DNA methylation-based profiling of horse archaeological remains for age-at-death and castration

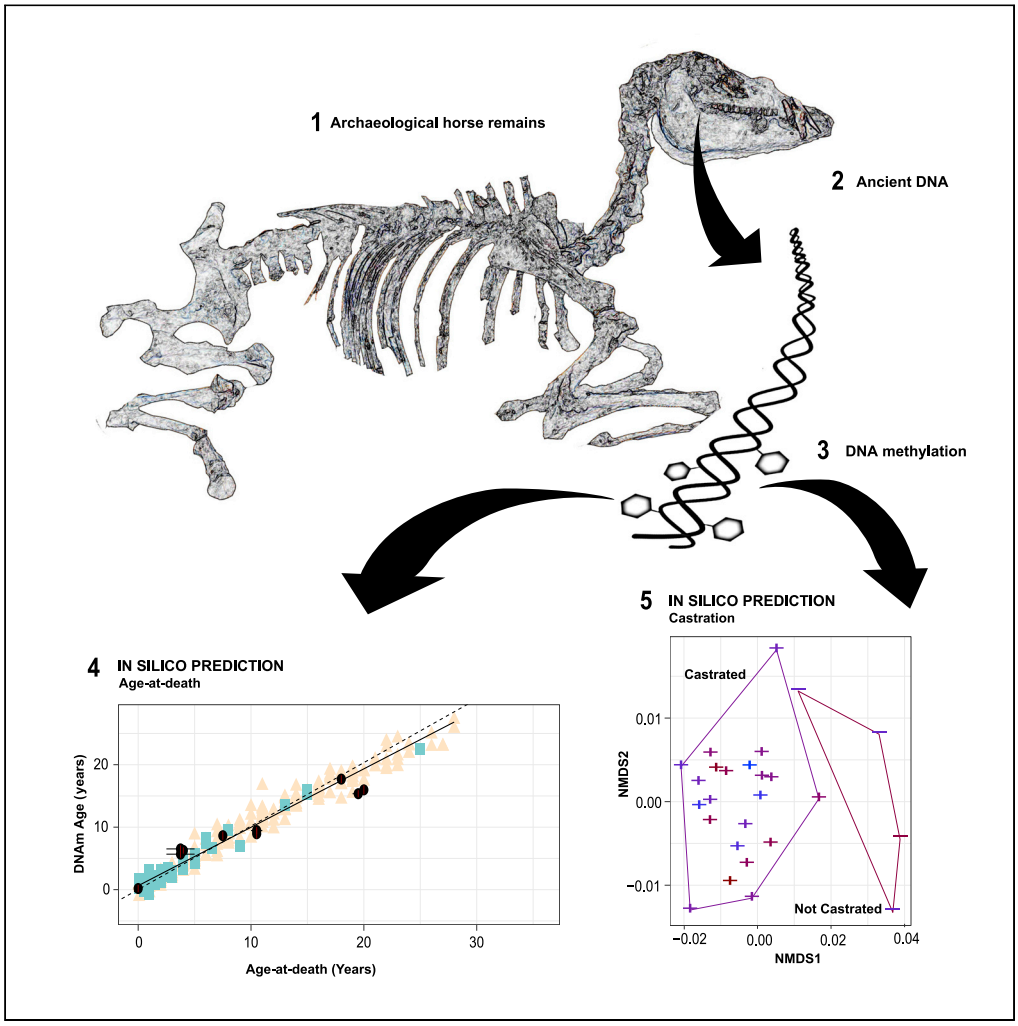
Xuexue Liu, Andaine Seguin- Orlando, Lorelei Chauvey, Gaëtan Tressières, Laure Tonasso-Calvie, Jean-Marc Aury, Aude Perdereau, Stefanie Wagner, Pierre Clavel, Oscar Estrada, et al.

iScience

CellPress
OPEN ACCESS

Article

DNA methylation-based profiling of horse archaeological remains for age-at-death and castration



Xuexue Liu, Andaine Seguin-Orlando, Lorelei Chauvey, ..., Yvette Running Horse Collin, Clio Der Sarkissian, Ludovic Orlando

ludovic.orlando@univ-tlse3.fr

Highlights

We develop reliable DNA methylation clocks for the age-at-death of ancient horses

We use a cost-effective capture assay to infer ancient DNA methylation at 1,611 CpGs

The ancient ritual and domestic waste deposits examined showed similar age profiles

Support Vector Machines achieve full accuracy for castration prediction

Liu et al., iScience 26, 106144
March 17, 2023
https://doi.org/10.1016/j.isci.2023.106144

iScience

CellPress
OPEN ACCESS

Article

DNA methylation-based profiling of horse archaeological remains for age-at-death and castration

Xuexue Liu,[1] Andaine Seguin-Orlando,[1] Lorelei Chauvey,[1] Gaëtan Tressières,[1] Stéphanie Schiavinato,[1] Laure Tonasso-Calvière,[1] Jean-Marc Aury,[2] Aude Perdereau,[2] Stefanie Wagner,[1,3] Pierre Clavel,[1] Oscar Estrada,[1] Jianfei Pan,[4] Yuehui Ma,[4] Jacob Enk,[5] Alison Devault,[5] Jennifer Klunk,[5] Sébastien Lepetz,[6] Benoit Clavel,[6] Lin Jiang,[4] Patrick Wincker,[2] Yvette Running Horse Collin,[1] Clio Der Sarkissian,[1] and Ludovic Orlando[1,7,*]

SUMMARY

Age profiling of archaeological bone assemblages can inform on past animal management practices, but is limited by the fragmentary nature of the fossil record and the lack of universal skeletal markers for age. DNA methylation clocks offer new, albeit challenging, alternatives for estimating the age-at-death of ancient individuals. Here, we take advantage of the availability of a DNA methylation clock based on 31,836 CpG sites and dental age markers in horses to assess age predictions in 84 ancient remains. We evaluate our approach using whole-genome sequencing data and develop a capture assay providing reliable estimates for only a fraction of the cost. We also leverage DNA methylation patterns to assess castration practice in the past. Our work opens for a deeper characterization of past husbandry and ritual practices and holds the potential to reveal age mortality profiles in ancient societies, once extended to human remains.

INTRODUCTION

Zooarchaeological research is aimed at understanding the diversity of human-animal relationships in the past.[1] It has traditionally built on the analysis of morphological variation, which allows for the taxonomic and sex identification of archaeological bone assemblages,[2] as well as their geospatial differentiation.[3] The presence of bone cut marks and tooth bit wear[4] also typically provide insights into butchery and bridling practice,[5] while isotopic signatures can reveal patterns of animal mobility throughout their lives,[6] including seasonal mobility and diet.[7]

Recent advances in ancient DNA research have democratized access to genome data from archaeological osseous material.[8] While most attention is focused on ancient human remains, providing fine-grained resolution into the atlas of past human migrations,[9] the genomic analysis of non-human archaeological remains has also improved our understanding of past human-animal relationships, especially in the context of animal domestication.[10] For example, whole-genome sequence data from ancient animals have clarified the debated origins of several species, including horses[11] and dogs,[12,13] as well as their further geographic spread post-domestication and differentiation into modern breeds.[14] While this research has typically concentrated on charting changing patterns of genetic ancestry through space and time, underpinning the expansion of particular bloodlines,[15] it has more recently focused on unfolding past breeding management practices.[16] In goats, the identification of elevated inbreeding and shared genetic relatedness among animals not otherwise associated with major skeletal changes has indicated earlier-than-expected attempts of animal management at Ganj Dareh, Iran, around ~10,200 years ago.[17] Complete genome sequencing of the ~2,300-year-old horses buried at Berel', Kazakhstan, has also illuminated the breeding and ritual practices of Scythian Pazyryk, within which diverse bloodlines of various coat colorations were formed and sacrificed in the absence of inbreeding and male-driven selection.[18] Moreover, genotyping data for loci underlying key traits-of-interest have opened for the prediction of phenotypes that do not otherwise fossilize, ranging from the physiological (e.g. speed[19]), to the esthetical (e.g. color[20] and height[21]) and behavioral (e.g. aggressivity[11]), and more (e.g. ambling[22]).

[1]Centre d'Anthropobiologie et de Génomique de Toulouse, CNRS UMR5288, Université Paul Sabatier, 31000 Toulouse, France

[2]Genoscope, Institut de biologie François Jacob, CEA, Université d'Évry, Université Paris-Saclay, 91042 Évry, France

[3]INRAE Division Ecology and Biodiversity (ECODIV), Castanet Tolosan, France

[4]Laboratory of Animal (Poultry) Genetics Breeding and Reproduction, Ministry of Agriculture, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing 100193, P. R. China

[5]Daicel Arbor Biosicences, Ann Arbor, MI, USA

[6]Archéozoologie, Archéobotanique: sociétés, pratiques et environnements (AASPE), Muséum National d'Histoire Naturelle, CNRS, CP 56, 55 Rue Buffon, 75005 Paris, France

[7]Lead contact

*Correspondence: ludovic.orlando@univ-tlse3.fr

https://doi.org/10.1016/j.isci.2023.106144

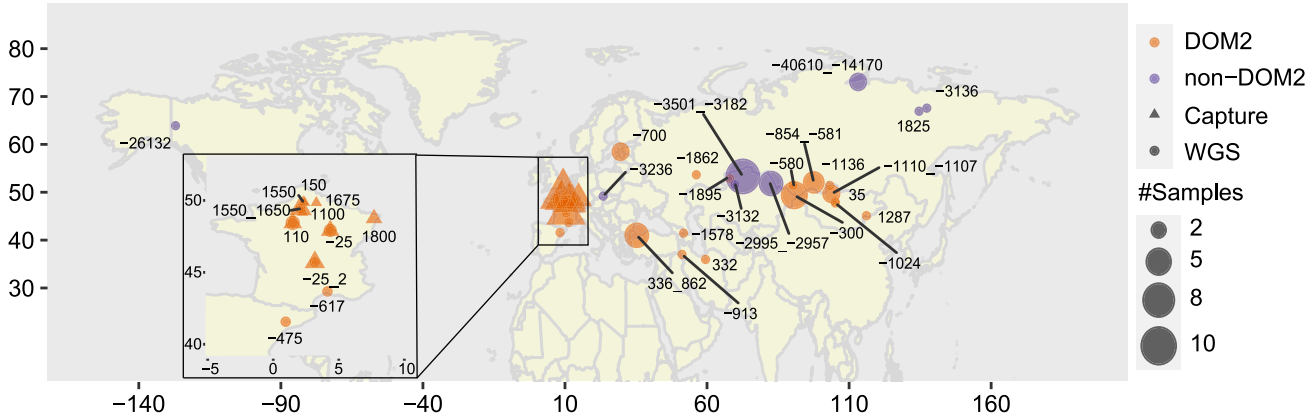


Figure 1. Archaeological samples

Samples subjected to target enrichment (Capture) or whole genome sequencing (WGS) are indicated with triangles and circles, respectively. Symbol sizes are proportional to the number of samples investigated, while the numbers reported indicate their age, in years BCE (Before Common Era), if negative, and years CE (Common Era), if positive. Colors refer to the genetic lineages investigated, with DOM2 referring to modern domestic horses and non-DOM2 to any other lineage. See also Table S1.

No matter how successful ancient DNA analyses have already been, they hold the potential to improve the characterization of past animal and breeding practice even further; since beyond the sole sequence variation, patterns of nucleotide misincorporation at sites that have been damaged after death carry epigenetic information.[23] In particular, following postmortem deamination, CpG sites are converted into TpG sites if methylated, but into UpG sites if not.[24,25] Enzymatic excision of Uracils prevents the incorporation of UpG sites into ancient DNA libraries. This provides an excess of CpG→TpG conversions specifically at those sites that were methylated when the animal died. Statistical packages, such as DamMet,[26] automate the inference of the fraction, F, of cells carrying DNA methylation at any given locus or genomic region of interest from patterns of CpG→TpG conversions in ancient DNA sequence data. Reliable inference requires high-coverage sequence data, which has hitherto limited the approach to only a handful of specimens whose genomes were characterized following extensive sequencing efforts, including anatomically modern humans,[24,27,28] Neanderthals,[25] and Denisovans.[29] Despite DNA methylation clocks characterized across several species, including humans,[30] equids,[31] and other vertebrates,[32] previous attempts to infer the age-at-death of ancient individuals from DNA methylation levels at key CpG loci have largely lacked precision.[27,33] Yet, the age profile of ancient animal communities can provide important insights into past breeding practices, with culling patterns indicating management optimization toward animal production.[34] In addition, differentially methylated regions have been recently reported in castrated and non-castrated males (e.g. in sheep[35]; in horses[36]). This opens the potential for the identification of the castration status of ancient animals on the basis of patterns of DNA methylation.

In this study, we developed the first reliable DNA methylation clock for ancient horses. We first validated our methodology on shotgun sequencing data from 49 ancient horses whose genome was previously characterized at moderate to high coverage. As shotgun sequencing ancient genomes to high coverage is generally not cost effective, we developed a new target-enrichment technology for characterizing 1,611 CpG sites at high coverage from limited DNA sequencing efforts. These sites were selected for showing age-dependent changes in DNA methylation levels and to encompass regions reported to be differentially methylated in castrated and non-castrated males.[36] Our approach is, thus, tailored to predict the age-at-death and the castration status of ancient horse remains from patterns of ancient DNA methylation. It can be extended to any other animal species and provides zooarchaeologists with an augmented range of predictable phenotypes informing on past breeding management strategies.

RESULTS AND DISCUSSION

Whole-genome sequencing (WGS)

Previous work has shown that high-quality sequence data are necessary for DNA methylation inference in ancient individuals.[27] We, thus, took advantage of 49 ancient horse genomes that were previously characterized at relatively high coverage (4.69- to 24.24-fold on average) (Figure 1; Table S1) to infer DNA

methylation levels for 31,836 CpG sites showing age-dependent profiles in modern horses[36] and plains zebras.[31] We also increased the average genome sequence coverage to 6.94- and 7.04-fold for two ancient specimens from Chartres, France, dating to the first–second century CE (GVA53_Fra_110 and GVA60_Fra_110), and sequenced three new genomes from first century BCE specimens from France at 6.63-, 7.13-, and 8.08-fold (GVA661_Fra_m25, GVA607_Fra_m25, and GVA602_Fra_m25, respectively) (Table S1).

DNA methylation levels were inferred using the DamMet statistical package,[26] which provides a probabilistic framework to assess the fraction of cells, F, sequenced showing DNA methylation at a given genomic coordinate. This procedure was previously reported to show best performance on sequence data produced following treatment of ancient DNA extracts with the USER enzymatic mix,[26,33] which largely restricts the sequencing of deaminated cytosines to methylated contexts only. In addition to sequence alignments, this method requires users to indicate the genomic windows of interest and the average fraction of methylated cells in the tissue analyzed. We aimed first to identify the best possible combination of parameters for retrieving accurate DNA methylation estimates, following previous work[26] and exploring F values ranging from 70% to 95%. We also used genomic windows centered on each single CpG site but showing different sizes and containing from a single CpG to up to 50 flanking CpG sites (Figure 2A). The parameter combination maximizing Spearman correlation coefficients between DNA methylation levels estimated for ancient individuals and those measured in modern horses of the same sex was used in downstream analyses (Table S2A). Our overall methodological approach is summarized in the left part of the flowchart shown in Figure 3, and involved further mathematical transformation of the F DNA methylation values returned by DamMet.

These analyses revealed that DNA methylation inference was largely insensitive to the average fraction of methylated cells considered, but could be significantly improved when using genomic windows containing 25–50 CpG sites, in line with previous reports in archaic hominins[25] (Table S2A). The distribution of DNA methylation values across the 31,836 genomic windows was, however, significantly different in ancient and modern individuals (Figure 2B), with the latter showing a lack of sites fully methylated (F = 1) and unmethylated (F = 0). To mitigate this, we built a first statistical model based on Random Forest (RF) regression to transform the ancient DNA methylation values inferred by DamMet,[26] according to the values measured in modern horses, and accounting for the sequence coverage achieved in each genomic window. We then followed previous work[37] and implemented a linear regression scheme relating modern and RF-transformed ancient DNA methylation values as a further mathematical transformation (Figure 2C). This two-steps procedure significantly improved the correlation between ancient and modern DNA methylation values (from 0.04 to 0.68 up to 0.83), despite maintaining a deficit of unmethylated and fully methylated sites.

Reassuringly, the transformed F-values proved sufficiently accurate to build a DNA methylation clock model showing high correlation between the known and the predicted ages of 10 ancient horses (Pearson correlation coefficient = 0.97; Figure 2D), with median absolute errors (MAE) of 1.94 years (Figure 2D). This error range corresponds to approximately ~6.5% of the average horse lifespan (~30 years), and ~3.1% of the longest-lived horse known to date (62 years). The underlying DNA methylation clock corresponds to a linear and additive combination of the DNA methylation levels (following the two-steps transformation) at 29,113 CpG sites for the 10 ancient horses of known age-at-death, as assessed from patterns of tooth wear,[2] and normalized DNA methylation values for 239 modern horses of known date-of-birth (0.005–34 years old). We assessed the prediction precision using bootstrap resampling of the CpG sites considered, given that ancient DNA data often show uneven sequence coverage and patterns of missingness (100 pseudo-replicates; Figure S1). It revealed overestimated ages for young horses below 5 years, and underestimated ages for adults over 12 years (Figure 2E). Applying this clock model to 54 ancient horses of unknown age predicted ages spanning 2 to 18 years, with corresponding error margins of 0.37–0.79 years, consistent across all age categories (Figure 2F).

Overall, the DNA methylation clock model reconstructed here shows good performance but overestimates the age of foals and underestimates the age of old adults. It, nonetheless, provides a reliable methodology to sort ancient fragmentary remains according to increasing age categories, even in the absence of dental remains allowing classical age determination in zooarchaeology.

In-solution target enrichment

DNA preservation of the vast majority of archaeological remains is not compatible with a cost-effective characterization of high-quality genomes using shotgun sequencing.[8] To expand the potential of our

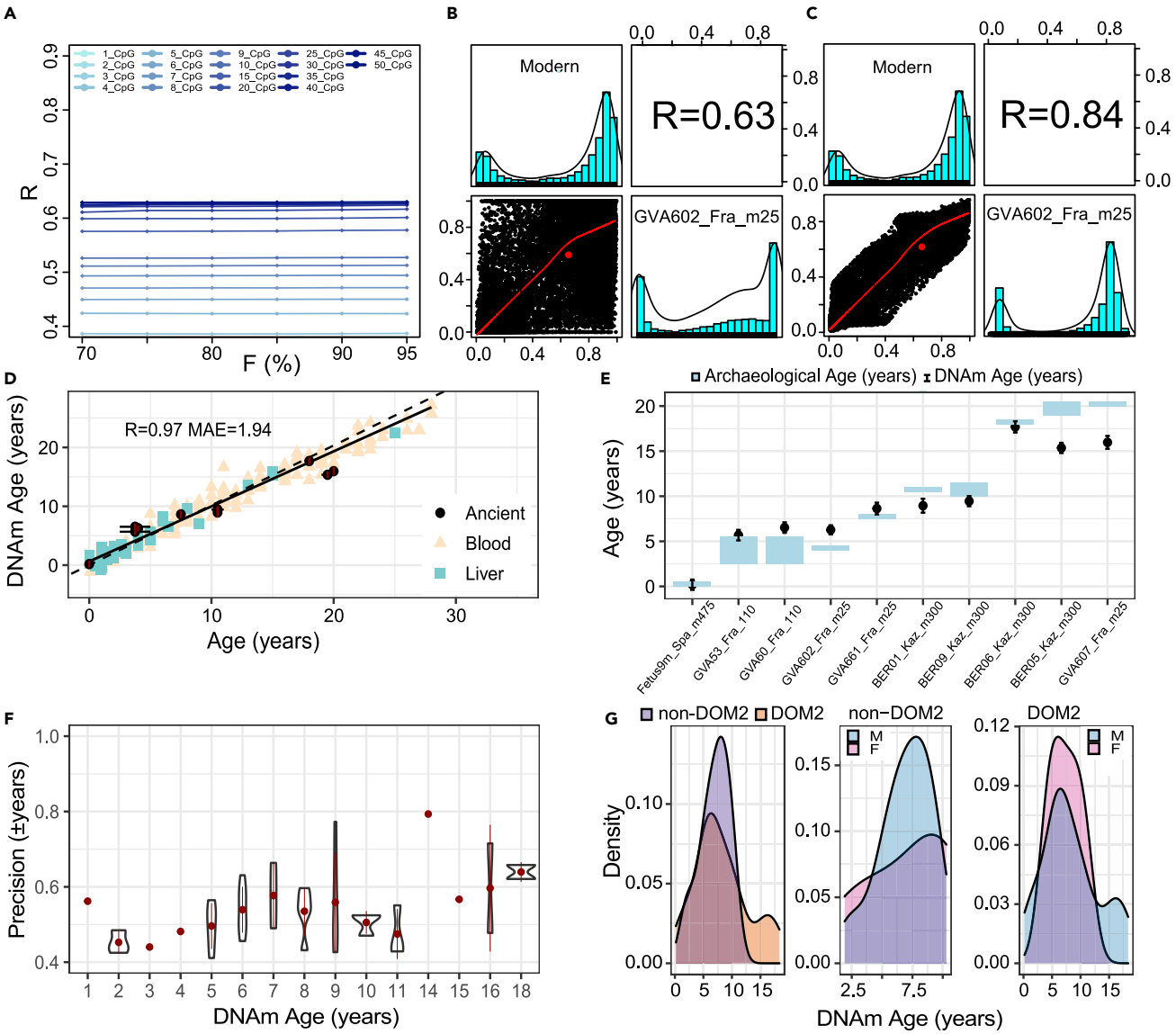


Figure 2. Age-at-death predictions based on WGS data

(A) Spearman correlation coefficient R between modern and raw ancient DNA methylation values (F%) centered at 31,836 CpG sites, for sample GVA602_Fra_m25 and window sizes including 1 to 50 CpGs.

(B) Distribution of F values averaged across modern male individuals and raw F values predicted for sample GVA602_Fra_m25, and their Spearman correlation coefficient R.

(C) Same as (B), after the two-step mathematical transformation of ancient F DNA methylation values.

(D) DNA methylation clock and performance in 10 ancient individuals of known age-at-death. Modern blood and liver DNA methylation values are shown in light yellow and blue, respectively. Black horizontal bars indicate the temporal span of individual archaeological ages. Red vertical bars provide the precision of predicted methylation age.

(E) Age prediction for 10 ancient individuals. Zooarchaeological ages-at-death are indicated in blue for comparison to those inferred using DNA methylation clocks (black).

(F) Precision of predicted age for 50 ancient individuals.

(G) Age mortality profiles in ancient DOM2 and non-DOM2 horses (left), ancient non-DOM2 males and females (center), and ancient DOM2 males and females (right).

See also Figure S1, Table S2.

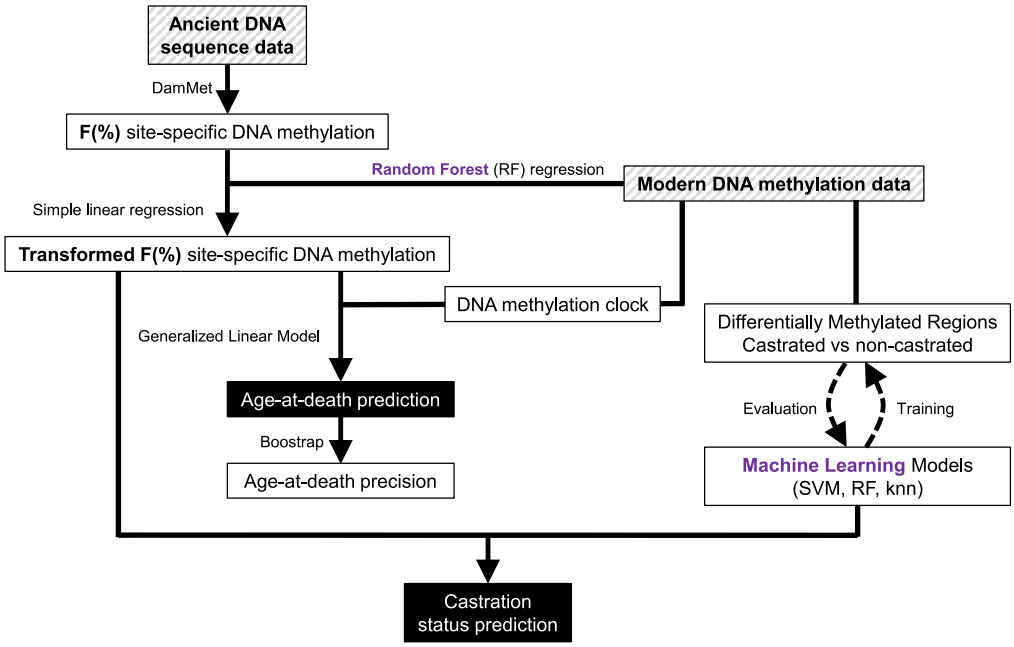


Figure 3. Overview of prediction methods

Ancient DNA sequence data are converted into DNA methylation F values for a number of preselected CpG sites with known DNA methylation levels in modern horses. The age and castration status of modern horses are also known. Ancient F values follow a two-steps mathematical transformation consisting of Random Forest (RF) and simple linear regressions, before age-at-death prediction using a Generalized Linear Model explaining age-at-death as an additive function of DNA methylation levels. The prediction of the castration status is based on several machine learning models trained on regions differentially methylated in modern horses between castrated and non-castrated horses. The model showing best prediction performance in a panel of 26 modern horses is finally applied on the transformed F values to predict the castration status of each individual ancient horse.

See also STAR Methods.

methodology to a wider range of ancient remains, we developed an in-solution target-enrichment assay capturing the ancient DNA fragments present in libraries for a subset of 1,611 of the 31,836 CpG sites considered above (~5%, Table S3A, Figure S2). This approach was expected to significantly reduce the experimental costs while increasing local sequence coverage by focusing sequencing efforts only on a fraction of the sites. DNA probes were designed following Haak and colleagues (2015),[38] and consisted of two 60-mers flanking each targeted CpG site, plus two additional ones centered on the CpG site, each carrying one of the two possible sequence variants (i.e. CpG and TpG). The latter two were selected since methylated cytosines that are deaminated postmortem at CpG sites are sequenced as TpGs following USER enzymatic treatment, while those left undamaged are sequenced as CpGs.[24]

Our assay retrieved good sequence coverage across all targeted CpG sites (17.20- to 195.34-fold) for a total of 33 ancient horse remains from France and spanning the last ~2,300 years. The sequence data proved sufficient to infer DNA methylation at 2,171 CpG sites, due to their physical proximity to the 1,611 CpG sites originally targeted. Repeating the methodology described above on the captured sequence data, all the findings could be recapitulated, excepting that the correlation between the modern and the transformed ancient DNA methylation values was stronger than when using WGS data (Spearman correlation coefficient = 0.92 vs. 0.84; Figures 2C and 4C). In addition, genomic windows containing 15–50 CpG sites (instead of 25–50) were found to be most appropriate for mathematical transformation, likely due to the higher coverage achieved at each site after target enrichment (Figures 4A–4C; Table S3B). Following target enrichment, the final DNA methylation clock was calibrated for a total of 2,171 CpG sites and assessed by comparing the known and predicted ages in a subset of 25 ancient horses (Figures 4D–4G). It showed reasonable performance (Pearson correlation coefficient = 0.77), while the trend to overestimate (underestimate) young (old) age categories persisted. MAE

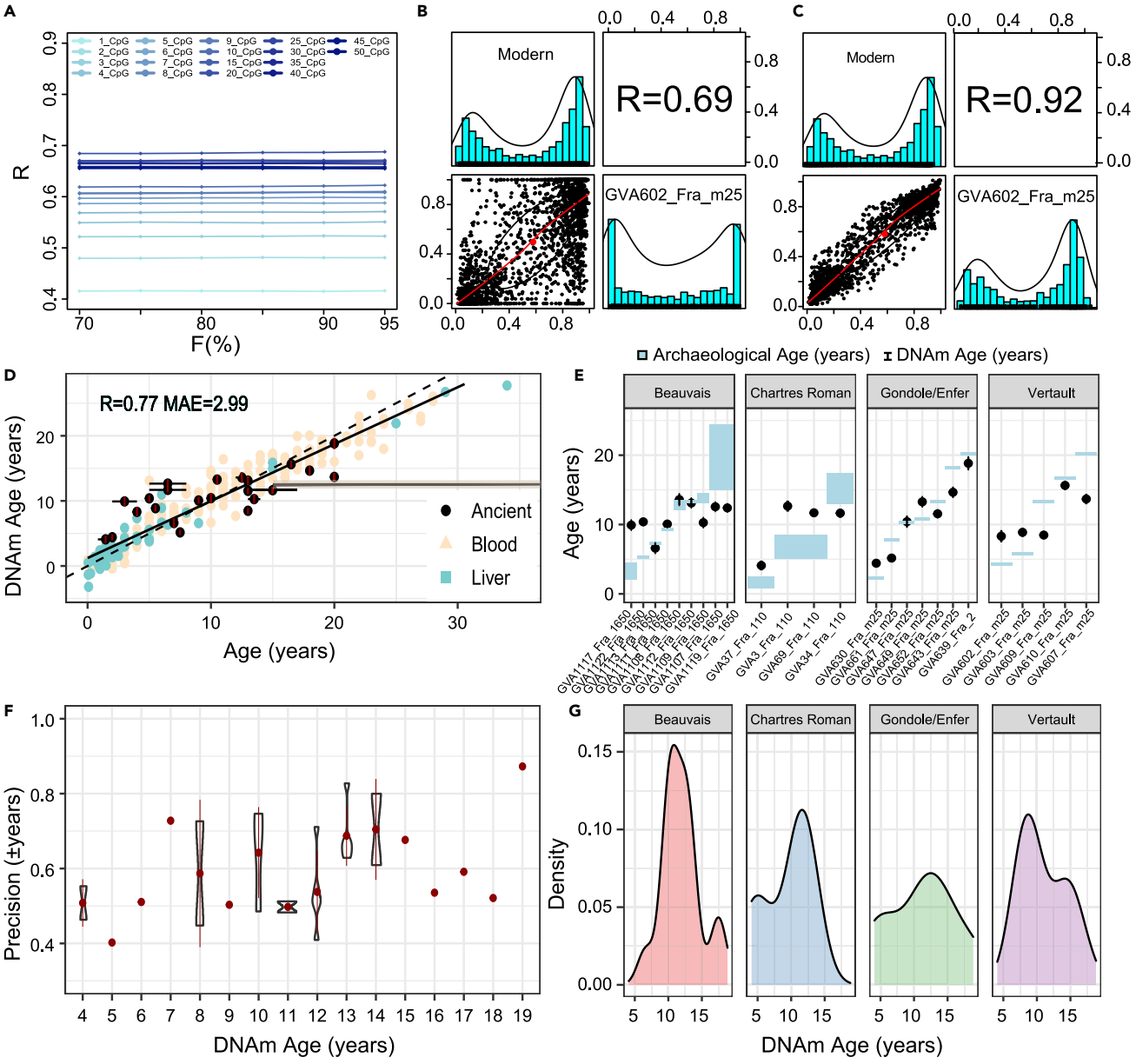


Figure 4. Age-at-death prediction based on target-enrichment data

(A–F) same as Figure 2, except that 33 ancient horses from France were analyzed. Panels B and C compare F DNA methylation values averaged across modern male individuals and those inferred prior to and after mathematical transformation in sample GVA602_Fra_m25.

(G) Age mortality profiles amongst ancient DOM2 horses found in four archaeological contexts.

See also Figures S1 and S2, and Table S3.

were, however, inflated when compared to the clock model constructed from WGS data (2.99 years vs. 1.94 years, corresponding to ~10.0% and ~6.5% of the average life expectancy in horses, respectively). This is due to the more limited number of sites that were included in the reconstructions (Figures 4D–4F). In the future, we expect that extending our target-enrichment assay to the whole set of 31,836 CpG could improve the prediction accuracy at rather moderate experimental costs. Our current assay provides, however, the first cost-effective experimental solution to assess age mortality profiles of ancient horse populations with a less-than-one-year precision across all age categories represented in our ancient panel (Figures 4G and S1).

Ancient age mortality profiles

Our DNA methylation clocks provided a first opportunity to explore the age mortality profiles of 84 ancient horse specimens, including 52 for which the age could not be determined from patterns of dental wear (Figures 1 and 2; Table S2B). WGS data did not support statistically different age mortality profiles in 34 domestic horses belonging to the DOM2 lineage, and 20 horses belonging to other genetic lineages (non-DOM2, excluding four showing missing data levels >10%; Figure 2G; Kruskal-Wallis test, p-value = 0.54). In addition, no statistically significant difference was found between sex in both genetic lineages considered (p-value = 0.78, and 0.69). The DOM2 lineage was domesticated in the lower Don-Volga region around ~4,200 years ago,[11] and rapidly spread across Eurasia together with new equestrian technologies, including warfare chariots, replacing almost all other ancient lineages. Their spread was accompanied with an over-representation of males in the archaeological assemblages,[39] possibly following the extension of rising gender inequalities in Bronze Age societies to prestige animals such as horses. With only 30 DOM2 males and 4 DOM2 females, all sampled across a wide variety of archaeological contexts, the statistical power for detecting possible difference in the age-at-death of DOM2 males and females is limited. Therefore, future work should examine a more extensive panel of ancient horses, especially from sacrificial DOM2 sites, before concluding that the over-representation of males detected in horse assemblages did not focus on specific age categories.

Similarly, target-enrichment sequence data indicated no statistical differences in the age mortality profiles of various Iron Age, Roman, Medieval, and pre-Modern archaeological contexts, encompassing a wide range of activities, from sacrificial burial sites (Gondole/L'Enfer and Vertault; N = 12) to food wastes, utilitarian, and secular deposits. At present, these data, thus, support Iron Age ritual sacrificial practices not focused on old animals but on animals that were on average 11.18 years old and could, thus, still provide important services to the community. In addition, while horses of almost all age categories (6 to 17 years old) were found in rendering waste deposits at Beauvais, their vast majority consisted of animals older than 10 years old (Figure 2G). These animals were likely even older given the trend of our methodology to underestimate the age of old animals. This finding indicates that horses were extensively used in Medieval and pre-Modern contexts up until they died.

Sex determination and castration

Males and females exhibit different DNA methylation patterns on the X chromosome,[40] which provides an additional opportunity to validate the mathematically transformed DNA methylation values inferred for the ancient horses. We used WGS data and X-to-autosomal coverage ratios to determine their genetic sex, following the study by Schubert et al.[41] Principal component analysis of DNA methylation at the 960 CpG sites located on the X chromosome present in our modern comparative panel revealed clear clustering of 42 ancient males and 8 ancient females with modern animals of the same sex. Two of the 333 modern horses that were previously reported as males[36] were, however, found to group with females, likely indicating erroneous sex assignment in the original data repository (Figure 5A). These samples were, thus, disregarded in the following analyses. Interestingly, modern males are characterized by an excess of hypomethylated CpG sites on the X chromosome, while females show a larger fraction associated with DNA methylation ranging between 20% and 80% (Figure 5B). The same profile was retrieved across the 50 ancient horses investigated, further confirming the quality of the DNA methylation values obtained following the two-steps transformation procedure described above (Figure 5C).

DNA methylation patterns along the X chromosome appeared, however, similar between uncastrated (i.e. intact stallions) and castrated males (Figures 5A and S3), precluding their use as castration markers. To assess the potential of DNA methylation for castration prediction, we instead relied on CpG sites identified in a previous EWAS analysis (epigenetic-wide association studies) to show different aging trajectories in uncastrated and castrated males[42] (Tables S4A–S4C). Multidimensional scaling of DNA methylation values in modern horses confirmed the non-overlapping clustering of intact stallions and castrated males aged above 11 years old (Figures 5D and S4). These sites failed, however, to discriminate younger males according to their castration status (Figure 5E). We, thus, restricted our analyses to the 8 males with WGS data that were inferred to have died at a minimal age of 11 on the basis of our DNA methylation clock (WGS data).

We next used three machine learning approaches aimed at classifying castrated males and intact stallions (Figure 3). These approaches included Support Vector Machine (SVM), Random Forest (RF), and Generalized k-Nearest Neighbors (GKNN) algorithms, and were trained on the panel of 26 modern males.[36] We

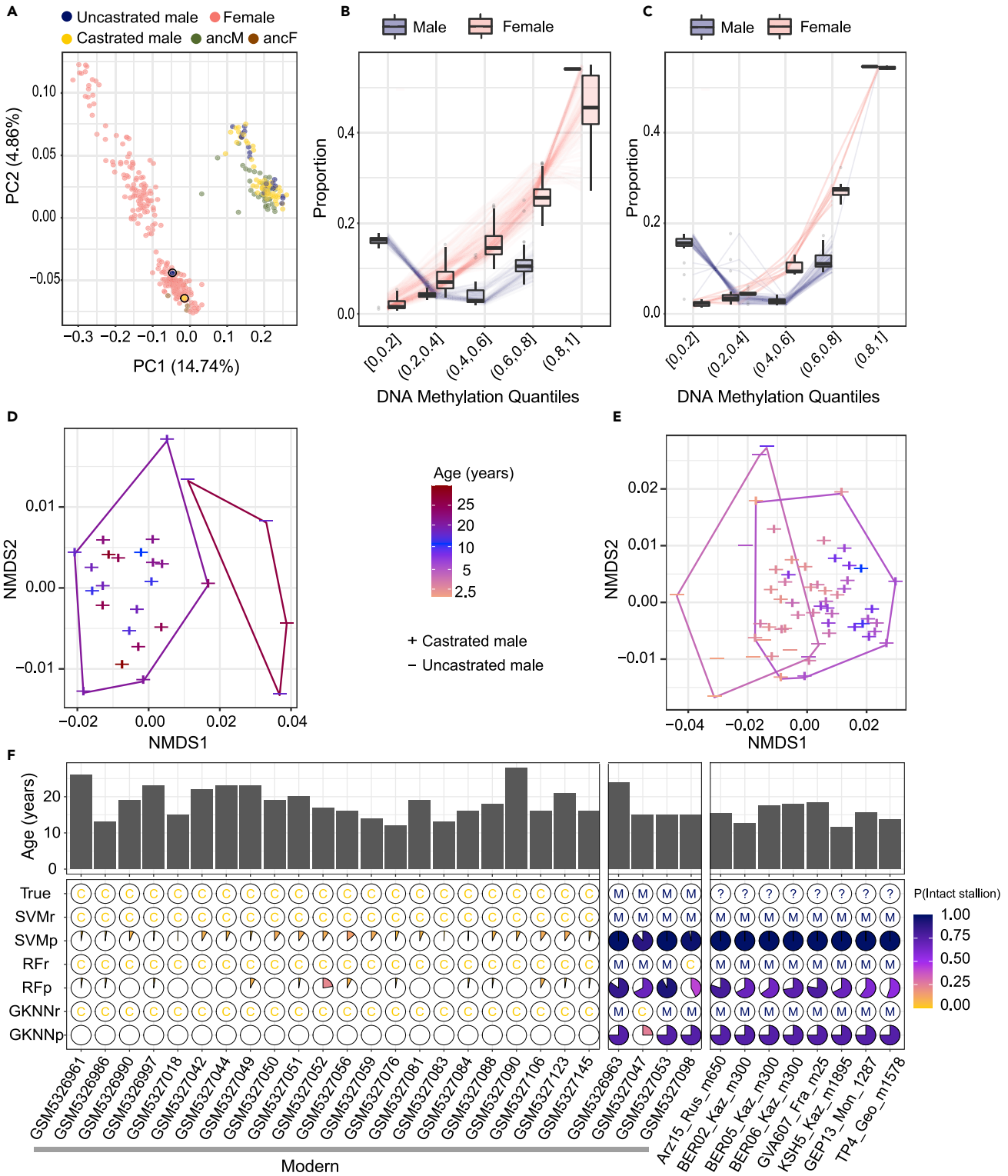


Figure 5. Sex and castration predictions based on ancient DNA methylation

(A) PCA on DNA methylation values at 960 CpG sites located on the X chromosome. Modern females, and (un)castrated males are indicated, as well as ancient males (ancM) and females (ancF). Two black circles in the central cluster at the bottom of the figure highlight those modern females likely mislabeled as males.

(B) Distribution of five DNA methylation quantiles in modern males and females for the X chromosome.

Figure 5. *Continued*
(C) Same as B), but reporting transformed F values for 50 ancient horses.
(D–F) Castration prediction based on ancient DNA methylation for the 59 candidate CpGs selected considering association level |R| ≥0.55. D) MDS clustering of modern castrated and uncastrated males (>11 years-old). (E) Same as (D), for younger horses. (F) Machine learning predictions of the male castration status. The barplot indicates the DNA methylation age of 26 modern and 8 ancient males. Pie chart proportions reflect the probability for a male to be castrated (C), or not (M).
See also Figures S3 and S4 and Table S4.

assessed the prediction accuracy using a cross-validation approach based on a training set of 80% of the individuals and performing predictions for the remaining 20% (Table S4C). We caution that the data available for training should be extended in a larger number of male individuals, especially in castrated males, before the exact performance of each methodology can be fully appreciated. Both RF and GKNN returned substantial classification errors, and average assignment probabilities below 0.25 and 0.425, respectively (Figure 5F). In contrast, SVM only returned the expected predictions, with high average assignment probabilities to both categories (0.93 for intact stallions, and 0.96 for castrated males; Figure 5F). Therefore, we used the SVM approach to assess the castration status of 8 ancient horses aged above 11 and showing sufficient WGS data. None of these specimens were predicted to be castrated, including three Pazyryk horses from Berel', Kazkhstan, a 2,300-year-old funerary ritual site known for the sacrifice of males.[43] These results, thus, contradict previous contention based on the slenderness of post-cranial skeletal elements,[44] which portrayed these males as castrated. Since the horses buried encompassed a full range of coat coloration and age categories ([18]; this study), and did not represent a full family genealogy, we conclude that the exact ritual practices underlying Pazyryk horse sacrifice remain to be deciphered.

Conclusions

In this study, we present both wet-lab and dry-lab procedures aimed at improving DNA methylation inference in ancient horse specimens. Our methodology returns estimates for the age-at-death that correlate with standard estimates based on patterns of dental wear. It shows reasonable precision, within less than one year based on both the full set of CpG sites and the reduced panel targeted in our DNA capture assay. In contrast to standard approaches, our methodology is not restricted to teeth only, but can be applied to other osteological remains, including the most fragmentary, provided that ancient DNA is preserved. We also extend the current zooarchaeological toolkit for identifying castrated individuals. This work paves the way for a more thorough characterization of past husbandry practices, taking full advantage of standard archaeological approaches as well as genetic evidence, revealing phenotypes, admixture and selection practices, and epigenetic inference, informing on age-at-death and castration. Extended to human archaeological remains, our approach holds the potential to measure changing age mortality profiles in past societies in the face of major epidemiological outbreaks, as well as cultural and demographic transitions.

Limitations of study

While the horse DNA methylation clocks considered here are largely consistent in two somatic tissues, DNA methylation values for modern horses are not yet available for osseous remains, such as those used to gather ancient DNA sequence data. Future work should focus on generating DNA methylomes from modern horse osseous and dental tissues to avoid possible tissue-specific changes in aging trajectories. Furthermore, the DNA methylation clocks considered here were constructed using the elastic linear regression between DNA methylation and age, as most commonly done. Whether other approaches, including nonlinear models, can improve prediction remains to be tested. Finally, castration predictions are based on a limited panel including 26 modern horse individuals, only four of which are castrated. More extensive panels must be considered before the performance of the method presented can be fully assessed.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Archaeological information

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106144.

ACKNOWLEDGMENTS

We thank Claudia Gillet and Dr Naveed Khan for technical assistance. This work was supported by CNRS and University Paul Sabatier (AnimalFarm International Research Program, IRP), the France Génomique "Grands Projets" program (BUCEPHALE), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 681605 - PEGASUS). Dr Xuexue Liu and Dr Yvette Running Horse Collin were supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreements 101027750 (HOPE) and 890702 (MethylRIDE), respectively. The Genoscope sequencing platform is partly funded by the France Génomique National infrastructure, funded as part of "Investissement d'avenir" program managed by Agence Nationale pour la Recherche (ANR-10-INBS-09).

AUTHOR CONTRIBUTIONS

L.O. conceived the project and designed research. J.-M.A., A.P., P.W., and L.O. provided material and reagents. S.L. and B.C. provided samples and information about archaeological contexts. L.O. inferred DNA methylation levels. X.L. carried out all other computational analyses, with input from L.O. X.L. drafted the STAR Methods. A.S.-O. and C.D.S. extensively revised the STAR Methods, with input from L.O. X.L., A.S.-O., G.T., S.D., L.T., L.C., S.W., C.G., P.C., O.E., and C.D.S. carried out ancient DNA laboratory work. J.E., A.D., and J.K. performed in-solution target enrichment at Daicel Arbor Biosciences. X.L., J.P., L.J., and Y.M. prepared the figures and tables. L.O. wrote the article, with input from X.L., A.S.-O., Y.R.H.C., C.D.S., and all co-authors.

DECLARATION OF INTERESTS

J.E., A.D., and J.K. are employed by Daicel Arbor Biosciences, who produced the targeted-capture panel and performed the capture laboratory services for this study.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location. We avoided "helicopter science" practices by including the participating local contributors from the region where we conducted the research as authors on the paper.

Received: September 27, 2022
Revised: January 2, 2023
Accepted: February 1, 2023
Published: February 5, 2023

REFERENCES

1. Steele, T.E. (2015). The contributions of animal bones from archaeological sites: the past and future of zooarchaeology. J. Archaeol. Sci. *56*, 168–176. https://doi.org/10.1016/j.jas.2015.02.036.

2. Levine, M. (1982). The use of crown height measurements and eruption-wear sequences to age horse teeth. In Ageing and Sexing Animal Bones from Archaeological Sites, pp. 223–255. https://doi.org/10.30861/9780860541929.

3. Toledo González, V., Ortega Ojeda, F., Fonseca, G.M., García-Ruiz, C., and Pérez-Lloret, P. (2021). Analysis of tooth mark patterns on bone remains caused by wolves (Canis lupus) and domestic dogs (Canis lupus familiaris) for taxonomic identification: a scoping review focused on their value as a forensic tool. Appl. Anim. Behav. Sci. *240*, 105356. https://doi.org/10.1016/j.applanim.2021.105356.

4. Taylor, W.T.T., Tuvshinjargal, T., and Bayarsaikhan, J. (2016). Reconstructing equine bridles in the Mongolian bronze age. J. Ethnobiol. *36*, 554–570. https://doi.org/10.2993/0278-0771-36.3.554.

5. Taylor, W., Fantoni, M., Marchina, C., Lepetz, S., Bayarsaikhan, J., Houle, J.-L., Pham, V., and Fitzhugh, W. (2020). Horse sacrifice and butchery in bronze age Mongolia. J. Archaeol. Sci. Rep. *31*, 102313. https://doi.org/10.1016/j.jasrep.2020.102313.

6. Wißing, C., Rougier, H., Baumann, C., Comeyne, A., Crevecoeur, I., Drucker, D.G., Gaudzinski-Windheuser, S., Germonpré, M., Gómez-Olivencia, A., Krause, J., et al. (2019). Stable isotopes reveal patterns of diet and mobility in the last Neandertals and first modern humans in Europe. Sci. Rep. *9*, 4433. https://doi.org/10.1038/s41598-019-41033-3.

7. Harju, S., Olson, C.V., Hess, J., and Webb, S.L. (2021). Isotopic analysis reveals landscape patterns in the diet of a subsidized predator, the common raven. Ecol. Solut. Evid. *2*, e12100. https://doi.org/10.1002/2688-8319.12100.

8. Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P.W., Ávila-Arcos, M.C., Fu, Q., Krause, J., Willerslev, E., Stone, A.C., and Warinner, C. (2021). Ancient DNA analysis. Nat. Rev. Methods Primers *1*, 14. https://doi.org/10.1038/s43586-020-00011-0.

9. Ermini, L., Der Sarkissian, C., Willerslev, E., and Orlando, L. (2015). Major transitions in human evolution revisited: a tribute to ancient DNA. J. Hum. Evol. *79*, 4–20. https://doi.org/10.1016/j.jhevol.2014.06.015.

10. Frantz, L.A.F., Bradley, D.G., Larson, G., and Orlando, L. (2020). Animal domestication in the era of ancient genomics. Nat. Rev. Genet. *21*, 449–460. https://doi.org/10.1038/s41576-020-0225-0.

11. Librado, P., Khan, N., Fages, A., Kusliy, M.A., Suchan, T., Tonasso-Calvière, L., Schiavinato, S., Alioglu, D., Fromentier, A., Perdereau, A., et al. (2021). The origins and spread of domestic horses from the Western Eurasian steppes. Nature *598*, 634–640. https://doi.org/10.1038/s41586-021-04018-9.

12. Frantz, L.A.F., Mullin, V.E., Pionnier-Capitan, M., Lebrasseur, O., Ollivier, M., Perri, A., Linderholm, A., Mattiangeli, V., Teasdale, M.D., Dimopoulos, E.A., et al. (2016). Genomic and archaeological evidence suggest a dual origin of domestic dogs. Science *352*, 1228–1231. https://doi.org/10.1126/science.aaf3161.

13. Bergström, A., Stanton, D.W.G., Taron, U.H., Frantz, L., Sinding, M.-H.S., Ersmark, E., Pfrengle, S., Cassatt-Johnstone, M., Lebrasseur, O., Girdland-Flink, L., et al. (2022). Grey wolf genomic history reveals a dual ancestry of dogs. Nature *607*, 313–320. https://doi.org/10.1038/s41586-022-04824-9.

14. Frantz, L.A.F., Haile, J., Lin, A.T., Scheu, A., Geörg, C., Benecke, N., Alexander, M., Linderholm, A., Mullin, V.E., Daly, K.G., et al. (2019). Ancient pigs reveal a near-complete genomic turnover following their introduction to Europe. Proc. Natl. Acad. Sci. USA *116*, 17231–17238. https://doi.org/10.1073/pnas.1901169116.

15. Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., McCrory Constantz, C., Gamba, C., Al-Rasheid, K.A.S., Albizuri, S., et al. (2019). Tracking five millennia of horse management with extensive ancient genome time series. Cell *177*, 1419–1435.e31. https://doi.org/10.1016/j.cell.2019.03.049.

16. Todd, E.T., Tonasso-Calvière, L., Chauvey, L., Schiavinato, S., Fages, A., Seguin-Orlando, A., Clavel, P., Khan, N., Pérez Pardal, L., Patterson Rosa, L., et al. (2022). The genomic history and global expansion of domestic donkeys. Science *377*, 1172–1180. https://doi.org/10.1126/science.abo3503.

17. Daly, K.G., Mattiangeli, V., Hare, A.J., Davoudi, H., Fathi, H., Doost, S.B., Amiri, S., Khazaeli, R., Decruyenaere, D., Nokandeh, J., et al. (2021). Herded and hunted goat genomes from the dawn of domestication in the Zagros Mountains. Proc. Natl. Acad. Sci. USA *118*. e2100901118. https://doi.org/10.1073/pnas.2100901118.

18. Librado, P., Gamba, C., Gaunitz, C., Der Sarkissian, C., Pruvost, M., Albrechtsen, A., Fages, A., Khan, N., Schubert, M., Jagannathan, V., et al. (2017). Ancient genomic changes associated with domestication of the horse. Science *356*, 442–445. https://doi.org/10.1126/science.aam5298.

19. Bower, M.A., McGivney, B.A., Campana, M.G., Gu, J., Andersson, L.S., Barrett, E., Davis, C.R., Mikko, S., Stock, F., Voronkova, V., et al. (2012). The genetic origin and history of speed in the Thoroughbred racehorse. Nat. Commun. *3*, 643. https://doi.org/10.1038/ncomms1644.

20. Marklund, L., Moller, M.J., Sandberg, K., and Andersson, L. (1996). A missense mutation in the gene for melanocyte-stimulating hormone receptor (MCIR) is associated with the chestnut coat color in horses. Mamm. Genome *7*, 895–899. https://doi.org/10.1007/s003359900264.

21. Liu, X., Zhang, Y., Liu, W., Li, Y., Pan, J., Pu, Y., Han, J., Orlando, L., Ma, Y., and Jiang, L. (2022). A single-nucleotide mutation within the TBX3 enhancer increased body size in Chinese horses. Curr. Biol. *32*, 480–487.e6. https://doi.org/10.1016/j.cub.2021.11.052.

22. Wutke, S., Benecke, N., Sandoval-Castellanos, E., Döhle, H.J., Friederich, S., Gonzalez, J., Hallsson, J.H., Hofreiter, M., Lõugas, L., Magnell, O., et al. (2016). Spotted phenotypes in horses lost attractiveness in the Middle Ages. Sci. Rep. *6*, 38548. https://doi.org/10.1038/srep38548.

23. Hanghøj, K., and Orlando, L. (2018). Ancient epigenomics. In Paleogenomics (Springer), pp. 75–111. https://doi.org/10.1007/13836_2018_18.

24. Pedersen, J.S., Valen, E., Velazquez, A.M.V., Parker, B.J., Rasmussen, M., Lindgreen, S., Lilje, B., Tobin, D.J., Kelly, T.K., Vang, S., et al. (2014). Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. Genome Res. *24*, 454–466. https://doi.org/10.1101/gr.163592.113.

25. Gokhman, D., Lavi, E., Prüfer, K., Fraga, M.F., Riancho, J.A., Kelso, J., Pääbo, S., Meshorer, E., and Carmel, L. (2014). Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. Science *344*, 523–527. https://doi.org/10.1126/science.1250368.

26. Hanghøj, K., Renaud, G., Albrechtsen, A., and Orlando, L. (2019). DamMet: ancient methylome mapping accounting for errors, true variants, and post-mortem DNA damage. GigaScience *8*, giz025. https://doi.org/10.1093/gigascience/giz025.

27. Seguin-Orlando, A., Donat, R., Der Sarkissian, C., Southon, J., Thèves, C., Manen, C., Tchérémissinoff, Y., Crubézy, E., Shapiro, B., Deleuze, J.-F., et al. (2021). Heterogeneous hunter-gatherer and steppe-related ancestries in late neolithic and bell beaker genomes from present-day France. Curr. Biol. *31*, 1072–1083.e10. https://doi.org/10.1016/j.cub.2020.12.015.

28. Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., et al. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature *463*, 757–762. https://doi.org/10.1038/nature08835.

29. Gokhman, D., Mishol, N., de Manuel, M., de Juan, D., Shuqrun, J., Meshorer, E., Marques-Bonet, T., Rak, Y., and Carmel, L. (2019). Reconstructing Denisovan anatomy using DNA methylation maps. Cell *179*, 180–192.e10. https://doi.org/10.1016/j.cell.2019.08.035.

30. Horvath, S. (2013). DNA methylation age of human tissues and cell types. Genome Biol. *14*, R115–R120. https://doi.org/10.1186/gb-2013-14-10-r115.

31. Larison, B., Pinho, G.M., Haghani, A., Zoller, J.A., Li, C.Z., Finno, C.J., Farrell, C., Kaelin, C.B., Barsh, G.S., Wooding, B., et al. (2021). Epigenetic models developed for plains zebras predict age in domestic horses and endangered equids. Commun. Biol. *4*, 1412. https://doi.org/10.1038/s42003-021-02935-z.

32. Horvath, S., Lu, A.T., Haghani, A., Zoller, J.A., Li, C.Z., Lim, A.R., Brooke, R.T., Raj, K., Serres-Armero, A., Dreger, D.L., et al. (2022). DNA methylation clocks for dogs and humans. Proc. Natl. Acad. Sci. USA *119*. e2120887119. https://doi.org/10.1073/pnas.2120887119.

33. Hanghøj, K., Seguin-Orlando, A., Schubert, M., Madsen, T., Pedersen, J.S., Willerslev, E., and Orlando, L. (2016). Fast, accurate and automatic ancient nucleosome and methylation maps with epiPALEOMIX. Mol. Biol. Evol. *33*, 3284–3298. https://doi.org/10.1093/molbev/msw184.

34. Brusgaard, N.Ø., Dee, M.W., Dreshaj, M., Erven, J., van den Hurk, Y., Raemaekers, D., and Çakırlar, C. (2022). Hunting before herding: a zooarchaeological and stable isotopic study of suids (Sus sp.) at Hardinxveld-Giessendam, The Netherlands (5450–4250 cal BC). PLoS One *17*, e0262557. https://doi.org/10.1371/journal.pone.0262557.

35. Sugrue, V.J., Zoller, J.A., Narayan, P., Lu, A.T., Ortega-Recalde, O.J., Grant, M.J., Bawden, C.S., Rudiger, S.R., Haghani, A., Bond, D.M., et al. (2021). Castration delays epigenetic aging and feminizes DNA methylation at androgen-regulated loci. Elife *10*, e64932. https://doi.org/10.7554/eLife.64932.

36. Horvath, S., Haghani, A., Peng, S., Hales, E.N., Zoller, J.A., Raj, K., Larison, B., Robeck, T.R., Petersen, J.L., Bellone, R.R., and Finno, C.J. (2022). DNA methylation aging and transcriptomic studies in horses. Nat. Commun. *13*, 40. https://doi.org/10.1038/s41467-021-27754-y.

37. Song, J. (2015). Bias corrections for Random Forest in regression using residual rotation. J. Korean Stat. Soc. *44*, 321–326. https://doi.org/10.1016/j.jkss.2015.01.003.

38. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature *522*, 207–211. https://doi.org/10.1038/nature14317.

39. Fages, A., Seguin-Orlando, A., Germonpré, M., and Orlando, L. (2020). Horse males became over-represented in archaeological assemblages during the Bronze Age. J. Archaeol. Sci. Rep. *31*, 102364. https://doi.org/10.1016/j.jasrep.2020.102364.

40. Sharp, A.J., Stathaki, E., Migliavacca, E., Brahmachary, M., Montgomery, S.B., Dupre, Y., and Antonarakis, S.E. (2011). DNA methylation profiles of human active and inactive X chromosomes. Genome Res. *21*, 1592–1600. https://doi.org/10.1101/gr.112680.110.

41. Schubert, M., Mashkour, M., Gaunitz, C., Fages, A., Seguin-Orlando, A., Sheikhi, S., Alfarhan, A.H., Alquraishi, S.A., Al-Rasheid, K.A., Chuang, R., et al. (2017). Zonkey: a simple, accurate and sensitive pipeline to genetically identify equine F1-hybrids in archaeological assemblages. J. Archaeol. Sci. *78*, 147–157. https://doi.org/10.1016/j.jas.2016.12.005.

42. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. *9*, 559. https://doi.org/10.1186/1471-2105-9-559.

43. Lepetz, S. (2013). Horse sacrifice in a Pazyryk culture Kurgan: the princely tomb of Berel'(Kazakhstan). selection Criteria and slaughter procedures. Anthropozoologica *48*, 309–321. https://doi.org/10.5252/az2013n2a9.

44. Chechushkov, I.V., Usmanova, E.R., and Kosintsev, P.A. (2020). Early evidence for horse utilization in the Eurasian steppes and the case of the Novoil'inovskiy 2 Cemetery in Kazakhstan. J. Archaeol. Sci. Rep. *32*, 102420. https://doi.org/10.1016/j.jasrep.2020.102420.

45. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics *29*, 1682–1684. https://doi.org/10.1093/bioinformatics/btt193.

46. Team, R.C. (2013). R: a language and environment for statistical computing. http://www.R-project.org/.

47. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes *9*, 88. https://doi.org/10.1186/s13104-016-1900-2.

48. Schubert, M., Ermini, L., Der Sarkissian, C., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M.D., Fernández, R., Kircher, M., McCue, M., et al. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nat. Protoc. *9*, 1056–1082. https://doi.org/10.1038/nprot.2014.063.

49. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

50. Lepetz, S., Clavel, B., Alioğlu, D., Chauvey, L., Schiavinato, S., Tonasso-Calvière, L., Liu, X., Fages, A., Khan, N., Seguin-Orlando, A., et al. (2021). Historical management of equine resources in France from the Iron age to the modern period. J. Archaeol. Sci. Rep. *40*, 103250. https://doi.org/10.1016/j.jasrep.2021.103250.

51. Clavel, B., Lepetz, S., Chauvey, L., Schiavinato, S., Tonasso-Calvière, L., Liu, X., Fages, A., Khan, N., Seguin-Orlando, A., Der Sarkissian, C., et al. (2022). Sex in the city: uncovering sex-specific management of equine resources from prehistoric times to the Modern Period in France. J. Archaeol. Sci. Rep. *41*, 103341. https://doi.org/10.1016/j.jasrep.2022.103341.

52. Foucras, S., Caillat, P., Goudemez, S., Nuviala, P., Balasse, M., Cabanis, M., and Ferret, C. (2019). Sépultures de chevaux devant Gergovie, archéozoologie des rituels gaulois. https://hal.archives-ouvertes.fr/hal-03218106.

53. Méniel, P., and Jouin, M. (2000). Les inhumations d'animaux de Vertault (Côte d'Or, début de notre ère), 2020/10/31 Edition (Université de Liège). https://doi.org/10.1093/bib/bbaa227.

54. Gamba, C., Hanghøj, K., Gaunitz, C., Alfarhan, A.H., Alquraishi, S.A., Al-Rasheid, K.A.S., Bradley, D.G., and Orlando, L. (2016). Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. Mol. Ecol. Resour. *16*, 459–469. https://doi.org/10.1111/1755-0998.12470.

55. Briggs, A.W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., and Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Res. *38*, e87. https://doi.org/10.1093/nar/gkp1163.

56. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. Philos. Trans. R. Soc. Lond. B Biol. Sci. *370*, 20130624. https://doi.org/10.1098/rstb.2013.0624.

57. Gaunitz, C., Fages, A., Hanghøj, K., Albrechtsen, A., Khan, N., Schubert, M., Seguin-Orlando, A., Owens, I.J., Felkel, S., Bignon-Lau, O., et al. (2018). Ancient genomes revisit the ancestry of domestic and Przewalski's horses. Science *360*, 111–114. https://doi.org/10.1126/science.aao3297.

58. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*. pdb.prot5448. https://doi.org/10.1101/pdb.prot5448.

59. Vershinina, A.O., Heintzman, P.D., Froese, D.G., Zazula, G., Cassatt-Johnstone, M., Dalén, L., Der Sarkissian, C., Dunn, S.G., Ermini, L., Gamba, C., et al. (2021). Ancient horse genomes reveal the timing and extent of dispersals across the Bering Land Bridge. Mol. Ecol. *30*, 6144–6161. https://doi.org/10.1111/mec.15977.

60. Kalbfleisch, T.S., Rice, E.S., DePriest, M.S., Jr., Walenz, B.P., Hestand, M.S., Vermeesch, J.R., O Connell, B.L., Fiddes, I.T., Vershinina, A.O., Saremi, N.F., et al. (2018). Improved reference genome for the domestic horse increases assembly contiguity and composition. Commun. Biol. *1*, 197. https://doi.org/10.1038/s42003-018-0199-z.

61. Felkel, S., Vogl, C., Rigler, D., Dobretsberger, V., Chowdhary, B.P., Distl, O., Fries, R., Jagannathan, V., Janečka, J.E., Leeb, T., et al. (2019). The horse Y chromosome as an informative marker for tracing sire lines. Sci. Rep. *9*, 6095. https://doi.org/10.1038/s41598-019-42640-w.

62. Poullet, M., and Orlando, L. (2020). Assessing DNA sequence alignment methods for characterizing ancient genomes and methylomes. Front. Ecol. Evol. *8*, 105. https://doi.org/10.3389/fevo.2020.00105.

63. Schmidt, M., Maixner, F., Hotz, G., Pap, I., Szikossy, I., Pálfi, G., Zink, A., and Wagner, W. (2021). DNA methylation profiling in mummified human remains from the eighteenth-century. Sci. Rep. *11*, 15493. https://doi.org/10.1038/s41598-021-95021-7.

64. Niiranen, L., Leciej, D., Edlund, H., Bernhardsson, C., Fraser, M., Quinto, F.S., Herzig, K.-H., Jakobsson, M., Walkowiak, J., and Thalmann, O. (2022). Epigenomic modifications in modern and ancient genomes. Genes *13*, 178. https://doi.org/10.3390/genes13020178.

65. Zhang, W., Spector, T.D., Deloukas, P., Bell, J.T., and Engelhardt, B.E. (2015). Predicting genome-wide DNA methylation using methylation marks, genomic position, and DNA regulatory elements. Genome Biol. *16*, 14–20. https://doi.org/10.1186/s13059-015-0581-9.

66. Breiman, L. (2001). Random forests. Mach. Learn. *45*, 5–32. https://doi.org/10.1023/A:1010933404324.

67. Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. *33*, 1–22. https://doi.org/10.18637/jss.v033.i01.

68. Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., and Altman, R.B. (2001). Missing value estimation methods for DNA microarrays. Bioinformatics *17*, 520–525. https://doi.org/10.1093/bioinformatics/17.6.520.

69. Gardner, S.T., Bertucci, E.M., Sutton, R., Horcher, A., Aubrey, D., and Parrott, B.B. (2022). Development of DNA methylation-based epigenetic age predictors in loblolly pine (Pinus taeda). Mol. Ecol. Resour. *23*, 131–144. https://doi.org/10.1101/2022.01.27.477887.

70. Horvath, S., and Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat. Rev. Genet. *19*, 371–384. https://doi.org/10.1038/s41576-018-0004-3.

71. Raznahan, A., Parikshak, N.N., Chandran, V., Blumenthal, J.D., Clasen, L.S., Alexander-Bloch, A.F., Zinn, A.R., Wangsa, D., Wise, J., Murphy, D.G.M., et al. (2018). Sex-chromosome dosage effects on gene expression in humans. Proc. Natl. Acad. Sci. USA *115*, 7398–7403. https://doi.org/10.1073/pnas.1802889115.

72. Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., and Weingessel, A. (2009). E1071: misc functions of the department of statistics (E1071), TU wien. https://CRAN.R-project.org/package=e1071.

73. Uddin, S., Haque, I., Lu, H., Moni, M.A., and Gide, E. (2022). Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. Sci. Rep. *12*, 6256. https://doi.org/10.1038/s41598-022-10358-x.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
osteological remain	this study	GVA1107_Fra_1650
osteological remain	this study	GVA1108_Fra_1650
osteological remain	this study	GVA1109_Fra_1650
osteological remain	this study	GVA1111_Fra_1650
osteological remain	this study	GVA1112_Fra_1650
osteological remain	this study	GVA1113_Fra_1650
osteological remain	this study	GVA1117_Fra_1650
osteological remain	this study	GVA1119_Fra_1650
osteological remain	this study	GVA1122_Fra_1650
osteological remain	this study	GVA3_Fra_110
osteological remain	this study	GVA34_Fra_110
osteological remain	this study	GVA37_Fra_110
osteological remain	this study	GVA602_Fra_m25
osteological remain	this study	GVA603_Fra_m25
osteological remain	this study	GVA607_Fra_m25
osteological remain	this study	GVA609_Fra_m25
osteological remain	this study	GVA610_Fra_m25
osteological remain	this study	GVA630_Fra_m25
osteological remain	this study	GVA639_Fra_2
osteological remain	this study	GVA643_Fra_m25
osteological remain	this study	GVA647_Fra_m25
osteological remain	this study	GVA649_Fra_m25
osteological remain	this study	GVA652_Fra_m25
osteological remain	this study	GVA661_Fra_m25
osteological remain	this study	GVA69_Fra_110
osteological remain	this study	GVA226_Fra_250
osteological remain	this study	GVA373_Fra_1550
osteological remain	this study	GVA758_Fra_1800
osteological remain	this study	GVA764_Fra_1800
osteological remain	this study	GVA768_Fra_1800
osteological remain	this study	GVA827_Fra_150
osteological remain	this study	GVA839_Fra_1675
osteological remain	this study	GVA980_Fra_1100
Chemicals, peptides, and recombinant proteins		
Proteinase K 100 MG	Thermo Fisher Scientific	Cat# 10103533
H2O, Molecular Biology Grade, Fisher BioReagents	Thermo Fisher Scientific	Cat# 10490025
Tween 20 100 ML	Thermo Fisher Scientific	Cat# 10113103
Ethanol, Absolute, Mol Biology Grade	Thermo Fisher Scientific	Cat# 10644795
5M Sodium Chloride 100 ML	Thermo Fisher Scientific	Cat# 10609823
USER Enzyme	New England Biolabs	Cat# M5505L
NEBNext End Repair Module	New England Biolabs	Cat# E6050L

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bst DNA Polymerase	New England Biolabs	Cat# M0275L
NEBNext Quick Ligation Module	New England Biolabs	Cat# E6056L
BSA Molecular Biology Grade	New England Biolabs	Cat# B9000S
N-Lauroylsarcosine solution 30% 500 mL	Dutscher	N-Lauroylsarcosine solution 30% 500 mL
ACCUPRIME PFX DNA POLYMERASE 100 µL	Thermo Fisher Scientific	Cat# 10472482
Agencourt AMPure XP - 60 mL	Beckman Coulter	Cat# A63881
Buffer PE	QIAGEN	Cat# 19065
Buffer PB	QIAGEN	Cat# 19066
Buffer EB	QIAGEN	Cat# 19086
EDTA 0.5M pH 8.0 Fisher	Thermo Fisher Scientific	Cat# 10182903
Tris HCl, 1M, pH 8.0, 100 ML	Thermo Fisher Scientific	Cat# 10336763
dNTP Set 100 mM 100 µL	Thermo Fisher Scientific	Cat# 10336653
Critical commercial assays		
Tapestation screenTape D1000 HS	Agilent	Cat# 5067-5584
MinElute PCR Purification kit	QIAGEN	MinElute PCR Purification kit
Mybaits R V5.02 target capture kit	Daicel Arbor Biosciences	Cat# 502
Deposited data		
ENA	this study	PRJEB56293
Software and algorithms		
DamMet	(Hanghøj et al., 2019)[26]	https://github.com/grenaud/gargammel
mapDamage2	(Jonsson et al., 2013)[45]	https://ginolhac.github.io/mapDamage
R	(Team, 2013)[46]	https://www.R-project.org/
AdapterRemoval2	(Schubert et al., 2016)[47]	https://github.com/MikkelSchubert/adapterremoval
PALEOMIX	(Schubert et al., 2014)[48]	https://github.com/MikkelSchubert/paleomix
Bowtie2	(Langmead et al., 2012)[49]	https://bowtie-bio.sourceforge.net/bowtie2/index.shtml

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources, material and reagents should be addressed and will be fulfilled by the lead contact, Ludovic Orlando ludovic.orlando@univ-tlse3.fr.

Materials availability

This study did not generate new unique reagents.

Data and code availability

The sequence data reported in this paper have been deposited to the European Nucleotide Archive (ENA: PRJEB56293). All other previously published genomic data used in this study are available at the sources referenced in the quantification and statistical analysis section. The code used in this study is listed in the key resources table. Code produced from this study is available via Github (https://github.com/xuefenfei712/MethylationAge).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Archaeological information

The 33 bone remains new to this study were collected from nine archaeological sites in present-day France (Table S1), whose archaeological contexts were fully described in previous work (Lepetz et al.,

2021): Vertault (25 BCE; N=5), Gondole/Enfer (Orcet - La Roche Blanche - L'Enfer, Le Cendre – Gondole – Les piots; 1st century BCE – 1st century CE; N=7), Chartres Roman (Boulevard de la Courtille C277.2; 1st-2nd century CE, ~110 CE; N=4), Amiens (Rue Legrand Daussy; 2nd-3rd century CE, ~150 CE; N=1), Boves (La Vallée de Glisy; 2nd-3rd century CE, Z250 CE; N=1), Chambly (Impasse du Moulin; 11th-13th century CE, ~1100 CE; N=1), Beauvais (Villiers-de-l'Isle Adam, 16 th-17th century CE, ~1550 CE; Abord du Théâtre, 17 th-18th century CE, ~1650 CE; N=10), Saint-Quentin (18 rue des Faucons; 17 th-18th century CE, Z1675 CE; N=1), and Bernolsheim Momenheim (Plateforme Départementale Activité zone Sud, 18th-19th century CE, ~1800 CE; N=3). In downstream analyses, samples showing similar dates and geographical locations were grouped together (i.e. samples from Chartres Roman, Amiens and Boves were labelled "Chartres Roman", and samples from Beauvais, Chambly, Saint-Quentin and Bernolsheim Momenheim were considered as the "Beauvais" group). All bone remains consisted of petrosal bones, excepting the radius bone from the Chambly site. The horse bones collected at the two Iron Age sites of Gondole/Enfer and Vertault were the only ones found within well-established horse sacrificial ritual deposits.[50–53] Bones collected at the urban sites of Beauvais (Abords du Théâtre) and Chartres Roman (Boulevard de la Courtille C277.2) were excavated within constructions aiming at draining humid zones.[50] Other horse bones were found in rural elite (Boves, secular deposit), rural civil (Chambly, food waste and Bernolsheim Momenheim, rendering waste) and urban civil (Amiens, food waste and Saint-Quentin, rendering waste) contexts.[50,51]

All samples were collected under the official agreement signed between the sample providers (SL and BC, Natural History Museum, Paris, France) and the Centre for Anthropobiology and Genomics of Toulouse (CAGT, France), where the analyses were carried out.

Age-at-death of horse individuals previously published[43] and those new to this study were estimated through morphological examination of the remains by SL and BC, following the methodology from,[2] based on patterns of tooth wear and dental eruption.

METHOD DETAILS

Ancient DNA extraction, library construction and sequencing

Wet-laboratory procedures were conducted at CAGT, following protocols outlined by Seguin-Orlando and colleagues,[27] Fages and colleagues[15] and Librado and colleagues.[11] Pre-PCR steps, including sample drilling, DNA extraction and sequencing library construction, were performed in lab facilities strictly dedicated to the analysis of ancient remains, and located in separate buildings from those where modern and post-PCR molecules are manipulated. In brief, around 210-790mg of bone powder was generated using the Mixel Mill MM200 (Retsch) Micro-dismembrator or a mortar; DNA was then extracted following the Y2 procedure recommended by Gamba and colleagues.[54] The DNA extracts were subjected to Uracil-Specific Excision Reagent (USER, NEB), removing uracil residues generated by the post-mortem deamination of non-methylated cytosines.[55] Double-stranded DNA templates were built into Illumina sequencing libraries, with two internal indexes added at the adapter ligation step and one external index incorporated during the PCR enrichment reactions, following a protocol described in.[56,57] After AMPure-bead purification and quantification using both the Qubit (HS assay) and TapeStation (D1000 assay) instruments, 25–50 DNA sequencing libraries were pooled and sequenced at low depth on a MiniSeq instrument at the CAGT (High-Output 80 Paired-End mode) to screen for equine library content. For samples GVA602, GVA607 and GVA661, extra sequencing libraries were produced and further deep sequenced on the Illumina HiSeq4000 instrument (76 Paired-End mode) at the Genoscope (France Génomique), resulting in 502.3, 378.8 and 415.7 million sequences, respectively (Table S1). Extra sequencing data were generated in the same conditions to increase the genome coverage for samples GVA53 and GVA60, previously published in,[15] resulting in 389.0 and 408.7 million sequences and a final average depth-of-coverage of 6.94- and 7.04-fold, respectively (Table S1).

In-solution target enrichment

DNA capture probes were designed at CAGT to target 1,611 genomic CpG sites showing age-dependent profiles in horses,[36] following the procedure described in Haak and colleagues (2015).[38] Those regions overlapped with a total 2,171 CpG sites informing on DNA methylation clock, 34% of which were distributed in introns, 24% in intergenic regions, and 11% in promoters (Figure S2A). Probes were synthesized as single-stranded biotinylated DNA oligonucleotides by Daicel Arbor Biosciences (Ann Arbor, MI, USA) (Table S3A).

Sequencing libraries showing endogenous horse DNA content above 50% were selected for target enrichment. To ensure sufficient amount of DNA templates for in-solution capture, each library was amplified in two consecutive rounds of eight parallel 25 μL PCR reactions. Aliquots of 2uL of the non-amplified library were PCR-amplified for a first round of 8 cycles, following the conditions described in[11]. After a MinElute column purification step (QIAGEN), aliquots of 1 to 3 μL of each of the 25uL eluates were subjected to a second round of 10 PCR cycles using IS5 and IS6 primers.[58] Final products were pooled, purified on a single MinElute column, using elution volumes of 25μL, and quantified using both the Qubit (HS assay) and TapeStation (D1000 assay) instruments. Amplified libraries were sent to Daicel Arbor Biosciences as pools of four libraries in six separate batches. Upon arrival, the pools were quantified via a spectrofluorometric assay to measure total DNA concentration and via a qPCR assay to quantify properly adapted molecules. At least 1 μg of each pool was taken through to capture following the protocol outlined in the myBaits Expert Human Affinites v1 manual (https://arborbiosci.com/wp-content/uploads/2021/03/myBaits_Expert_HumanAffinities_v1.0_Manual.pdf).

After the first hybridization, the capture supernatant was collected and saved. After the second round of capture, the pools were amplified using i5 and i7 indexing primers to re-write the external indexes of the libraries. The amplified captured library pools were visualized on the TapeStation (Agilent) instrument. For each batch, a sequencing pool was prepared from all captures in equimolar ratios. Sequencing was performed on the NovaSeq 6000 platform (S4 lanes, 150 Paired-End mode), producing 9.8–55.6 million sequences per library (Table S1). Demultiplexed, but otherwise raw, FASTQ files were delivered electronically to the CAGT.

QUANTIFICATION AND STATISTICAL ANALYSIS

Whole genome sequencing data collection

We gathered a total of 54 ancient horse genomes characterized at medium-to-high-coverage. These represented 51 genomes that were obtained from previously published data,[11,15,18,57,59] including three newly sequenced for this study and two for which additional sequence data were generated. Detailed information is provided in Table S1.

Read processing, trimming and alignment

Sequencing reads were processed as described by Librado and colleagues.[11] Briefly, fastQ sequencing reads were demultiplexed, collapsed and trimmed using AdapterRemoval2 (version 2.3.0; –barcode-mm-r[12] 1 –minlength 25 –trimns –trimqualities –minadapteroverlap 3 –mm 5).[47] In PALEOMIX version 1.2.13.2,[48] collapsed, truncated and paired reads were aligned against the horse reference genome EquCab3,[60] supplemented with the Y-chromosomal contigs from Felkel et al. 2019.[61] This was achieved using Bowtie2[49] and the parameters from.[62] BAM alignments were realigned locally around indels. Reads showing mapping quality below 25 and PCR duplicates were filtered out. Post-mortem DNA damage levels were measured on 100,000 random reads per library using the mapDamage2 software (version 2.0.8)[45] to verify that the nucleotide mis-incorporation and fragmentation profiles exhibited features expected for USER-treated ancient DNA molecules.

Detection of DNA methylation levels

We assessed DNA methylation levels using DamMet,[26] with genomic windows centered on each of the 31,836 (WGS data) or 2,171 (capture data) CpG sites underlying the HorvathMammalMethylChip40 mammalian methylation array.[36] Calculations were performed from cytosine deamination profiles at CpG and Cp[ACT] sites, accounting for sequencing errors, and considering genomic windows containing 1 and up to 50 CpGs (Figures 2A and 4A; 1_CpG, 2_CpG, 3_CpG, 4_CpG, 5_CpG, 6_CpG, 7_CpG, 8_CpG, 9_CpG, 10_CpG, 15_CpG, 20_CpG, 25_CpG, 30_CpG, 35_CpG, 40_CpG, 45_CpG and 50_CpG). Increasing percentages for the expected average cellular methylation levels were also considered from 70% to 95% (step-size = 5%).

Ancient DNA methylation inference

The normalized methylation data for modern horses based on the HorvathMammalMethylChip40 array[36] were downloaded from Gene Expression Omnibus (Accession Number GSE174767). The ancient genomes used in the present study are sequenced from tooth or bone DNA extracts. As no reference methylation maps for modern horses are available for these specific tissues, we used averaged values over all tissues,

based on the assumption that bone and tooth methylation levels are similar to mean methylation levels across tissues.[63] Spearman correlation levels between untransformed ancient and modern DNA methylation levels for different average cellular methylation fractions (70%–95%), and CpG window sizes (from 1 to 50 CpGs), are shown in Table S2A.

Mathematical transformation of ancient DNA methylation values

The major obstacles to correlating modern and ancient DNA methylation levels pertain to the biases introduced by the different detection methods used (DNA sequencing for the former and DNA methylation array for the latter), but also to the lack of fine-scale resolution of ancient methylation maps.[64] To improve ancient DNA methylation inference, we applied mathematical transformation of DamMet methylation scores based on a two-step procedure. This first step built on a previously-described machine learning approach,[65] which we extended to output DNA methylation percentage values instead of binary 0 and 1 categories. To achieve this, we used the randomforest (RF) package[66] in R4.1[67] and regressed the DNA methylation values inferred by DamMet[26] from ancient DNA sequence data against those measured in modern horses, accounting for the sequencing depth of the ancient data for each CpG site. Separate RF models were constructed for ancient WGS and capture data. We used a 10-fold cross-validation approach, with 500 *ntree* (the total number of trees to grow in the forest) and *mtry* (the number of variables randomly sampled as candidates at each split) set as default, to obtain the best regression parameters. More specifically, the parameter (X) leading to highest regression for a given individual were obtained for the following function:

$$(modF)^X = (ancF)^2 + ancF + Dep + ancF * Dep \quad \text{(Equation 1)}$$

with: *modF* representing the matrix of DNA methylation values averaged for modern males (females), if the ancient individual considered is a male (female), *ancF* the DNA methylation values inferred by DamMet for a given ancient individual, *Dep* the sequencing depth for each site. The interaction effect between ancF and Dep (ancF*Dep in R) was also included in the model.

We built separate RF models for ancient WGS and capture data using the same procedures. For modelling, we excluded age-dependent CpG sites as well as those diagnostic for castration, leading to input datasets of 29,655 sites for WGS and 1,902 sites for capture data. Input data were randomly split into a training set (90% of the samples) and a test set (10%) for validation. The best X parameter identified and Equation 1 were the used to transform DNA methylation values for the 31,836 sites considered for WGS data, or the 2,171 sites considered for DNA capture.

As a further transformation step, we conducted a Simple Linear Regression (SLR),[37] following the formula below, where *Predict(RF)* represents the predicted value obtained from the model optimized following RF and described in 1:

$$\text{modF} = a \times \text{predict(RF)} + b \quad \text{(Equation 2)}$$

Sites with residual errors above one standard error were excluded for later analysis. Furthermore, four individuals showing >10% missingness across the CpG sites considered were disregarded (CGG101397_Rus_1825, CGG10022_Rus_m40610, CGG10023_Rus_m14170, and Batagai_Rus_m3136). Additionally, a total of 2,723 sites that were missing after transformation in more than 10% of the remaining ancient individuals (WGS data) were also disregarded, resulting in a final set of 29,113 CpG sites. For capture data, no individuals were dismissed, but 112 sites were excluded due to missingness. When less than 10% of the sites were missing for a given individual, missing values were imputed using k-nearest neighbor (KNN), with default parameters.[68] This provided the transformed ancient DNA methylation values that were considered in downstream analyses.

Principal component analysis

Principal component analysis (PCA) of DNA methylation values was performed using the R function prcomp,[46] on the merged dataset of samples passing the filtering criteria (333 modern individuals and 84 ancient individuals, Figure S2B). DNA methylation values for modern individuals corresponded to those measurements returned by the HorvathMammalMethylChip40 array,[36] while those of ancient specimens were the values inferred by DamMet,[26] following the transformation procedures described above.

Building DNA methylation clocks

We developed two DNA methylation clocks, one based on WGS data (29,113 CpG sites, after filtering for missingness) and one based on capture data (2,059 CpG sites, after filtering for missingness). They were both built using generalized linear models through the glmnet package in R.[67] Modern DNA methylation data for 239 individuals of known date-of-birth were retrieved from blood and liver tissues only, as these were previously reported to provide highly accurate modern epigenetic clocks.[36] DNA methylation available for other tissues were, thus, disregarded.

We first used DNA methylation data to optimize alpha, the elastic net mixing parameter of the generalized linear model relating DNA methylation and biological age, by investigating values ranging from 0 (Ridge) to 1 (Lasso), using 0.1 increments. Those ancient individuals for which biological ages could be determined based on anatomical features were retained for DNA methylation clock calibration, together with the modern individuals. Transformed DNA methylation values for the 10 horses with WGS data and the 25 horses with capture data were merged with the untransformed DNA methylation values for 239 modern horses, resulting in two respective matrices underlying two different DNA methylation clocks (Tables S2B and S3C). The returned optimized value of 0.9 was used in downstream analyses. Penalty parameters lambda minimizing mean errors were determined automatically by using a 10-fold internal cross-validation (cv.glmnet),[69] following.[36] To assess the accuracy of our DNA methylation age estimators, we used the "leave one out cross-validation" (LOOCV) scheme, applying the glmnet with the best optimal parameters leaving one sample out and predicting the age of that sample. The procedure was then iterated over all samples. All sites showing glmnet coefficients equal to zero were removed for predictions, following previous work.[35]

Both Pearson's correlation coefficients and Mean Absolute Errors (MAE) were used to measure the robustness and accuracy of the DNA methylation clocks reconstructed. The Pearson's correlation R was calculated between biological and methylation age, following.[70] The MAE metric measures the average absolute differences between predictions and true values. Bootstrapping across CpG sites allowed us to evaluate the 95% Confidence Intervals of age predictions (calculated with the Rmisc package; N=100). The age precisions for modern and ancient samples are shown in Figures 3E, 4E, and S1.

Sex detection from DNA methylation data

Due to X-chromosome inactivation in females,[71] males and females can be differentiated on the basis of their levels of DNA methylation along this chromosome. We used DNA methylation data from 333 modern horses of known sex, and 50 ancient WGS data whose molecular sex was determined by the X-to-autosomal coverage ratio, following[41] (Table S1). We first calculated the ratio of DNA methylation values between autosomes and X-chromosome in different categories of modern horses, considering females, castrated males and intact stallions (i.e. non-castrated males) (Figure S3). We then estimated the distribution of DNA methylation values within 20% intervals, which was used for PCA to classify horses according to their labelled category (Figure 5A). We found one modern horse that was labeled as an intact stallion and one modern horse reported as castrated clustering with females (Figure S3), which likely identifies mislabeling issues in the original data repository.

Machine learning classification of castrated and males

We explored the possibility of detecting castration in ancient individuals from the delay in their epigenetic age incurred from castration. Previous work on modern horses reported that castration shows a significant impact on DNA methylation age in individuals older than 11 years old.[36] Thus, to select sites that are differentially methylated in castrated and intact males, we only retained modern horses older than 11 years old, resulting in a total of 22 castrated males and 4 intact stallions, for which blood DNA methylation data were available from HorvathMammalMethylChip40 array.[36]

We then carried out some Epigenome-wide association studies of castration (EWAS) using the R function "standardScreeningBinaryTrait" from the WGCNA R package,[42] categorizing males as either intact stallions (uncastrated) or castrated. CpG sites with Pearson's correlation $|R| \geq 0.5$ (Figure S4), or ≥ 0.55 (Figure 5), were included in the underlying machine learning models. These corresponded to 122 and 59 CpG sites, respectively (Tables S3A and S3B and Figure S3).

We further attempted to classify castrated males and intact stallions based on DNA methylation values at these CpG sites. For this, we explored the performance of three machine learning methods: Random Forest (RF), Support Vector Machine (SVM) and Generalized k-Nearest Neighbors (GKNN) (Figure 3). We divided the original dataset providing the DNA methylation values of 26 modern samples for the 122 and 59 CpG sites considered into a training (80%) and a test (20%) dataset. The RF-based classifier was built using the randomForest R package based on the algorithm of Breiman and Cutler.[66] A first preliminary RF model was built for the association between methylation values and castration status, for all sites in the training set, using *ntree*=200 and default values for *mtry* and *cv.fold*. The optimized value for the *mtry* parameter was determined by exploring *mtry* values comprised between 1 and the number of sites (122 or 59), and selecting the one minimizing the error rate across all sites. We optimized the *ntree* number following a similar procedure, considering *ntree* values between 1 and 200. The optimal values of *mtry*=1 (122 sites), *mtry*=5 (59 sites) and *ntree*=120 and 60 were then implemented in the final RF model used for prediction. The SVM procedure was implemented in the e1071 package in R,[72] using the *svm* function in classification mode (castrated males vs. intact stallions), with a linear function kernel and the following parameters: *cost*=2^(-9) and *cross*=10. For GKNN, we used the *gknn* function in e1071, with the *manhattan* method for distance computation[73] and default parameters. The classifications returned by the three models are presented on an MDS plot and a pie chart reporting individual assignment probabilities (Figure 5 and Table S4).

ADDITIONAL RESOURCES

Our study has not generated or contributed to a new website/forum and it is not part of a clinical trial.