



**HAL**  
open science

## Assessing the impact of USER-treatment on hyRAD capture applied to ancient DNA

Tomasz Suchan, Lorelei Chauvey, Marine Poulet, Laure Tonasso-Calvière, Stéphanie Schiavinato, Pierre Clavel, Benoit Clavel, Sébastien Lepetz, Andaine Seguin-Orlando, Ludovic Orlando

► **To cite this version:**

Tomasz Suchan, Lorelei Chauvey, Marine Poulet, Laure Tonasso-Calvière, Stéphanie Schiavinato, et al.. Assessing the impact of USER-treatment on hyRAD capture applied to ancient DNA. *Molecular Ecology Resources*, 2022, 22 (6), pp.2262-2274. 10.1111/1755-0998.13619 . hal-03827861

**HAL Id: hal-03827861**

**<https://hal.science/hal-03827861>**

Submitted on 20 Jun 2023




**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# Assessing the impact of USER-treatment on hyRAD capture applied to ancient DNA

Tomasz Suchan<sup>1,2</sup> | Lorelei Chauvey<sup>1</sup> | Marine Poulet<sup>1</sup>  | Laure Tonasso-Calvière<sup>1</sup> |  
Stéphanie Schiavinato<sup>1</sup> | Pierre Clavel<sup>1</sup> | Benoit Clavel<sup>3</sup> | Sébastien Lepetz<sup>3</sup> |  
Andaine Seguin-Orlando<sup>1</sup>  | Ludovic Orlando<sup>1</sup> 

<sup>1</sup>Centre d'Anthropobiologie et de Génomique de Toulouse (CAGT), Faculté de Santé, Université Paul Sabatier, Toulouse, France

<sup>2</sup>W. Szafer Institute of Botany, Polish Academy of Sciences, Kraków, Poland

<sup>3</sup>Archéozoologie, Archéobotanique: sociétés, pratiques et environnements (AASPE), Muséum National d'Histoire Naturelle, CNRS, Paris, France

## Correspondence

Ludovic Orlando, Centre d'Anthropobiologie et de Génomique de Toulouse (CAGT), Université Paul Sabatier, Faculté de Santé, 37 allées Jules Guesde, Bâtiment A, 31000 Toulouse, France.  
Email: [ludovic.orlando@univ-tlse3.fr](mailto:ludovic.orlando@univ-tlse3.fr)

## Funding information

University Paul Sabatier IDEX Chaire d'Excellence (OURASI); CNRS Programme de Recherche Conjoint (PRC); CNRS International Research Project (IRP AMADEUS); European Union's Horizon 2020 Research and Innovation Programme, Grant/Award Number: 681605; Marie Skłodowska-Curie, Grant/Award Number: 797449; European Research Council (ERC)

**Handling Editor:** Sebastien Calvignac-Spencer

## Abstract

Ancient DNA preservation in subfossil specimens provides a unique opportunity to retrieve genetic information from the past. As ancient DNA extracts are generally dominated by molecules originating from environmental microbes, capture techniques are often used to economically retrieve orthologous sequence data at the population scale. Post-mortem DNA damage, especially the deamination of cytosine residues into uracils, also considerably inflates sequence error rates unless ancient DNA extracts are treated with the USER enzymatic mix prior to library construction. While both approaches have recently gained popularity in ancient DNA research, the impact of USER-treatment on capture efficacy still remains untested. In this study, we applied hyRAD capture to eight ancient equine subfossil specimens from France (1st–17th century CE), including horses, donkeys and their first-generation mule hybrids. We found that USER-treatment could reduce capture efficacy and introduce significant experimental bias. It differentially affected the size distribution of on-target templates following capture with two distinct hyRAD probe sets in a manner that was not driven by differences in probe sizes and DNA methylation levels. Finally, we recovered unbalanced proportions of donkey-specific and horse-specific alleles in mule capture sequence data, due to the combined effects of USER-treatment, probe sets and reference bias. Our work demonstrates that while USER-treatment can improve the quality of ancient DNA sequence data, it can also significantly affect hyRAD capture outcomes, introducing bias in the sequence data that is difficult to predict based on simple molecular probe features. Such technical batch effects may prove easier to model and correct for using capture with synthetic probes of controlled sizes and diversity content.

## KEYWORDS

ancient DNA, DNA capture, DNA damage, DNA methylation, equids, reference bias

Tomasz Suchan and Lorelei Chauvey equally contributed.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

The survival of ancient DNA molecules within subfossil specimens provides a unique opportunity to gather genetic information from past, and even extinct, organisms within the last 1.6 million year time range (van der Valk et al., 2021). During the last decade, the experimental costs underlying the sequencing of ancient genomes have dramatically declined as high-throughput DNA sequencing instruments have become increasingly performant (Goodwin et al., 2016). The ultra-fragmented and degraded nature of ancient DNA, which is more difficult to manipulate than fresh DNA contaminants, represents a major technical limitation to ancient genome characterization (Dabney, Knapp, et al., 2013). In fact, the vast majority of subfossil DNA extracts are dominated by large proportions of exogenous DNA molecules, mostly derived from various environmental microbial sources (Green et al., 2009). Therefore, shotgun sequencing does not generally provide a cost-effective solution, except for material such as petrosal bones (Gamba et al., 2014), ossicles (Sirak et al., 2020) and tooth cementum (Damgaard et al., 2015), which can show better DNA preservation and limited microbial DNA content.

Over the last years, an increasing number of molecular techniques have enriched the ancient DNA toolkit with the aim to improve DNA extraction (Boessenkool et al., 2016; Dabney et al., 2013; Gamba et al., 2016; Korlević & Meyer, 2019), as well as DNA library conversion (Gansauge et al., 2017, 2020; Kapp et al., 2021; Meyer et al., 2012) and immortalization (Dabney & Meyer, 2012). Other approaches have been developed to focus sequencing efforts on endogenous fragments of interest (Carpenter et al., 2013; Enk et al., 2014; Mathieson et al., 2015), while reducing the proportion of nucleotide mis-incorporations resulting from post-mortem DNA damage (Briggs et al., 2010; Rohland et al., 2015). The latter can be achieved by treating ancient DNA extracts with the USER mix, which combines Uracil DNA glycosylase (UDG) and Endonuclease VIII (EndoVIII) enzymes. UDG excises uracil residues that are formed following post-mortem deamination of cytosines (Briggs et al., 2007). The EndoVIII ultimately cleaves the phosphodiester bonds at the resulting abasic sites, which introduces a one-nucleotide gap along the corresponding DNA strand. USER-treated ancient DNA fragments are, thus, on average shorter and largely deprived of uracil residues, which translates into less inflated sequencing error rates, as uracils represent thymine analogs that introduce an excess of C to T substitutions during sequencing (Briggs et al., 2007).

Focusing sequencing efforts on target regions of interest can be achieved using various in-solution capture techniques, including both fully commercial and home-made solutions (Carpenter et al., 2013; Maricic et al., 2010; Suchan et al., 2021). The myBaits Expert Human Affinities kit from Arbor Bioscience provides an example of commercial capture reagents that target approximately 2 million loci spread across the human genome. hyRAD capture provides an example of procedures reducing experimental costs related to probe synthesis and allowing users to prepare RNA probes covering a defined fraction of the genome (Suchan et al., 2016). Biotinylated hyRAD RNA probes are prepared from genomic DNA extracts of

extant organisms, following double enzymatic restriction, library conversion and in vitro transcription. hyRAD RNA probes are then used in standard capture reactions to retrieve those DNA fragments present in ancient DNA libraries and showing high sequence complementarity (Suchan et al., 2021). This methodology is scalable and has been successfully used to improve the characterization of the DNA variation that was present in 4.8%–7.2% of the horse genome and shared across multiple individuals dating back to beyond the radiocarbon range (Suchan et al., 2021).

Capture techniques can achieve on-target enrichment folds of several orders of magnitude (Ávila-Arcos et al., 2011; Furtwängler et al., 2020) but are limited by a number of factors, including the probe molecular features (Cruz-Dávalos et al., 2017), the DNA adapter length (Rohland et al., 2015), as well as the size of the ancient DNA templates (Suchan et al., 2021). While USER treatment reduces the impact of post-mortem damage on the sequencing data (Rohland et al., 2015), it does not repair those cytosines that were deaminated after death but cleaves instead the DNA strand at their location. By reducing the length of ancient DNA templates and therefore potentially the probe-template heterodimer stability, this treatment may, thus, impede the efficacy of target-enrichment techniques. Conversely, the removal of post-mortem damage limits the probe-template edit distance, which may facilitate probe-template annealing and, ultimately, improve capture success. How ancient DNA capture is affected by DNA methylation, which protects deaminated cytosines from USER treatment (Hanhøj et al., 2016; Pedersen et al., 2014), and thus can increase both read edit distances and sequence length, remains unclear. Surprisingly, the impact of USER treatment on the capture efficacy has not been tested, despite both techniques gaining popularity in ancient DNA research, and possibly far-ranging effects.

In this study, we selected eight ancient equine bone specimens from present-day France dating to the 1st–17th centuries CE (Common Era). These specimens were selected for their high equine DNA content (Clavel et al., 2021, 2022; Lepetz et al., 2021), which allowed us to characterize the aforementioned methodological bias with minimal sequencing efforts. We constructed DNA libraries in the absence of or following USER treatment of ancient DNA extracts. These ancient DNA libraries were then subjected to shotgun sequencing as well as parallel hyRAD capture reactions, in which RNA probes were prepared either from horse or donkey fresh DNA extracts. We found that USER treatment impacts the capture efficacy. It led to an over-representation of shorter or longer ancient DNA templates, depending on the probe set used for capture. While methylated DNA fragments that are damaged post-mortem are refractory to USER-treatment and, thus, strand cleavage, ancient DNA templates tended to contain reduced DNA methylation levels post-capture. This suggests that the lower edit distance associated with DNA hypomethylation and lower deamination rates, more than the size increment gained from USER-protection at methylated sites, offers an advantage during capture. Therefore, following USER-treatment, capture sequencing data can reduce the power to access to hypermethylated DNA fragments, which yet preserve

post-mortem DNA damage signatures that are important for data authentication (Orlando et al., 2021; Rohland et al., 2015).

The ancient specimens analysed included horses, donkeys as well as their first-generation mule hybrids. Interestingly, our data revealed unbalanced proportions of horse- and donkey-specific alleles in mules, depending on the reference genome used for sequence alignment, the probe set used and whether the DNA was USER-treated or not. Combined, these results caution that the probability to detect highly divergent alleles in an ancient individual, thus, the ancestry contribution from potential divergent ghost-lineages, is not only sensitive to the reference genome used for sequence alignment and USER-treatment but also to probe molecular features that may be difficult to control for with current hyRAD procedures.

## 2 | MATERIALS AND METHODS

### 2.1 | Data generation

In a previous study, we used shallow DNA sequencing to screen a total of 874 equine subfossil specimens for DNA preservation levels (Lepetz et al., 2021). The sequence data collected were processed within the Zonkey pipeline (Schubert et al., 2017) to identify horses, donkeys and mules showing high levels of endogenous DNA (Table S1). We selected two horses, three donkeys and three mules with sufficient endogenous DNA to ensure that sufficient sequence data could be obtained from minimal sequencing efforts.

Probes production followed protocol by Suchan et al. (2021), with slight modifications, starting from 200 to 500 ng of horse or donkey DNA, digested for 3 h at 37°C with 3 U of MseI and 6 U of PstI-HF (New England Biolabs – NEB) in 15 µl reactions. Enzymes were inactivated for 20 min at 65°C before the reaction was supplemented with ATP (final concentration = 6.6 mM), T4 DNA ligase (NEB, 400 U) and the P1 and P2 adapters (0.5 µM for each) and incubated for 3 h at 16°C. The reactions were purified with AMPure beads with a bead-to-DNA volume ratio of 1.5:1 and eluted in 30 µl of 10 mM Tris. The samples were then size-selected for 190–390 bp size range using the Blue Pippin instrument and 2% agarose cassettes with external marker (Sage Science). Fragments containing P2 adapters were next separated using biotinylated MyOne C1 Dynabeads (Invitrogen) as in Suchan et al. (2021) and eluted in 30 µl of 10 mM Tris. The resulting fragments were amplified in three parallel PCR reactions, using KAPA HiFi HotStart ReadyMix (Roche), 0.5 µM of IS4 primer and indexing primer (Meyer & Kircher, 2010), and a 5 µl aliquot of the Dynabeads-purified solution obtained at the previous step. The PCR consisted of denaturation for 3 min at 95°C, followed by 15 cycles (denaturation for 20 s at 98°C, annealing for 15 s at 60°C, and; elongation for 30 s at 72°C), and a final elongation step for 5 min at 72°C. Amplification triplicates were pooled, purified with AMPure beads using a bead-to-buffer volume ratio of 1:1 and eluted in 60 µl of 10 mM Tris. The resulting products were digested for 3 h at 37°C, using 50 µl of the purified product and 7.5 U of MseI (NEB) in 75 µl reactions. The RNA probes were prepared

using HiScribe T7 High Yield RNA Synthesis Kit (NEB), according to the manufacturer's instructions for RNA synthesis with biotinylated-UTP, using 1 µl of the template and following overnight incubation, and the DNA template was removed using 4 U of TURBO DNase in 30 min reaction at 37°C. Finally, RNA probes were purified using RNEasy Mini kit (Qiagen) with the standard protocol, excepting that 665 µl ethanol were added to the RNA+RTL mix. The probes were then diluted to around 100 ng/µl. Blocking RNAs were prepared as described in Suchan et al. (2021).

Ancient DNA extraction, USER-treatment and library construction procedures followed the work from Fages et al. (2019). For each ancient DNA extract, one aliquot was prepared and subjected to USER-treatment before preparing a triple-indexed DNA library while a second aliquot was directly used for library construction. A volume of 14.9 µl of raw DNA extract was used for library construction (final volume = 25 µl), together with adapters including unique 7 bp indices located at the ends undergoing ligation with ancient DNA templates (Rohland et al., 2015). Two aliquots containing 3 µl of DNA libraries were amplified in two parallel 25 µl reactions, containing 0.4 µl AccuPrime Pfx polymerase (ThermoFisher Scientific), 1 mg/ml BSA, 200 mM of the inPE1 (5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT) and external index (5'-CAA GCA GAA GAC GGC ATA CGA GAT NNN NNN GTG ACT GGA GTT CAG ACG TGT), where NNN NNN represents the external 6 bp library index. The amplification conditions consisted of a first denaturation step for 5 min at 95°C, and were followed by 12 cycles (denaturation for 15 s at 95°C, annealing for 30 s at 60°C, and; elongation for 30 s at 68°C) and a final elongation step for 5 min at 68°C. PCR amplifications were purified using AMPure XP beads (Beckman Coulter), with a bead-to-buffer ratio of 1.4:1, and eluted in 25 µl EB (Qiagen). Two aliquots of each library amplification (7.5 µl each) were then reamplified in parallel reactions using the same conditions as above, except that the reaction volume was 50 µl and that 1 µM of the primers IS5 (5'-AAT GAT ACG GCG ACC ACC GA) and IS6 (5'-CAA GCA GAA GAC GGC ATA CGA) were used (Meyer & Kircher, 2010). The required number of PCR cycles were determined using qPCR and ranged from 7 to 10 cycles. The two parallel reactions were purified on MinElute columns (Qiagen) and eluted in 11 µl EB before being mixed to a single tube, thus, providing 22 µl of amplified library. Library concentrations were estimated using Qubit (ThermoFisher Scientific) between 180 and 216 ng/µl.

Capture conditions followed those described by Suchan and colleagues (Suchan et al., 2021), and included a volume of 7 µl of each amplified library and 550 ng hyRAD RNA probes. The hybridization reaction was carried out for 40 h at 55°C in a 30 µl volume. Dynabeads M-280 Streptavidin (ThermoFisher Scientific) were prepared following first three successive washing steps with 200 µl of TEN buffer, followed by a final elution in 70 µl of TEN buffer. Hybridization reactions were then purified on an Opentrons OT2 liquid-handling instrument using the automated procedure described by Suchan et al. (2021), with the slight improvements provided in the Data S1. The final elution step was carried out in 30 µl of 10mM Tris. Two 8 µl aliquots of this first hybridization round were amplified in parallel 50 µl

reactions using the KAPA HiFi HotStart ReadyMix (Kapa Biosystems) and the IS5 and IS6 PCR primers (Meyer & Kircher, 2010, 500 nM each). Amplification conditions consisted of a first denaturation step for 3 min at 95°C, followed by 9–12 cycles (denaturation for 20 s at 98°C, annealing for 15 s at 60°C, and; elongation for 30 s at 72°C) and a final elongation step for 1 min at 65°C. Parallel amplification reactions were pooled and purified on MinElute columns (Qiagen) and eluted in 10 µl EB. A total of 7 µl of such purified amplifications were used in a second hybridization reaction, followed by Dynabeads purification using the same conditions as above, except that the hybridization lasted for 18 h. The capture product was amplified in PCR as above, for three cycles, and purified using AMPure XP beads (Beckman Coulter), with a bead-to-buffer ratio of 1:1, and eluted in 30 µl EB tween 0.05%. Library profiles and concentrations were estimated using the TapeStation 4200 instrument (Agilent) and Qubit HS dsDNA assay (Invitrogen), before pooling at equimolar concentration for sequencing.

Sequencing was carried out on the Illumina MiniSeq instrument available at CAGT using the paired-end sequence mode (2 × 81 bp, except for the horse probe library, which was sequenced using 2 × 156 bp reactions).

## 2.2 | Data analysis

Probe DNA sequences were demultiplexed using CutAdapt v2.10 (Martin, 2011; `--discard-untrimmed -g "TAATACGACTCACTA TAGGGCGG;max_error_rate=0.15;min_overlap=20" -G "^TAA;max_error_rate=0;min_overlap=3"`). Raw sequence fastq files generated from ancient DNA libraries were demultiplexed on the basis of the two internal indices using AdapterRemoval2 (Schubert et al., 2016; `--barcode-mm-r1 1 --barcode-mm-r2 1 --minlength 25 --trimms --trimqualities --minadapteroverlap 3 --mm 5`). This pipeline also collapses paired-end reads showing significant sequence overlap into single end reads, creating so-called “collapsed” reads, or “collapsed truncated” reads, if ends show low base quality scores and require further trimming. Both collapsed, collapsed truncated and those paired-end reads showing no significant overlap were then aligned against reference genomes using Bowtie2 v2.3.5.1 (Langmead & Salzberg, 2012) and the Paleomix v1.2.13.2 pipeline (Schubert et al., 2014). Read alignments were carried out following the optimized parameters from Poullet and Orlando (2020) and repeated against the horse reference genome (EquCab3, Kalbfleisch et al., 2018), supplemented by the Y-chromosome contigs from Felkel and colleagues (Felkel et al., 2019), and the donkey reference genome (Wang et al., 2020). Read alignments were also carried out against the horse (Accession no. = NC\_001640) and the donkey (JADWZWO10000001.1) mitochondrial genomes. PCR duplicates were removed using Paleomix and alignments showing mapping qualities strictly inferior to 25 were disregarded. Similar mapping parameters were applied to the demultiplexed set of probe paired-end sequences and on-target regions were identified applying Bedtools v2.27.1 `bamtoBed` to the alignments obtained from the probe sequence libraries. The extent of daisy chaining (Cruz-Dávalos

et al., 2017) was investigated by following the drop of sequence coverage within the 250 bp flanking on-target regions, within intervals of 50 bp (Figures S1 and S2). As coverage was substantially inflated in the regions immediately flanking probes, we defined off-target regions as located at a conservative distance of at least 250 bp away from on-target regions. On-target and off-target read alignments were identified using `samtools view v1.11`, considering a minimum of 1 bp sequence overlap.

In silico digestion of the horse and the donkey nuclear genomes were carried out using the script `fragmatic.pl` (Chafin et al., 2018), with `-r "CTGCA^G T^TAA"`. This provided the expected fraction of each genome represented in each size class (Figure S3). These size distributions were then compared to those measured empirically from the sequence data generated from the probe libraries, using the `bed coordinates` of on-target regions. We obtained the genomic coordinates of the in silico predicted fragments using `PatMaN (v1.2.2, -e 0 -g 0; Prüfer et al., 2008)`, as these are not returned by `fragmatic.pl`. The overlap of the in-silico predicted fragments with those regions empirically identified as on- and off-target was assessed using `Bedtools v2.27.1 intersect` (see above). The nonredundant probe library sequence content was estimated using the `Preseq v2_0 lc_extrap` command (Daley & Smith, 2014) and the limited sequence data generated from the probe DNA libraries. Size, %GC and %CpG distributions were estimated from the fraction of collapsed reads to ensure that ancient DNA templates were characterized across their full sequence length. The size of each DNA template was obtained from the `samtools view` command, while %GC and %CpG distributions of the underlying genomic regions were obtained by applying `seqtk v1.3-r117-dirty comp` (Li, 2013), with default parameters, to the reference genome of interest, considering genomic windows of 100 bp. Post-mortem DNA damage signatures, especially C to T misincorporations within and outside CpG dinucleotides, were obtained from `PMDtools v0.60` (Skoglund et al., 2014). Calculations and statistical tests were carried out using standard functions in R (R Core Team, 2014), and plots were generated using the R `ggplot2` package (Wickham, 2009). F-methylation scores were calculated using `DamMet v1` (Hanghøj et al., 2019), with standard parameters.

Horse-specific and donkey-specific alleles were identified using published genome sequence data for a total of 27 horses and 27 donkeys, respectively, which encompassed a whole range of breeds in both species (for horses: Akhal-Teke (Jagannathan et al., 2019), Duellmeyer (Schrimpf et al., 2016), Franches Montagnes (Jagannathan et al., 2019), Haflinger (Jagannathan et al., 2019), Hanoverian (Schrimpf et al., 2016), Holsteiner (Jagannathan et al., 2019), Icelandic (Andersson et al., 2012), Jeju (Lee et al., 2018), Lipizzan (Wallner et al., 2017), Marwari (Jun et al., 2014), Mixed UK Warmblood (Jagannathan et al., 2019), Mongolian (Do et al., 2014), Noriker (Jagannathan et al., 2019), Painted (Jagannathan et al., 2019), Quarter (Jagannathan et al., 2019), Reit (Jagannathan et al., 2019), Shetland (Jagannathan et al., 2019), Sorraia (Metzger et al., 2015), Standardbred (Cosgrove et al., 2020), Swiss Warmblood (Jagannathan et al., 2019), Thoroughbred (Cosgrove et al., 2020), Trakener (Jagannathan et al., 2019), Egyptian Arabian (Cosgrove



et al., 2020), Welsh (Jagannathan et al., 2019), Westfale (Jagannathan et al., 2019), and Yakutian (Librado et al., 2015); for donkeys: Au-1, Ch-by1, Ch-by3, Ch-gl1, Ch-ht1, Ch-jm1, Ch-kl2, Ch-qy2, Ch-tlf1, Ch-xj1, Ch-yn3, Eg-1, Eg-3, Et-4, Et-8, Ir-7, Ke-16, Ky-6, Ni-6, Sp-10, Sp-12, Sp-15, Sp-17, Sp-2, Sp-5, Sp-7, Ti-4; Wang et al., 2020). Sequence data were aligned following the procedure described above, and resulting BAM alignment files were processed in ANGSD (-minQ 30, -minMapQ 25 -baq 0 -rmTriallelic 1e-4 -SNP\_pval 1e-6 -C 50; Korneliusson et al., 2014) to identify minor and major alleles at sites that were polymorphic and covered in at least 40 individuals. ANGSD output files were next restricted to those sites showing minimal genotype quality scores of 0.99 in all individuals covered to identify positions where allelic frequency differences between horses and donkeys were at least 95%. This provided a provisional list of genome positions in which different variants neared fixation in donkeys and horses.

### 3 | RESULTS

#### 3.1 | Experimental design and hyRAD probes

A previous study by Lepetz et al. (2021) allowed us to identify a total of eight ancient equine subfossils from France (1st–17th centuries CE) showing excellent DNA preservation, including two horses, three mules and three donkeys (Table S1). Their high endogenous DNA content (37.9%–78.1%) guaranteed that a large number of equine DNA templates could be identified, and thus, the effect of capture following USER treatment assessed, even from limited sequencing efforts (289,599–1,997,334 reads; Table S1). We used fresh DNA extracts from modern horses and donkeys to prepare two hyRAD probe libraries following the methodology from Suchan and colleagues (Suchan et al., 2021). We produced a total of 6.41 and 12.71 million paired-end sequences to characterize the horse and donkey probe library content. The sequence data generated from the horse probe library covered approximately 118.99 Mb (~4.73%) of the horse reference genome. The donkey sequence probes obtained covered 110.50 Mb (~4.54%) of the donkey reference genome. In the following, these sequences were used to identify those ancient DNA read alignments located on-target.

It is important to notice that further sequencing of DNA probe libraries would have extended the on-target range (Figure S2, S3). In silico digestion with *fragmatic.pl* (Chafin et al., 2018) indeed indicates that probe regions could encompass 377.57 Mb (15.01%) and 360.89 Mb (14.83%) of the horse and donkey genomes, respectively. Similar estimates were obtained using PatMaN (Prüfer et al., 2008), assuming perfect match (473.64 Mb, 18.83% and 364.59 Mb, 14.99%, respectively). Random sampling across the genome (N = 100) regions of similar size than those predicted by PatMaN to be on-target indicated that only 43.74–44.12 Mb and 56.60–57.14 Mb would intercept by chance those regions currently defined as off-target for the horse and the donkey, respectively. The observed effective overlap between in silico digests and off-target

regions is considerably larger (300.54 Mb and 154.71, respectively), which demonstrates the presence of a statistically significant fraction of not-yet-sequenced probes within those regions currently considered off-target (empirical *p*-values <1%). It follows that the calculations presented in the next sections, including enrichment folds and the differences detected between on- and off-target regions, should be considered as conservative.

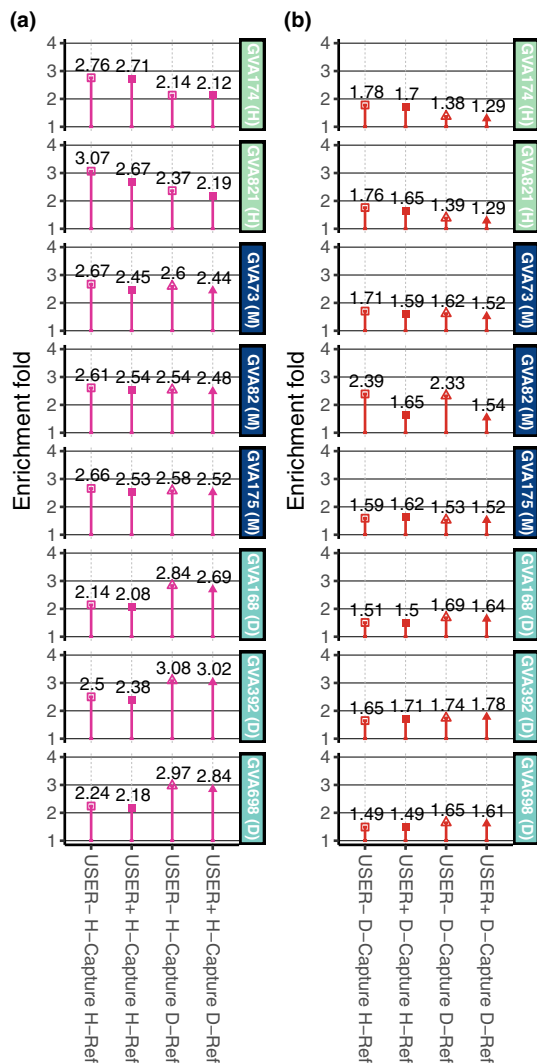
Interestingly, Preseq (Li, 2013) predicted that an average 25–30 million unique templates were present in the horse probe library versus 15–25 million for the donkey library (Figure S3). This is in line with a previous study using the same combination of enzymes for hyRAD probe preparation (HspI-MseI, Suchan et al., 2021), and suggests 50–100 million read pairs as a conservative sequencing effort to characterize the full mappable probe library content.

#### 3.2 | The effect of USER-treatment on hyRAD capture

The ratio of the total number of unique read alignments overlapping on-target regions following hyRAD capture or shotgun sequencing showed enrichment-folds of 2.08–3.08 for the horse probes (Figure 1a, Figure S4a). Enrichment-folds were more limited when using the donkey probes (1.29–2.39; Figure 1b, Figure S4b). While the enrichment-folds obtained are conservative, they may seem limited. It is, however, important to consider that only samples characterized by high endogenous content were considered in our experiments. The probability that shotgun data overlap a probe genomic region by chance is considerably increased in such specimens relative to standard subfossil DNA extracts, which are dominated by environmental microbial DNA. Suchan et al. (2021) reported that up to 146.5-fold enrichment could be achieved on ancient DNA libraries containing only marginal endogenous DNA content (0.34%). Therefore, higher enrichment-folds are expected when working on material with a more limited DNA preservation.

Interestingly, higher enrichment-folds were obtained when capturing DNA libraries constructed on raw DNA extracts with horse probes than when capturing DNA libraries prepared from USER-treated DNA extracts (Wilcoxon signed rank test,  $p = 0.014$ – $0.036$ ; Figure 1a). The same was true for seven (five) out of the eight samples analysed here, when donkey probes were used for capture and when the sequence data were mapped against the donkey (horse) genome (Wilcoxon signed rank test,  $p$ -values = 0.039 and 0.151; Figure 1b).

Our hyRAD probe sequences represented a non-random fraction of the genome and were characterized by both a higher %GC and %CpG dinucleotide content (Figure S5; Kolmogorov-Smirnov test,  $p$ -values <  $2.2 \cdot 10^{-16}$ ). The %GC and %CpG dinucleotide content of the ancient DNA sequence data were also significantly increased following capture, relative to shotgun sequencing (Figure 2b,c, Figure S6a,b; Kolmogorov-Smirnov test,  $p < 2.2 \cdot 10^{-16}$ ). This effect was generally more pronounced in on-target regions than in off-target regions, in line with the latter including only a limited fraction



**FIGURE 1** On-target enrichment-folds. (a) Capture with horse hyRAD probes (pink). (b) Capture with donkey hyRAD probes (red). Ancient DNA libraries from two horses (light green, [H]), three mules (deep blue, [M]) and three donkeys (light blue, [D]) specimens were subjected to capture with hyRAD horse (H-capture) or donkey (D-capture) probes. Ancient DNA libraries were prepared from extracts that were treated with the USER enzymatic mix (USER+, filled symbols) or from raw extracts (USER-, open symbols). Reads were aligned against the horse reference genome (H-Ref; Kalbfleisch et al., 2018), supplemented with the Y-chromosome contigs from Felkel et al. (2019), and filtered for PCR duplicates and minimal mapping quality of 25. Reads were also aligned the donkey reference genome and subjected to similar quality filters (D-Ref; Wang et al., 2020)

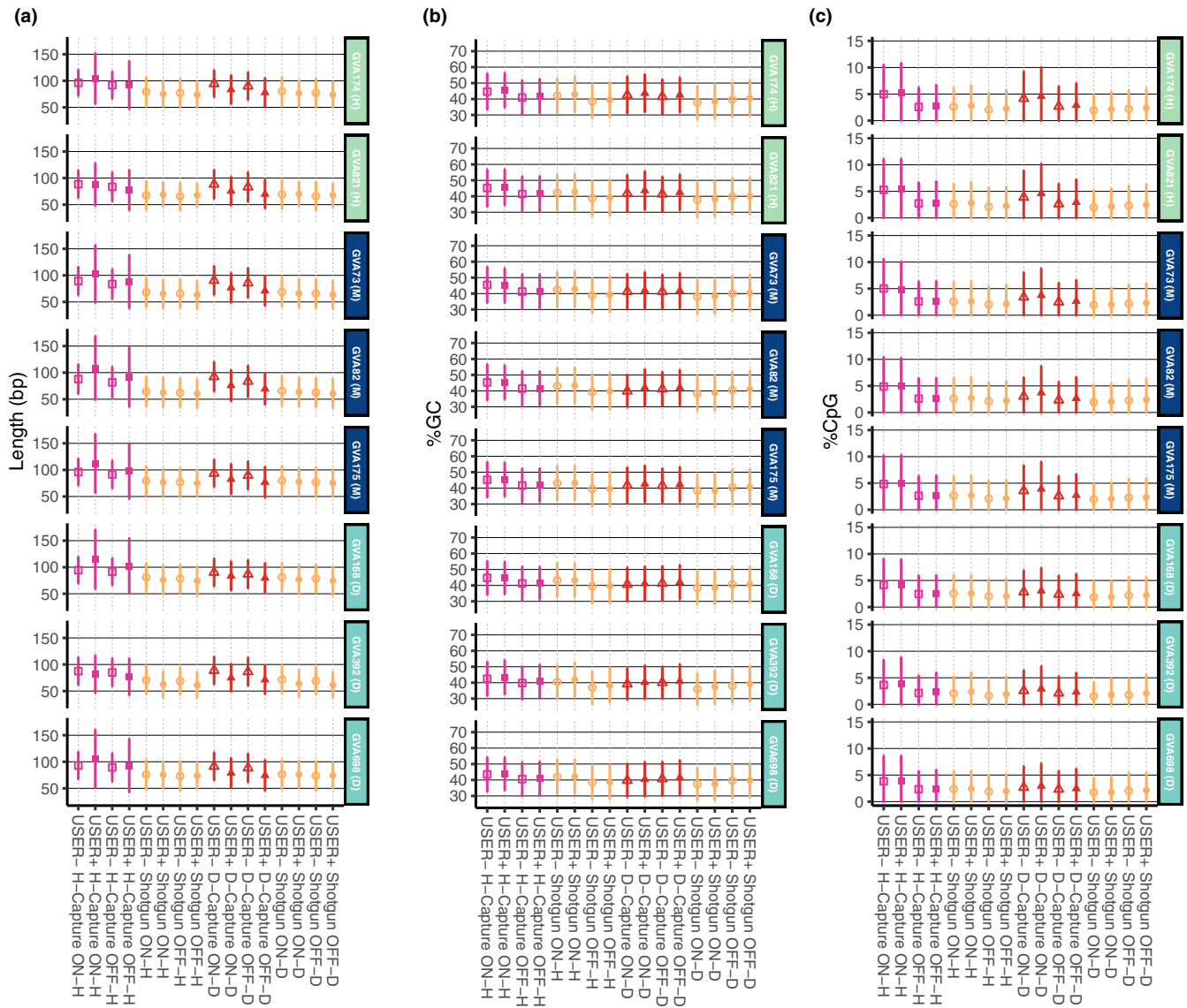
of yet uncharacterized probes. Additionally, in the absence of USER-treatment, those ancient DNA templates sequenced were also significantly longer post-capture (Figure 2a; Kolmogorov-Smirnov test,  $p < 2.2 \cdot 10^{-16}$ ), especially in on-target regions, in line with previous reports. That the sequence data (1) are more often located in on-target regions following capture and (2) comprise a larger fraction of longer fold templates while (3) mirror base compositional features characteristics of the probes demonstrates that the hyRAD capture procedure was applied successfully.

We further assessed the efficacy of the USER-treatment carried out by calculating C to T mis-incorporation rates at the first read position, where post-mortem DNA damage is maximal, and within the following 24 read positions (Figures S7 and S8). We conditioned these analyses to non-CpG dinucleotides as sites mostly unaffected by DNA methylation (Bird, 2002) (Figures S7 and S8). We found that C to T mis-incorporation rates were higher in the absence of USER-treatment than following USER-treatment in all experimental conditions and read positions considered (Wilcoxon signed rank test,  $p = 1.554 \cdot 10^{-4} - 0.014$ ). This indicated that the USER treatment was applied successfully. Combined, these results suggested that the sequence data generated in this study could be used to assess the effect of USER-treatment on capture efficacy.

Interestingly, when capture was carried out with horse probes, we observed an upward shift in the size distribution of the equine DNA material sequenced (Figure 2a). This was observed whether or not PCR duplicates were removed (data not shown). This size increment of on-target sequence data was generally more pronounced following USER-treatment than in the absence of USER treatment (Figure 2a, Figure S6a; Kolmogorov-Smirnov test,  $p < 2.2 \cdot 10^{-16}$ ). That DNA libraries constructed on USER-treated extracts showed larger size distributions post-capture than those constructed on raw extracts was unexpected since USER-treatment cleaves off unmethylated cytosine residues that were deaminated into uracils post-mortem, hence, reduce the average template size (Briggs et al., 2010). The size shift effect was, however, reversed (Kolmogorov-Smirnov test,  $p < 2.2 \cdot 10^{-16}$ ), when donkey probes were used for capture (Figure 2a, Figure S6a). This suggests that USER-treatment can differentially impact on the size distribution of the captured templates, depending on probe-specific features. We, thus, next explored probe size and DNA methylation levels as potential driving features.

We noticed that the size profile of the donkey and horse hyRAD probes was markedly different, with the former significantly shorter (median = 101–103 bp, depending on the reference genome used for alignments) than the latter (median = 185–186 bp), despite similar experimental production procedures were applied (Figure S9; Kolmogorov-Smirnov test,  $p < 2.2 \cdot 10^{-16}$ ). In particular, while the donkey probes contained at least 10% of molecules shorter than 76 nucleotides, this size category was almost fully absent from the horse probe set. We hypothesized that the shorter molecules present in the donkey probe set have allowed annealing to those more fragmented ancient DNA templates post-USER-treatment, resulting in an overall size reduction of the on-target sequence data. However, those equine templates overlapping shorter donkey probes (40–60 bp) were found to be on average longer than those overlapping longer probes (Figures S10 and S11). Similarly, those equine templates overlapping longer horse probes were not disproportionately longer (Figures S10 and S11). This suggests molecular features other than size driving the differential size shift observed when donkey and horse probes following USER-treatment.

We next tested whether donkey and horse probes differed in their DNA methylation levels, as DNA methylation protects



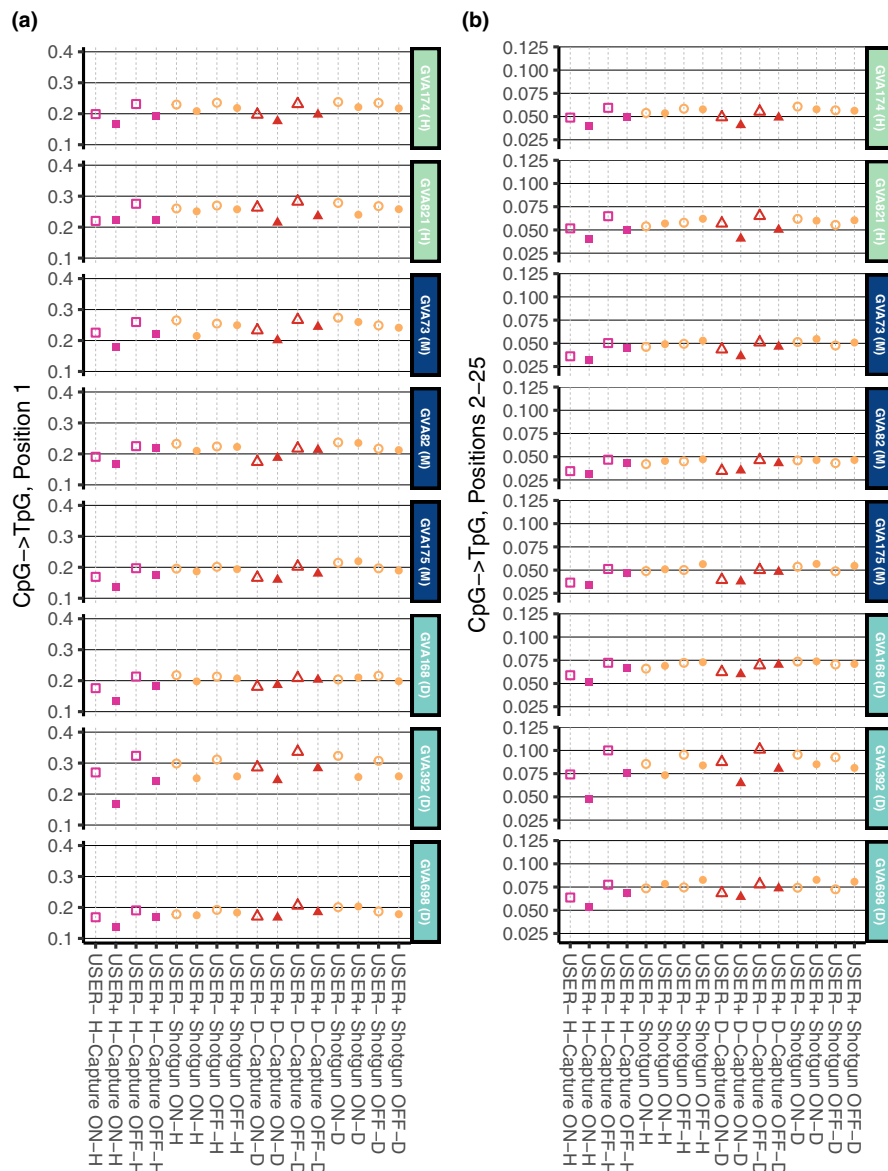
**FIGURE 2** Size and base composition of ancient DNA templates following shotgun sequencing and hyRAD capture. (a) Size distribution of endogenous DNA templates. (b) %GC content of the genomic regions covered by endogenous DNA templates. (c) Proportions of CpG dinucleotides present in the genomic regions covered by endogenous DNA templates. Read pairs that could not be collapsed into full-length templates were not used for mapping and only unique, high-quality alignments were considered. The horse reference genome (Kalbfleisch et al., 2018), supplemented with the Y-chromosome contigs from Felkel et al. (2019), was used for read alignment. Ancient DNA libraries were prepared from extracts that were treated with the USER enzymatic mix (USER+, filled symbols) or from raw extracts (USER-, open symbols). Ancient DNA libraries from two horses (light green, [H]), three mules (deep blue, [M]) and three donkeys (light blue, [D]) specimens were shotgun sequenced (Shotgun, yellow) or subjected to capture with hyRAD horse (H-capture, pink) or donkey (D-capture, red) probes

deaminated CpG dinucleotides from USER treatment and can, thus, result in differential capture outcomes. For example, if the genomic regions underlying hyRAD probes showed higher DNA methylation levels in horses than in donkeys, they would be less prone to USER-cleavage and would, thus, remain longer on average. To test this, we first calculated CpG to TpG conversion rates as a proxy for DNA methylation levels in USER-treated sequence data (Hanghøj et al., 2016; Pedersen et al., 2014). We contrasted the first read position, where post-mortem cytosine deamination is maximal (Briggs et al., 2007) but where molecular breathing can limit the efficacy of USER-treatment (Hanghøj et al., 2019; Rohland et al., 2015), and the

following 24 positions in order to gain resolution. These analyses revealed reduced on-target CpG to TpG mis-incorporation rates post-capture both in the presence and the absence of USER-treatment, and regardless of the probe set considered (Wilcoxon signed rank test,  $p = 3.052.10^{-5}$ ). This was equally true for the first read position as well as the following 24 nucleotides (i.e., positions 2–25, Figure 3a,b, Figure S12). This indicated that both probe sets overlap genomic regions showing more limited DNA methylation levels (Hanghøj et al., 2016). We further confirmed this finding using the F-scores of on-target and off-target regions returned by DamMet, as another DNA methylation proxy (Figure S13; Hanghøj et al.,



**FIGURE 3** CpGTPG conversion rates. (a) First read position. (b) Read positions 2–25. Ancient DNA libraries from two horses (light green, [H]), three mules (deep blue, [M]) and three donkeys (light blue, [D]) specimens were shotgun sequenced (Shotgun, yellow) or subjected to capture with hyRAD horse (H-capture, pink) or donkey (D-capture, red) probes. Reads were aligned against the horse reference genome (H-Ref; Kalbfleisch et al., 2018), supplemented with the Y-chromosome contigs from Felkel et al. (2019), and filtered for PCR duplicates and minimal mapping quality of 25. The proportions of CpGTPG conversions report the relative fraction of read alignments in which a CpG dinucleotide is found in the horse reference genome but a TpG dinucleotide is found in the sequence data. Calculations were conditioned on those read alignments located on-target (ON-H, when horse hyRAD probes were used for capture, and ON-D, when donkey probes were used) or off-target (OFF-H and OFF-D, respectively). Ancient DNA libraries were prepared from extracts that were treated with the USER enzymatic mix (USER+, filled symbols) or from raw extracts (USER-, open symbols)

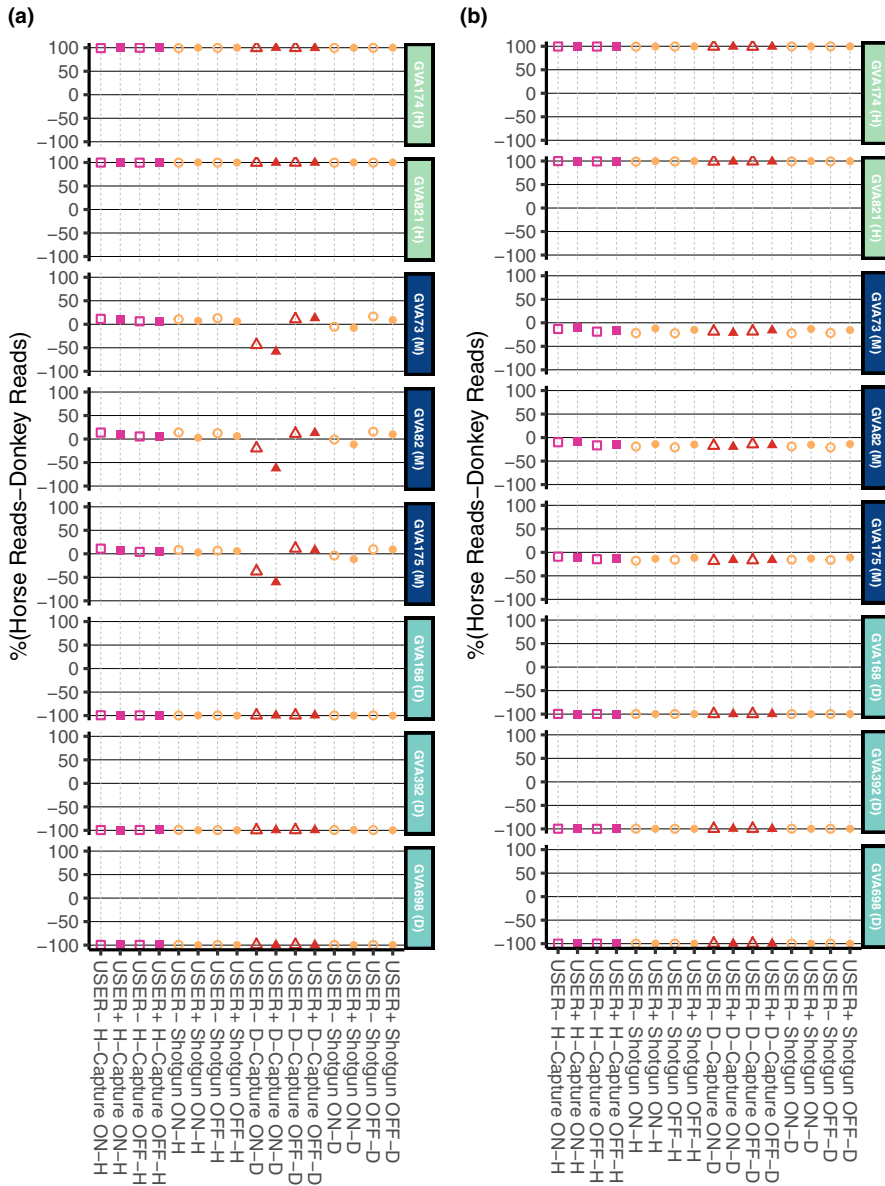


2019). More specifically, on-target regions were found to be associated with lower DNA methylation values than off-target regions as the fraction of CpG dinucleotides characterized by at least 50% DNA methylation levels was consistently reduced at such regions. This was true for both horse and donkey probes, and regardless of whether capture or shotgun DNA sequencing were performed (Wilcoxon signed rank test,  $p = 1.397.10^{-9} - 5.122.10^{-8}$ ). Therefore, the probe regions were characterized by lower DNA methylation than the remaining genomic fraction in both probe sets, which rules out DNA hypermethylation driving the increment in size observed in post-capture with horse hyRAD probes.

### 3.3 | The effect of hyRAD capture on reference bias

We next leveraged our experimental design to assess whether horse-specific alleles and donkey-specific alleles have had equal

chances to be detected following capture with horse or donkey hyRAD probes, respectively. Species-specific alleles for horses and donkeys were identified as those alleles nearing fixation in a worldwide panel representing 27 genomes for both species. Regardless of the reference genome considered, we found that the sequence data generated from the two horse specimens almost exclusively showed horse-specific alleles (Figure 4a,b). The reverse was true for the sequence data generated from the three donkey specimens, which supports the validity of our SNP assignment panel as well as the previous taxonomic identification of these specimens (Lepetz et al., 2021). Interestingly, the shotgun sequence data from the three mule specimens showed almost balanced proportions of horse- and donkey-specific alleles. However, mule shotgun sequence alignments against the horse reference genome generally slightly over-represented horse-specific alleles (Figure 4a) while those against the donkey reference genome were slightly enriched in donkey-specific alleles (Figure 4b). This is characteristic of limited but consistent reference bias in our analyses.



**FIGURE 4** Proportions of horse-specific and donkey-specific alleles identified following shotgun and hyRAD capture. (a) Alignments against the horse reference genome (Kalbfleisch et al., 2018), supplemented with the Y-contigs from Felkel and colleagues (Felkel et al., 2019). (b) Alignments against the donkey reference genome (Wang et al., 2020). Ancient DNA libraries from two horses (light green, [H]), three mules (deep blue, [M]) and three donkeys (light blue, [D]) specimens were shotgun sequenced (Shotgun, yellow) or subjected to capture with hyRAD horse (H-capture, pink) or donkey (D-capture, red) probes. Calculations indicate the difference between the number of reads carrying horse-specific and donkey-specific alleles, respectively. Only transversion alleles were considered. Positive (negative) fractions indicate an overrepresentation of horse-specific (donkey-specific) alleles in the sequence data. Calculations were conditioned on those read alignments against the reference genome that were located on-target (ON-H when horse hyRAD probes were used for capture, and ON-D when donkey probes were used) or off-target (OFF-H and OFF-D, respectively). Ancient DNA libraries were prepared from extracts that were treated with the USER enzymatic mix (USER+, filled symbols) or from raw extracts (USER-, open symbols)

Importantly, this reference bias was generally more limited following USER-treatment, and was generally not impacted by hyRAD capture since allele representation closely followed that observed with shotgun data. However, the mule sequence alignments following capture with donkey hyRAD probes and alignment against the horse reference genome stand as an exception. Here, the reference bias was over-compensated, especially after USER-treatment, resulting in an overrepresentation of donkey-specific alleles (Figure 4a). This was observed whether or not transition SNPs were filtered (Figure 4a, Figure S14a), which rules out DNA damage as a potential driver. This indicates that the combined effects of the probe set, USER-treatment and the reference genome used for alignment can magnify the extent of the reference bias, with potentially significant impact on downstream sequence data analyses investigating population structure, if overlooked.

## 4 | DISCUSSION

In this study, we extracted ancient DNA from eight subfossils of horses, donkeys and mules and constructed double-stranded DNA libraries using both raw DNA extracts and DNA extracts that were subjected to USER-treatment. Amplified DNA libraries were then shotgun sequenced or captured using the hyRAD technology with probes that were prepared from fresh horse or donkey DNA extracts. Our experimental design allowed us to assess, for the first time, whether USER-treatment impacted the capture performance and provided equal chances to identify both parental alleles in hybrid individuals.

We found that the sequence data generated following capture mapped more often to the probe regions than when performing shotgun sequencing. Additionally, on-target sequence data recapitulated the molecular features of the probes,

including increased %GC and %CpG dinucleotide content relative to the remaining fraction of the genome, and decreased DNA methylation levels. This ensured that the hyRAD procedures implemented here succeeded in facilitating the sequence characterization of a pre-selected genome fraction. Interestingly, the size distribution of the ancient DNA fragments sequenced over their full size was increased in all capture conditions. However, USER-treatment further increased the recovery of long ancient DNA fragments when using the horse probe set for capture. This contrasted with the results obtained using donkey probes, which led to equine material encompassing a shorter size range. This could not be explained by differences in DNA methylation levels of the underlying genomic regions, nor the shorter size of the donkey probes used in our experiments. Therefore, other probe features, possibly including species-dependent diversity content, are responsible for the effects reported. In any case, our findings highlight probe preparation as a critical step in the hyRAD procedure. Strikingly, despite the same methodology was used for preparing both probe sets, we yet recovered probe spanning different size ranges. Improving the reproducibility of this step, and identifying quality checks for diversity content, including allele representation, will be critical before the hyRAD capture can be applied across large sample panels without risking to introduce uncontrolled but important technical batch effects. No such difficulties are expected with capture protocols using synthetic oligonucleotides of controlled sizes (e.g., 52-mers in Slon et al., 2017, 60-mers in Cruz-Dávalos et al., 2017 and 120-mers in Ramos-Madrigal et al., 2019).

While mules carry even numbers of donkey and horse autosomal genetic material, horse-specific alleles were slightly over-represented when the horse reference genome was used for sequence alignments. Conversely, donkey-specific alleles were slightly over-represented when aligning the mule sequence data against the donkey reference genome. This is typical of a reference bias (Günther & Nettelblad, 2019), favouring the parental species used as reference for sequence alignment. We expect such effects to be commensurate with the sequence divergence when species others than the horse and the donkey are investigated. Importantly, USER-treatment slightly reduced this bias when using horse probes. This was true regardless of the reference genome considered for mapping. However, the donkey probes disproportionately favored the detection of donkey alleles when aligning mule sequences against the horse genome but not the donkey genome. This effect was magnified following USER-treatment. Therefore, while mixing horse and donkey probes may have seemed the most optimal strategy for limiting the extent of reference bias, our results suggest complex capture outcomes depending on the capture probe set and reference genomes considered. We, thus, recommend merging probe panels prepared from divergent lineages only after confirming their respective impact on reference bias. The analytical framework implemented here, in which the sequencing data were aligned against multiple reference genomes from closely

related species, provides an example of how to do so. More generally, our experiments measured different fractions of the most divergent alleles (i.e., donkey alleles in mule sequence data mapped against the horse reference genome), depending on whether USER-treatment was used or not. This implies that combining capture data obtained from extracts that were or were not subjected to USER treatment in single analyses can come with a risk when assessing the genetic ancestry contribution of divergent, ghost lineages in a given individual.

Post-mortem DNA damage signatures at CpG sites also indicated that the on-target ancient DNA data retrieved post-USER treatment carried fewer deaminated cytosines following capture than shotgun sequencing. Therefore, those ancient DNA templates sequenced post-capture were characterized by lower-than-average DNA methylation levels. This indicates that, in the experimental conditions tested, DNA methylation protection against USER-cleavage at deaminated CpG sites offered no particular advantage for capture, despite preserving longer templates.

Finally, the data retrieved following USER-treatment were characterized by different cytosine deamination rates following shotgun sequencing and hyRAD capture, despite being collected from the same individual (Figure 3, Figures S7, S8, S12). This confirms that the capture procedures do not provide a random subset of the pool of ancient DNA library templates covering the regions of interest but can instead overrepresent those showing molecular features increasing probe-template affinities, including those least deaminated and with lower methylation rates. This has important ramifications for data authentication, especially following mild-USER treatment (Rohland et al., 2015), which relies on the partial preservation of particular post-mortem DNA damage signatures.

## ACKNOWLEDGEMENTS

We thank all members of the AGES research group at CAGT for fruitful discussions. This project has received funding from: the University Paul Sabatier IDEX Chaire d'Excellence (OURASI); the CNRS Programme de Recherche Conjoint (PRC); the CNRS International Research Project (IRP AMADEUS); the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement no 797449; the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant agreement 681605).

## AUTHOR CONTRIBUTIONS

Tomasz Suchan and Ludovic Orlando designed and conceived the study. Tomasz Suchan, Lorelei Chauvey, Laure Tonasso-Calvière, Stéphanie Schiavinato, Andaine Seguin-Orlando performed wet-laboratory work. Benoit Clavel, Sébastien Lepetz and Ludovic Orlando provided samples, material and reagents. Ludovic Orlando analysed the data. Tomasz Suchan and Ludovic Orlando interpreted the data. Ludovic Orlando wrote the article with input from all coauthors.

## OPEN RESEARCH BADGES



This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. Raw DNA sequences have been deposited to the European Nucleotide Archive and are available for download under the PRJEB50952 project.

## DATA AVAILABILITY STATEMENT

Raw DNA sequences have been deposited to the European Nucleotide Archive and are available for download under the PRJEB50952 project.

## ORCID

Marine Poulet <https://orcid.org/0000-0003-2514-2024>

Andaine Seguin-Orlando <https://orcid.org/0000-0002-8265-3229>

Ludovic Orlando <https://orcid.org/0000-0003-3936-1850>

## REFERENCES

- Andersson, L. S., Larhammar, M., Memic, F., Wootz, H., Schwochow, D., Rubin, C.-J., Patra, K., Arnason, T., Wellbring, L., Hjälm, G., Imsland, F., Petersen, J. L., McCue, M. E., Mickelson, J. R., Cothran, G., Ahituv, N., Roepstorff, L., Mikko, S., Vallstedt, A., ... Kullander, K. (2012). Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. *Nature*, *488*(7413), 642–646. <https://doi.org/10.1038/nature11399>
- Ávila-Arcos, M. C., Cappellini, E., Romero-Navarro, J. A., Wales, N., Moreno-Mayar, J. V., Rasmussen, M., Fordyce, S. L., Montiel, R., Vielle-Calzada, J., Willerslev, E., & Gilbert, M. T. (2011). Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Scientific Reports*, *1*(1). <https://doi.org/10.1038/srep00074>
- Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes & Development*, *16*(1), 6–21. <https://doi.org/10.1101/gad.947102>
- Boessenkool, S., Hanghøj, K., Nistelberger, H. M., Der Sarkissian, C., Gondek, A. T., Orlando, L., Barrett, J. H., & Star, B. (2016). Combining bleach and mild predigestion improves ancient DNA recovery from bones. *Molecular Ecology Resources*, *17*(4), 742–751. <https://doi.org/10.1111/1755-0998.12623>
- Briggs, A. W., Stenzel, U., Johnson, P. L., Green, R. E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M., & Pääbo, S. (2007). Patterns of damage in genomic DNA sequences from a neandertal. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(37), 14616–14621. <https://doi.org/10.1073/pnas.0704665104>
- Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., & Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Research*, *38*(6), e87. <https://doi.org/10.1093/nar/gkp1163>
- Carpenter, M. L., Buenostro, J. D., Valdiosera, C., Schroeder, H., Allentoft, M. E., Sikora, M., Rasmussen, M., Gravel, S., Guillén, S., Nekhrizov, G., Leshtakov, K., Dimitrova, D., Theodossiev, N., Pettener, D., Luiselli, D., Sandoval, K., Moreno-Estrada, A., Li, Y., Wang, J., ... Bustamante, C. D. (2013). Pulling out the 1%: Whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *The American Journal of Human Genetics*, *93*(5), 852–864. <https://doi.org/10.1016/j.ajhg.2013.10.002>
- Chafin, T. K., Douglas, M. R., & Douglas, M. E. (2018). MrBait: Universal identification and design of targeted-enrichment capture probes. *Bioinformatics*, *34*(24), 4293–4296. <https://doi.org/10.1093/bioinformatics/bty548>
- Clavel, B., Lepetz, S., Chauvey, L., Schiavinato, S., Tonasso-Calvière, L., Liu, X., Fages, A., Khan, N., Seguin-Orlando, A., Der Sarkissian, C., Clavel, P., Estrada, O., Alioğlu, D., Gaunitz, C., Aury, J.-M., Barme, M., Bodu, P., Olive, M., Bignon-Lau, O., ... Orlando, L. (2022). Sex in the city: Uncovering sex-specific management of equine resources from prehistoric times to the modern period in France. *Journal of Archaeological Science: Reports*, *41*, 103341. <https://doi.org/10.1016/j.jasrep.2022.103341>
- Clavel, P., Dumoncel, J., Der Sarkissian, C., Seguin-Orlando, A., Calvière-Tonasso, L., Schiavinato, S., Chauvey, L., Perdereau, A., Aury, J., Wincker, P., Onar, V., Clavel, B., Lepetz, S., Braga, J., & Orlando, L. (2021). Assessing the predictive taxonomic power of the BONY labyrinth 3D shape in horses, donkeys and their F1-hybrids. *Journal of Archaeological Science*, *131*, 105383. <https://doi.org/10.1016/j.jas.2021.105383>
- Cosgrove, E. J., Sadeghi, R., Schlamp, F., Holl, H. M., Moradi-Shahrbabak, M., Miraei-Ashtiani, S. R., Abdalla, S., Shykind, B., Troedsson, M., Stefaniuk-Szmukier, M., Prabhu, A., Bucca, S., Bugno-Poniewierska, M., Wallner, B., Malek, J., Miller, D. C., Clark, A. G., Antczak, D. F., & Brooks, S. A. (2020). Genome diversity and the origin of the Arabian horse. *Scientific Reports*, *10*(1). <https://doi.org/10.1038/s41598-020-66232-1>
- Cruz-Dávalos, D. I., Llamas, B., Gaunitz, C., Fages, A., Gamba, C., Soubrier, J., Librado, P., Seguin-Orlando, A., Pruvost, M., Alfarhan, A. H., Alquraishi, S. A., Al-Rasheid, K. A., Scheu, A., Beneke, N., Ludwig, A., Cooper, A., Willerslev, E., & Orlando, L. (2017). Experimental conditions improving in-solution target enrichment for ancient DNA. *Molecular Ecology Resources*, *17*(3), 508–522. <https://doi.org/10.1111/1755-0998.12595>
- Dabney, J., Knapp, M., Glocke, I., Gansauge, M., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J., & Meyer, M. (2013). Complete mitochondrial genome sequence of a middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(39), 15758–15763. <https://doi.org/10.1073/pnas.1314445110>
- Dabney, J., & Meyer, M. (2012). Length and GC-biases during sequencing library amplification: A comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *BioTechniques*, *52*(2), 87–94. <https://doi.org/10.2144/000113809>
- Dabney, J., Meyer, M., & Paabo, S. (2013). Ancient DNA damage. *Cold Spring Harbor Perspectives in Biology*, *5*(7), a012567. <https://doi.org/10.1101/cshperspect.a012567>
- Daley, T., & Smith, A. D. (2014). Modeling genome coverage in single-cell sequencing. *Bioinformatics*, *30*(22), 3159–3165. <https://doi.org/10.1093/bioinformatics/btu540>
- Damgaard, P. D., Margaryan, A., Schroeder, H., Orlando, L., Willerslev, E., & Allentoft, M. E. (2015). Improving access to endogenous DNA in ancient bones and teeth. *Scientific Reports*, *5*, 11184. <https://doi.org/10.1101/014985>
- Do, K., Kong, H., Lee, J., Lee, H., Cho, B., Kim, H., Ahn, K., & Park, K. (2014). Genomic characterization of the Przewalski's horse inhabiting mongolian steppe by whole genome re-sequencing. *Livestock Science*, *167*, 86–91. <https://doi.org/10.1016/j.livsci.2014.06.020>
- Enk, J. M., Devault, A. M., Kuch, M., Murgha, Y. E., Rouillard, J., & Poinar, H. N. (2014). Ancient whole genome enrichment using baits built from modern DNA. *Molecular Biology and Evolution*, *31*(5), 1292–1294. <https://doi.org/10.1093/molbev/msu074>
- 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., Alfarhan, A. H., Allentoft, M., Alquraishi, S., Anthony, D., Baimukhanov, N., Barrett, J. H., Bayarsaikhan, J., Benecke, N., Bernáldez-Sánchez, E., ... Orlando, L. (2019). Tracking five millennia of horse management with extensive ancient genome time series. *Cell*, 177(6), 1419–1435.e31. <https://doi.org/10.1016/j.cell.2019.03.049>
- Felkel, S., Vogl, C., Rigler, D., Dobretsberger, V., Chowdhary, B. P., Distl, O., Fries, R., Jagannathan, V., Janečka, J. E., Leeb, T., Lindgren, G., McCue, M., Metzger, J., Neuditschko, M., Rattei, T., Raudsepp, T., Rieder, S., Rubin, C.-J., Schaefer, R., ... Wallner, B. (2019). The horse Y chromosome as an informative marker for tracing sire lines. *Scientific Reports*, 9(1), 6095. <https://doi.org/10.1038/s41598-019-42640-w>
- Furtwängler, A., Neukamm, J., Böhme, L., Reiter, E., Vollstedt, M., Arora, N., Singh, P., Cole, S. T., Knauf, S., Calvignac-Spencer, S., Krause-Kyora, B., Krause, J., Schuenemann, V. J., & Herbig, A. (2020). Comparison of target enrichment strategies for ancient pathogen DNA. *BioTechniques*, 69(6), 455–459. <https://doi.org/10.1101/2020.07.09.195065>
- Gamba, C., Hanghøj, K., Gaunitz, C., Alfarhan, A. H., Alquraishi, S. A., Al-Rasheid, K. A., Bradley, D. G., & Orlando, L. (2016). Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Molecular Ecology Resources*, 16(2), 459–469. <https://doi.org/10.1111/1755-0998.12470>
- Gamba, C., Jones, E. R., Teasdale, M. D., McLaughlin, R. L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kóvári, I., Pap, I., Anders, A., Whittle, A., Dani, J., Raczy, P., Higham, T. F., Hofreiter, M., Bradley, D. G., & Pinhasi, R. (2014). Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications*, 5(1). <https://doi.org/10.1038/ncomms6257>
- Gansauge, M., Aximu-Petri, A., Nagel, S., & Meyer, M. (2020). Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nature Protocols*, 15(8), 2279–2300. <https://doi.org/10.1038/s41596-020-0338-0>
- Gansauge, M., Gerber, T., Glocke, I., Korlević, P., Lippik, L., Nagel, S., Riehl, L. M., Schmidt, A., & Meyer, M. (2017). Single-stranded DNA library preparation from highly degraded DNA using T4DNA ligase. *Nucleic Acids Research*, gkx033. <https://doi.org/10.1093/nar/gkx033>
- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333–351. <https://doi.org/10.1038/nrg.2016.49>
- Green, R. E., Briggs, A. W., Krause, J., Prüfer, K., Burbano, H. A., Siebauer, M., Lachmann, M., & Pääbo, S. (2009). The Neandertal genome and ancient DNA authenticity. *The EMBO Journal*, 28(17), 2494–2502. <https://doi.org/10.1038/emboj.2009.222>
- Günther, T., & Nettelblad, C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLOS Genetics*, 15(7), e1008302. <https://doi.org/10.1371/journal.pgen.1008302>
- Hanghøj, K., Renaud, G., Albrechtsen, A., & Orlando, L. (2019). DamMet: Ancient methylome mapping accounting for errors, true variants, and post-mortem DNA damage. *GigaScience*, 8(4). <https://doi.org/10.1093/gigascience/giz025>
- Hanghøj, K., Seguin-Orlando, A., Schubert, M., Madsen, T., Pedersen, J. S., Willerslev, E., & Orlando, L. (2016). Fast, accurate and automatic ancient nucleosome and methylation maps with epiPALEOMIX. *Molecular Biology and Evolution*, 33(12), 3284–3298. <https://doi.org/10.1093/molbev/msw184>
- Jagannathan, V., Gerber, V., Rieder, S., Tetens, J., Thaller, G., Drögemüller, C., & Leeb, T. (2019). Comprehensive characterization of horse genome variation by whole-genome sequencing of 88 horses. *Animal Genetics*, 50(1), 74–77. <https://doi.org/10.1111/age.12753>
- Jun, J., Cho, Y. S., Hu, H., Kim, H., Jho, S., Gadhvi, P., Park, K. M., Lim, J., Paek, W. K., Han, K., Manica, A., Edwards, J. S., & Bhak, J. (2014). Whole genome sequence and analysis of the Marwari horse breed and its genetic origin. *BMC Genomics*, 15(S9). <https://doi.org/10.1186/1471-2164-15-s9-s4>
- Kalbfleisch, T. S., Rice, E. S., DePriest, M. S., Walenz, B. P., Hestand, M. S., Vermeesch, J. R., O'Connell, B. L., Fiddes, I. T., Vershina, A. O., Saremi, N. F., Petersen, J. L., Finno, C. J., Bellone, R. R., McCue, M. E., Brooks, S. A., Bailey, E., Orlando, L., Green, R. E., Miller, D. C., ... MacLeod, J. N. (2018). Improved reference genome for the domestic horse increases assembly contiguity and composition. *Communications Biology*, 1(1), 1–8. <https://doi.org/10.1038/s42003-018-0199-z>
- Kapp, J. D., Green, R. E., & Shapiro, B. (2021). A fast and efficient single-stranded genomic library preparation method optimized for ancient DNA. *Journal of Heredity*, 112(3), 241–249. <https://doi.org/10.1093/jhered/esab012>
- Korlević, P., & Meyer, M. (2019). Pretreatment: Removing DNA contamination from ancient bones and teeth using sodium hypochlorite and phosphate. *Methods in Molecular Biology*, 15–19. [https://doi.org/10.1007/978-1-4939-9176-1\\_2](https://doi.org/10.1007/978-1-4939-9176-1_2)
- Korneliusson, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, 15(1), 356. <https://doi.org/10.1186/s12859-014-0356-4>
- Langmead, B., & Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Lee, W., Park, K., Taye, M., Lee, C., Kim, H., Lee, H., & Shin, D. (2018). Analysis of cross-population differentiation between thoroughbred and Jeju horses. *Asian-Australasian Journal of Animal Sciences*, 31(8), 1110–1118. <https://doi.org/10.5713/ajas.17.0460>
- Lepetz, S., Clavel, B., Alioğlu, D., Chauvey, L., Schiavinato, S., Tonasso-Calvière, L., Liu, X., Fages, A., Khan, N., Seguin-Orlando, A., Der Sarkissian, C., Clavel, P., Estrada, O., Gaunitz, C., Aury, J.-M., Barthe, M., Boulbes, N., Bourgeois, A., Decanter, F., ... Orlando, L. (2021). Historical management of equine resources in France from the Iron Age to the modern period. *Journal of Archaeological Science: Reports*, 40, 103250. <https://doi.org/10.1016/j.jasrep.2021.103250>
- Li, H. (2013). *Seqtk: A fast and lightweight tool for processing FASTA or FASTQ sequences*. <https://github.com/lh3/seqtk>
- Librado, P., Der Sarkissian, C., Ermini, L., Schubert, M., Jónsson, H., Albrechtsen, A., Fumagalli, M., Yang, M. A., Gamba, C., Seguin-Orlando, A., Mortensen, C. D., Petersen, B., Hoover, C. A., Lorente-Galdos, B., Nedoluzhko, A., Boulygina, E., Tsygankova, S., Neuditschko, M., & Jagannathan, V. ... Orlando, L. (2015). Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proceedings of the National Academy of Sciences of the United States of America*, 112(50). <https://doi.org/10.1073/pnas.1513696112>
- Maricic, T., Whitten, M., & Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One*, 5(11), e14004. <https://doi.org/10.1371/journal.pone.0014004>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, 17(1), 10. <https://doi.org/10.14806/embnet.17.1.200>
- Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., Sirak, K., Gamba, C., Jones, E. R., Llamas, B., Dryomov, S., Pickrell, J., Arsuaga, J. L., de Castro, J. M. B., Carbonell, E., ... Reich, D. (2015). Genome-wide patterns of selection in 230 ancient eurasians. *Nature*, 528(7583), 499–503. <https://doi.org/10.1038/nature16152>
- Metzger, J., Karwath, M., Tonda, R., Beltran, S., Águeda, L., Gut, M., Gut, I. G., & Distl, O. (2015). Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genomics*, 16(1), 1–4. <https://doi.org/10.1186/s12864-015-1977-3>
- Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring*



- Harbor Protocols, 2010(6). [pdb.prot5448](https://doi.org/10.1101/pdb.prot5448). <https://doi.org/10.1101/pdb.prot5448>
- Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J. G., Jay, F., Prüfer, K., de Filippo, C., Sudmant, P. H., Alkan, C., Fu, Q., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R. E., Bryc, K., ... Pääbo, S. (2012). A high-coverage genome sequence from an archaic Denisovan individual. *Science*, 338(6104), 222–226. <https://doi.org/10.1126/science.1224344>
- Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P. W., Ávila-Arcos, M. C., Fu, Q., Krause, J., Willenslev, E., Stone, A. C., & Warinner, C. (2021). Ancient DNA analysis. *Nature Review Methods Primers*, 1, 15. <https://doi.org/10.1038/s43586-021-00016-3>
- 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., Andersson, R., Jones, P. A., Hoover, C. A., Tikhonov, A., Prokhortchouk, E., Rubin, E. M., Sandelin, A., Gilbert, M. T. P., Krogh, A., ... Orlando, L. (2014). Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Research*, 24(3), 454–466. <https://doi.org/10.1101/gr.163592.113>
- Poulet, M., & Orlando, L. (2020). Assessing DNA sequence alignment methods for characterizing ancient genomes and methylomes. *Frontiers in Ecology and Evolution*, 8, 105. <https://doi.org/10.3389/fevo.2020.00105>
- Prüfer, K., Stenzel, U., Dannemann, M., Green, R. E., Lachmann, M., & Kelso, J. (2008). PatMaN: rapid alignment of short sequences to large databases. *Bioinformatics*, 24(13), 1530–1531. <https://doi.org/10.1093/bioinformatics/btn223>
- R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Ramos-Madrigrá, J., Runge, A. K., Bouby, L., Lacombe, T., Samaniego Castruita, J. A., Adam-Blondon, A., Figueiral, I., Hallavant, C., Martínez-Zapater, J. M., Schaal, C., Töpfer, R., Petersen, B., Sicheritz-Pontén, T., This, P., Bacilieri, R., Gilbert, M. T., & Wales, N. (2019). Palaeogenomic insights into the origins of French grapevine diversity. *Nature Plants*, 5(6), 595–603. <https://doi.org/10.1038/s41477-019-0437-5>
- Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., & Reich, D. (2015). Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1660), 20130624. <https://doi.org/10.1098/rstb.2013.0624>
- Schrumpf, R., Gottschalk, M., Metzger, J., Martinsson, G., Sieme, H., & Distl, O. (2016). Screening of whole genome sequences identified high-impact variants for stallion fertility. *BMC Genomics*, 17(1), 288. <https://doi.org/10.1186/s12864-016-2608-3>
- Schubert, M., Ermini, L., Sarkissian, C. D., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M. D., Fernández, R., Kircher, M., McCue, M., Willerslev, E., & Orlando, L. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature Protocols*, 9(5), 1056–1082. <https://doi.org/10.1038/nprot.2014.063>
- Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes*, 9(1), 88. <https://doi.org/10.1186/s13104-016-1900-2>
- 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., Ermini, L., Gamba, C., Weinstock, J., Vedat, O., & Orlando, L. (2017). Zonkey: A simple, accurate and sensitive pipeline to genetically identify equine F1-hybrids in archaeological assemblages. *Journal of Archaeological Science*, 78, 147–157. <https://doi.org/10.1016/j.jas.2016.12.005>
- Sirak, K., Fernandes, D., Cheronet, O., Harney, E., Mah, M., Mallick, S., Rohland, N., Adamski, N., Broomandkhoshbacht, N., Callan, K., Candilio, F., Lawson, A. M., Mandl, K., Oppenheimer, J., Stewardson, K., Zalzal, F., Anders, A., Bartík, J., & Coppa, A. ... Pinhasi, R. (2020). Human auditory ossicles as an alternative optimal source of ancient DNA. *Genome Research*, 30(3), 427–436. <https://doi.org/10.1101/654749>
- Skoglund, P., Northoff, B. H., Shunkov, M. V., Derevianko, A. P., Pääbo, S., Krause, J., & Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a siberian neanderthal. *Proceedings of the National Academy of Sciences of the United States of America*, 111(6), 2229–2234. <https://doi.org/10.1073/pnas.1318934111>
- Slon, V., Hopfe, C., Weiß, C. L., Mafessoni, F., de la Rasilla, M., Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M. V., Miller, R., Stewart, J. R., Derevianko, A. P., Jacobs, Z., Li, B. O., Roberts, R. G., Shunkov, M. V., de Lumley, H., Perrenoud, C., Gušić, I., ... Meyer, M. (2017). Neandertal and Denisovan DNA from Pleistocene sediments. *Science*, 356(6338), 605–608. <https://doi.org/10.1126/science.aam9695>
- Suchan, T., Kusliy, M. A., Khan, N., Chauvey, L., Tonasso-Calvière, L., Schiavinato, S., Southon, J., Keller, M., Kitagawa, K., Krause, J., Bessudnov, A. N., Bessudnov, A. A., Graphodatsky, A. S., Valenzuela-Lamas, S., Wilczyński, J., Pospuła, S., Tunia, K., Nowak, M., Moskal-delHoyo, M., ... Orlando, L. (2021). Performance and automation of ancient DNA capture with RNA hyRAD probes. *Molecular Ecology Resources*, 22(3), 891–907. <https://doi.org/10.1111/1755-0998.13518>
- Suchan, T., Pitteloud, C., Gerasimova, N. S., Kostikova, A., Schmid, S., Arrigo, N., Pajkovic, M., Ronikier, M., & Alvarez, N. (2016). Hybridization capture using RAD probes (hyRAD), a new tool for performing Genomic analyses on collection specimens. *PLoS One*, 11(3), e0151651. <https://doi.org/10.1371/journal.pone.0151651>
- van der Valk, T., Pečnerová, P., Díez-del-Molino, D., Bergström, A., Oppenheimer, J., Hartmann, S., Xenikoudakis, G., Thomas, J. A., Dehasque, M., Sağlıcan, E., Fidan, F. R., Barnes, I., Liu, S., Somel, M., Heintzman, P. D., Nikolskiy, P., Shapiro, B., Skoglund, P., Hofreiter, M., ... Dalén, L. (2021). Million-year-old DNA sheds light on the genomic history of mammoths. *Nature*, 591(7849), 265–269. <https://doi.org/10.1038/s41586-021-03224-9>
- Wallner, B., Palmieri, N., Vogl, C., Rigler, D., Bozlak, E., Druml, T., Jagannathan, V., Leeb, T., Fries, R., Tetens, J., Thaller, G., Metzger, J., Distl, O., Lindgren, G., Rubin, C.-J., Andersson, L., Schaefer, R., McCue, M., Neuditschko, M., ... Brem, G. (2017). Y chromosome uncovers the recent oriental origin of modern stallions. *Current Biology*, 27(13), 2029–2035.e5. <https://doi.org/10.1016/j.cub.2017.05.086>
- Wang, C., Li, H., Guo, Y. U., Huang, J., Sun, Y., Min, J., Wang, J., Fang, X., Zhao, Z., Wang, S., Zhang, Y., Liu, Q., Jiang, Q., Wang, X., Guo, Y., Yang, C., Wang, Y., Tian, F., Zhuang, G., ... Zhong, J. (2020). Donkey genomes provide new insights into domestication and selection for coat color. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-020-19813-7>
- Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Suchan, T., Chauvey, L., Poulet, M., Tonasso-Calvière, L., Schiavinato, S., Clavel, P., Clavel, B., Lepetz, S., Seguin-Orlando, A., & Orlando, L. (2022). Assessing the impact of USER-treatment on hyRAD capture applied to ancient DNA. *Molecular Ecology Resources*, 22, 2262–2274. <https://doi.org/10.1111/1755-0998.13619>