioinformatics/btt193

- [58] Orlando, L. et al. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499, 74–78 (2013). DOI: 10.1038/nature12323
- [59] Jónsson, H. et al. Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proc. Natl Acad. Sci. USA* 111, 18655–18660 (2014). DOI: 10.1073/pnas.1412627111
- [60] Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* 15, 356 (2014). DOI: 10.1186/s12859-014-0356-4
- [61] Skoglund, P. et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl Acad. Sci. USA* 111, 2229–2234 (2014). DOI: 10.1073/pnas.1318934111
- [62] Librado, P. et al. Ancient genomic changes associated with domestication of the horse. *Science* 356, 442–445 (2017). DOI: 10.1126/science.aam5298
- [63] Der Sarkissian, C. et al. Evolutionary genomics and conservation of the endangered Przewalski's horse. *Curr. Biol.* 25, 2577–2583 (2015). DOI: 10.1016/j.cub.2015.08.032
- [64] Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780 (2013). DOI: 10.1093/molbev/mst010
- [65] Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313 (2014). DOI: 10.1093/bioinformatics/btu033
- [66] Bouckaert, R. et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 15, e1006650 (2019). DOI: 10.1371/journal.pcbi.1006650
- [67] Heller, R., Chikhi, L. & Siegmund, H. R. The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. *PLoS ONE* 8, e62992 (2013). DOI: 10.1371/journal.pone.0062992
- [68] Keane, T. M., Creevey, C. J., Pentony, M. M., Naughton, T. J. & McInerney, J. O. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol. Biol.* 6, 29 (2006). DOI: 10.1186/1471-2148-6-29
- [69] 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). DOI: 10.1093/molbev/msi103
- [70] Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol* 67, 901–904 (2018). DOI: 10.1093/sysbio/syy032
- [71] Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274 (2015). DOI: 10.1093/molbev/msu300
- [72] Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522 (2018). DOI: 10.1093/molbev/msx281
- [73] Lefort, V., Desper, R. & Gascuel, O. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* 32, 2798–2800 (2015). DOI: 10.1093/molbev/msv150
- [74] Paradis, E., & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528 (2019). DOI: 10.1093/bioinformatics/bty633
- [75] Schraiber, J. Assessing the relationship of ancient and modern populations. *Genetics* 208, 383–398 (2018). DOI: 10.1534/genetics.117.300448
- [76] Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575 (2007). DOI: 10.1086/519795
- [77] Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664 (2009). DOI: 10.1101/gr.094052.109
- [78] Cheng, J. Y., Mailund, T. & Nielsen, R. Fast admixture analysis and population tree estimation for SNP and NGS data. *Bioinformatics* 33, 2148–2155 (2017). DOI: 10.1093/bioinformatics/btx098
- [79] Lawson, D. J., van Dorp, L. & Falush, D. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nat. Commun.* 9, 3258 (2018). DOI: 10.1038/s41467-018-05257-7
- [80] Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. Robust demographic inference from genomic and SNP data. *PLoS Genet.* 9, e1003905 (2013). DOI: 10.1371/journal.pgen.1003905
- [81] Gerritsen, H. mapplots: data visualisation on maps. R package version 1.5.1 <https://CRAN.R-project.org/package=mapplots> (2018).
- [82] Bjornstad, O. N. & Cai, J. ncf: spatial covariance functions. R package version 1.2-9 <http://ento.psu.edu/directory/onb1> (2020).
- [83] Loog, L. et al. Estimating mobility using sparse data: application to human genetic variation. *Proc. Natl Acad. Sci. USA* 114, 12213–12218 (2017). DOI: 10.1073/pnas.1703642114
- [84] Hijmans, R. J., Williams, E. & Vennes, C. E. geosphere: spherical trigonometry. R package version 1.5.1 (2019).
- [85] Boyle, J. GeoRange: calculating geographic range from occurrence data. R package version 0.1.0. (2017).
- [86] Hahne, F. & Ivanek, R. Visualizing genomic data using Gviz and Bioconductor. *Methods Mol. Biol.* 1418, 335–351 (2016). DOI: 10.1007/978-1-4939-3578-9\_16

- [87] Renaud, G. et al. Improved de novo genomic assembly for the domestic donkey. *Sci. Adv.* 4, eaaq0392 (2018). DOI: 10.1126/sciadv.aaq0392
- [88] Jagannathan, V. et al. Comprehensive characterization of horse genome variation by whole-genome sequencing of 88 horses. *Anim. Genet.* 50, 74–77 (2019). DOI: 10.1111/age.12753
- [89] Andersson, L. S. et al. Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. *Nature* 488, 642–646 (2012). DOI: 10.1038/nature11399
- [90] Teufer, M. Ein Scheibenknebel aus Dzarkutan (Süduzbekistan). *Archäologische Mitteilungen aus Iran und Turan*. Band 31, 69–142 (1999).
- [91] Chechushkov, I. V. Wheel Complex of the Late Bronze Age Era of Steppe and Forest-Steppe Eurasia (from Dnieper to Irtysh). PhD thesis. Department of Archeology and Ethnography of the Federal State Budgetary Institution of Science, Institute of History and Archeology of the Ural Branch of the Russian Academy of Sciences (2013).