

This is a repository copy of *Paging through history : parchment as a reservoir of ancient DNA for next generation sequencing*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/82812/>

Version: Submitted Version

Article:

Teasdale, Matthew David orcid.org/0000-0002-7376-9975, van Doorn, NL, Fiddymment, S et al. (5 more authors) (2015) *Paging through history : parchment as a reservoir of ancient DNA for next generation sequencing*. *Philosophical Transactions Of The Royal Society Of London Series B - Biological Sciences*. 20130379. ISSN 1471-2970

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Paging through history: parchment as a reservoir of ancient DNA for next generation sequencing

MD Teasdale¹, NL van Doorn², S Fiddymment², CC Webb³, T O'Connor², M Hofreiter^{2,4}, MJ Collins², DG Bradley¹

¹Smurfit Institute of Genetics, University of Dublin, Trinity College, Dublin 2, Ireland.

²BioArCh, and ³Borthwick Institute for Archives, University of York, York, YO10 5DD, England.

⁴Institute for Biochemistry and Biology, Faculty of Natural Sciences, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany

Correspondence: dbradley@tcd.ie

ABSTRACT

Parchment represent an invaluable cultural reservoir. Retrieving an additional layer of information from these abundant, dated livestock-skins via the use of ancient DNA (aDNA) sequencing has been mooted by a number of researchers. However, prior PCR-based work has indicated that this may be challenged by cross-individual and cross-species contamination, perhaps from the bulk parchment preparation process. Here we apply next generation sequencing to two parchments of 17th and 18th century northern English provenance. Following alignment to the published sheep, goat, cow and human genomes it is clear that the only genome displaying substantial unique homology is sheep and this species identification is confirmed by collagen peptide mass spectrometry. Only

4% of sequence reads align preferentially to a different species indicating low contamination across species. Moreover, mitochondrial DNA sequences suggest an upper bound of contamination at 5%. Over 45% of reads aligned uniquely to the sheep genome, and even this limited sequencing exercise yields 9% and 7% of each sampled sheep genome post filtering, allowing the mapping of genetic affinity to modern British sheep breeds. We conclude that parchment represents an excellent substrate for genomic analyses of historical livestock.

KEY WORDS

Parchment, Next generation sequencing, ancient DNA, ZooMS, sheep

INTRODUCTION

Before the mass production of paper, parchment was the major medium for codices and until the widespread adoption of typewriters, they were a clerk's preferred medium for many formal legal documents and records [1]. There are several aspects of parchments that mark them as compelling substrates for DNA extraction and analysis. Firstly, parchments are made from the skins of domestic animals, particularly cattle, sheep and goats, which are dehaired, stretched, dried, scraped and pounced [1,2]. This manufacturing process results in robust artefacts, which can survive intact for many centuries [1,2]. Secondly, parchments/parchment manuscripts are not only abundant and widespread, but because of their enduring legal and evidential value they have typically been carefully managed throughout their lives and, in the twentieth century, curated and

protected from both high temperatures and fluctuating humidity. Indeed, the number of skins is truly staggering, even if, as is likely, a high percentage of documents have been destroyed. In the UK alone, assuming the number of sheep slaughtered annually remained constant at 15M from 1150-1850 [3], then if only 1% of all the skins became parchment and only 4% survived this would equate to 4.2M animals' skins [3,4]. The Norfolk Record Office alone estimates that it holds 20M parchment items spanning 1,000 years (A. Curtis *pers comm*). If this number is multiplied across the UK and European archives, there may be as many surviving skins as there are animals living in Europe today. Thirdly, unlike bone remains, of which only a fractional percentage survives and much less have been excavated, all the skins are above ground, archived and in the case of legal documents, directly dated to specific calendar years and, usually, precise days, which is a level of resolution not readily achievable with any other historic DNA source. Even documents that do not carry a direct date can be dated palaeographically to a resolution better than radiocarbon dating (without the expense of this process) [5]. Finally, there is enormous interest in the genetics of the main parchment species, cattle, sheep and goat, each with vibrant research communities investigating both geographical and temporal genetic variation [6–8].

Though understudied, parchment has been the subject of prior ancient DNA research. For example, the Dead Sea scrolls have yielded mitochondrial DNA PCR fragments, which have been identified as ibex and goat [9,10]. Similarly, Poulakakis *et al.* [11] identified three 13th-16th century Greek Parchments as of goat origin. Promisingly, Burger *et al.* [12,13] also recovered autosomal DNA amplicons in addition to mtDNA from

parchments, suggesting the possibility of high-resolution genetic inference. However, one recent study has been less encouraging. Campana *et al.* [2] investigated 18th-19th century British parchments and found that a majority of these gave heterogeneous mtDNA amplification products with signatures from multiple individuals and species, in addition to a high proportion of chimeric PCR artifacts. This result was attributed to cross contamination in the industrial parchment production process, during which multiple animal skins may have been washed, cured and depilated together [2].

However, PCR-based ancient DNA research has well documented deficiencies, particularly with regard to controlling and estimating contamination [14,15]. A central issue is that PCR favours longer, less damaged templates and thus has a bias for contaminant over endogenous DNA, making a representative sampling of target molecules impossible. In contrast, next generation sequencing (NGS) of ancient DNA works well with shorter fragments, generates many orders of magnitude more data, shows greater sensitivity, including the analysis of autosomal DNA, and is less prone to the chimeric artefactual sequences that can emerge from PCR [16]. Already, next generation sequencing of ancient nuclear DNA has provided insights into the evolutionary history of both extant and extinct species [17–21], but the routine analysis of ancient nuclear genomes is limited by the availability of well-preserved historic and archaeological samples [22,23]. The main limitation for the analysis of bone specimens lies in the fact that most paleontological and archaeological samples are found to contain high levels of bacterial and low levels of endogenous DNA (~0.1-5%) [18,19,21,24], although there are some notable high profile exceptions [20].

Here we present a molecular archaeological analysis of parchment using two historical samples from the Borthwick Institute for Archives at The University of York dated palaeographically to the 17th (PA1) and 18th (PA2) century, respectively. Both of these samples are identified as sheep and give high proportions of endogenous DNA with very low contamination from other species or co-specific individuals, suggesting parchment as an excellent source of historic DNA. Following a conventional agricultural history narrative, PA1 predates the livestock ‘Improvements’ driven by Bakewell and Ellman amongst others, whilst PA2 falls into the period when the new breeds of sheep were being developed and spread. The variation uncovered shows the potential of genetics in localizing the geographic origins of artefacts.

MATERIALS AND METHODS

Zooarchaeology by Mass Spectrometry (ZoomS)

Parchment samples of circa 5 x 5 mm from PA1 (17th century) and PA2 (18th century), were obtained from the Borthwick Institute for Archives parchment discards, and incubated twice for one hour in 50 mM ammonium bicarbonate buffer (pH 8.0) at 65°C following the method of Van Doorn *et al.* [25]. The first extract was discarded and the second extract was trypsinated overnight at 37°C. The tryptic digest was transferred over C18 resin to desalt and concentrate peptides by washing with 0.1% trifluoroacetic acid (TFA). Peptides were eluted in a final volume of 10 µl of 50% acetonitrile (ACN) / 0.1%

TFA (v/v). 1 µl of elute was mixed on a ground steel plate with 1 µl α-cyano-4-hydroxycinnamic acid matrix solution (1% in 50% ACN / 0.1% TFA (v/v/v)) and air-dried. Samples were analysed using a calibrated Ultraflex III MALDI-TOF instrument in reflector mode. Peptides were identified manually according to Buckley *et al.* [26,27]. Campana *et al.* [28] have confirmed the ability of collagen to discriminate sheep and goat [27] using DNA sequencing.

Ancient DNA extraction

All DNA extractions were performed in a dedicated ancient DNA laboratory at Trinity College Dublin on the same parchment samples used for ZooMS. Given the pilot nature of this analysis, the maximum amount of starting material available was used for each extraction, with DNA retrieved from 2x2cm² (~0.05g) pieces cut from both parchments, prepared using procedures previously described by our group [29,30].

Illumina sequencing library preparation

Illumina single read sequencing libraries were produced for each of the two parchment samples via PCR amplification of end repaired adaptor-ligated DNA templates following [29]. Samples were indexed following the Craig *et al.* [31] method of barcoding. Two PCR amplifications (20µl) were performed for each enrichment step comprising 3µl of end-repaired-adaptor-ligated parchment DNA, 10µl Phusion® high-fidelity PCR master

mix with HF buffer 6.2µl ddH₂O and 0.4µl each of both the forward and reverse primers. Amplification reactions per library consisted of an initial denaturation step of 98°C for 30s, then 12 cycles of 98°C for 10s, 65°C for 30s and 72°C for 30s, followed by a final extension step of 72°C for 5 min. The final PCR products from each sample (2 each) were then pooled and visualized on an Agilent 2100 Bioanalyzer using a DNA 1000 chip. Both samples were then combined in equimolar ratios and sequenced together on a single (SE 49bp) lane of an Illumina HiSeq2000 at BGI.

Initial parchment sequencing quality control

Initial quality control of sequencing reads were performed using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Adapter sequences were trimmed from the 3' end of the reads using CUTADAPT [32]. CUTADAPT's default settings were modified to require only a 1bp overlap between the 3' sequence of a read and an adapter sequence for it to be trimmed. This highly conservative approach will lead to a high proportion of reads being trimmed due to spurious matches. However, this approach was selected to try to guard against subsequent adapter sequence poisoning of downstream analyses, which can lead to poor alignment of reads and mis-identification of sequence polymorphisms.

FastQ Screen

FastQ Screen (http://www.bioinformatics.babraham.ac.uk/projects/fastq_screen) is a Perl

wrapper script, which allows for the same sequencing library to be easily aligned to multiple reference genomes using Bowtie [33]. The percentage of raw parchment reads that aligned uniquely to a single genome, to multiple places in the same genome, uniquely in multiple genomes and to multiple places in multiple genomes can then be assessed. FastQ Screen was used in this analysis to align the trimmed reads (30bp minimum read length) to 4 genomes, sheep (oviAri3), cow (bosTau7), human (hg19) and goat (chir1). FastQ Screen alignment settings were modified to use Bowtie's "end-to-end" algorithm and to allow the number of mismatches between the read and the reference genome to vary between 0 and 3.

BWA Sequence Alignment

The raw trimmed reads (30bp minimum read length) were also aligned to the Sheep reference genome (oviAri3) minus the mitochondrial genome [21] using BWA [34]. Standard alignment settings for the use of BWA with ancient DNA were used [35]. Aligned reads were then further filtered for a minimum mapping quality of 30 and redundant clonal PCR amplified reads removed using the SAMtools rmdup command [36]. Uniquely aligned reads were then identified by the XT, X1 tags, produced from the BWA alignment.

Alignment to human genome and variant calling

All sequencing reads were further aligned to the human genome (hg19) using BWA and

SAMtools with identical parameters to the sheep (oviAri3) alignments above. Any non-clonal reads that aligned to the human genome were then removed from subsequent analysis irrespective of mapping quality. SNPs were called using established protocols for aDNA NGS data [18,24]. Briefly, SNPs were called for all positions in which the shotgun sequencing of the parchment overlapped with the positions of SNPs in the ovine HapMap dataset (oviAri3 alignment), requiring a minimum base quality of 15 and mapping quality of 30. If multiple reads overlapped a SNP position, one read was then taken at random and used for base calling, C/T and G/A SNPs were also removed.

SNP merging and Allele sharing

SNPs called from PA1 and PA2 were then doubled to create pseudo-diploid data and merged with data from the sheep HapMap (ovine 50K panel, supplemental table 2) [6] using PLINK [37]. PA1 and PA2 SNPs were flipped to match the orientation of the HapMap with A/T and G/C SNPs removed from the analysis. Allele sharing distances were then calculated using a custom Python script and visualised using the ArcMap software in the ArcGis suite (Environmental Systems Research Institute).

Whole Mitochondrial genomes

Full mitochondrial genomes were produced from alignments of the parchment reads to a modern mitochondrial reference genome (HM236176) with BWA. Redundant reads were then removed using SAMtools. Modern sheep mitochondrial reference genomes were downloaded from GenBank (n=23) to allow the placement of the parchment samples

within this dataset. Multiple sequence alignments of both the modern and parchment mtDNA genomes were completed using the MUSCLE alignment algorithm [38] implemented in SeaView [39]. Neighbour joining trees (1,000 bootstrap replicates) were then produced from the alignment data in Seaview using the default Jukes and Cantor model and ignoring gapped sites.

RESULTS

Species Identification

Identification of the source species of each parchment was completed using a combined proteomic/genomic approach. The results of both these analyses identified sheepskin as the likely origin. After re-alignment to the sheep genome and filtering for mapping quality and contamination (reads that aligned the human genome) a final set of 6,047,847 (35.5%) reads were retained for PA1 and 5,256,723 (16.7%) for PA2; which equates to a retrieval of 9.4% and 7.9% of the sheep genome in PA1 and PA2 respectively. Both parchments were also identified as being produced from ewes via the analysis of the ratio of X chromosome to autosome reads [21].

Mitochondrial DNA analysis

In total 25,314 reads aligned to the ovine mitochondrial genome after duplicate removal (Table 1) (PA1 = 11,271, PA2 = 14,043). This allowed for the production of whole

mitochondrial genomes from both samples with an average read depth of 28X and 33X, respectively. These genomes were then compared to a modern reference dataset (n=23) (supplemental table 1) including 16 domestic sheep and 6 other ovine samples. Both parchments were found to locate within the domestic sheep mitochondrial haplogroup B (supplemental figure 3), which is predominant in both the modern and ancient sheep populations of Europe [40,41].

The high level of mitochondrial genome coverage achieved in the parchment sequencing allows for a rough estimation of the contamination rate in these samples to be calculated (both historic and modern). To do this, haplotype informative polymorphic positions in both samples, outside of the difficult to align 75-76bp repeat motif located within the control region of the ovine mtDNA [42,43], that may be due to contamination, sequencing error, heteroplasmy or DNA damage were analyzed [24]. At these positions we found the sample consensus base in 96% of sequences (282/291) for PA1 and 95% (381/398) of sequences for PA2, giving maximum estimated contamination rates of 4% for PA1 and 5% for PA2. These figures are slightly less than in the study by Sánchez-Quinto et al [24] of a complete ancient human, mitochondrial genome, which using similar methods estimated exogenous contamination at 8%.

SNP analysis

Both parchment samples were genotyped following established protocols for aDNA data [18,24]. 3,168 SNPs from the ovine 50K panel were called for PA1 and 2,291 for PA2

after filtering and merging with the modern genotype data. Average allele-sharing scores between each parchment genotype and extant geographically sampled populations were calculated and are summarized graphically in the interpolated contour maps in figure 2. Despite the limited recovery of SNP genotypes, a localization towards the British Isles is seen.

DISCUSSION

Parchment is ubiquitous in the historical record and is an attractive source for aDNA analysis to elucidate the history of domestic species and address questions intrinsic to the material origins of documents. However, published reports of first generation parchment aDNA investigations indicate conflicting results, reporting both unique PCR amplification products from a single species [11,13,44] and, more recently, multiple sequences from multiple species [2]. The latter finding seemed to imply cross-contamination between skins during the preparation process, where multiple animals would have been co-treated, a result potentially fatal for successful analysis.

In contrast to these results, our samples showed no signs of manufacture contamination in either proteomic or aDNA approaches. ZooMS identified sheepskin as the likely origin for both parchments and failed to retrieve any taxon-specific collagen masses for goat, pig or cow. FastQ Screen analysis of the raw DNA sequencing data identified the sheep genome as the most likely source for the majority of the sequences. This analysis also identified relatively few molecules that aligned uniquely to either the bovine, goat or

human genome at zero mismatches; 1.0%, 2.4% and 0.02% of all reads, respectively (figure 1, supplemental table 3). These non-source species alignment percentages are likely caused by homology between the ruminant genomes as well as damage in the aDNA molecules producing a better alignment to non-source species. Moreover, these percentages are likely to be inflated by the differences in the completeness of genome builds between the species [8], with the ovine genome being one of the least complete.

The high coverage mitochondrial genomes also allowed for the estimation of a within species contamination rate for both samples, which were found to be lower than that previously reported for human aDNA samples at 8% (4% for PA1 and 5% for PA2) [24]. Undetermined mismatch errors due to sequence misreads and DNA damage may also have contributed to these proportions, suggesting that the real contamination rate is even lower. In summary, the analysis of our shotgun data suggests that they were, at most, affected by low levels of contamination either from different individuals of the same species (sheep) or some of the most commonly contaminating species human, goat and cow. It should also be noted that both the above analyses (FastQ Screen and mtDNA) were completed on unfiltered data and are therefore likely to be unbiased by the high level of filtering conducted for the SNP analysis. Thus, our results are in agreement with some previous studies suggesting parchment as a valuable source for historic DNA sequences but contrast with the most recent work [2] that suggested high levels of contamination and artifacts in PCR based DNA analyses from parchment. It should be noted that many contaminant sequences detected by Campana *et al.* [2] were unusual, and the authors note the possibility of artifactual origin such as jumping PCR.

In order to provide as pure a sheep dataset as possible, stringent filters were applied in a re-alignment of all reads to the sheep genome to remove putative human contaminants. For both parchment samples more than 45% of reads aligned to the ovine genome once clonally amplified products were removed (Table 1). These numbers decrease to 35.5% (PA1) and 16.7% (PA2) when filters for mapping quality (minimum mapping quality = 30) and uniquely aligned reads are applied. These values however, still equate to a retrieval of 9.4 and 7.9% of the sheep genome in PA1 and PA2, respectively. These mapping percentages are still very high in comparison to those obtained for bone extracts, which are typically in the range of 0.1 to 5%. This large percentage of endogenous DNA is likely facilitated by the parchments' lower age, the nature of the source material (skin instead of bone) and the preferential conditions in which they were stored. This analysis suggests that parchments are an excellent source of historic DNA and are superior to bones and teeth of a similar age in the amount of endogenous DNA retrieved and resolution of the dating available. Reliable dating being critical in the analysis of the effects of agricultural selection.

If other parchments show similar levels of endogenous DNA content, DNA sequencing of historic domesticated animal genomes over a range of time periods could be accomplished, providing insights into the breeding history of domesticated animals; for example into sheep breeding before, during and after the agricultural improvements of the 18th century that led to the emergence of regional breeds of sheep in Britain.

Indeed it is intriguing to note that the two items, both of them records held in the Borthwick Institute for Archives in York, seem to have different heat map distributions, which may represent different breeding strategies employed between these two time points [45,46]. The visualisation of the genetic affinities of our two samples offers an illustration of the potential of co-analysis of dense modern SNP genotypes with next generation ancient DNA data. PA1 shows a strong affinity with northern Britain, specifically the region in which black-faced breeds such as Swaledale, Rough Fell and Scottish Blackface have a deep history [45]. The affinity with modern Texel sheep is intriguing in view of a rare retroviral insertion event seen in Texels and in two historic northern English breeds [47]. PA2 shows closer affinity with the Midlands and southern Britain, the region in which the livestock Improvements of the later 18th century were most active. Although this is somewhat speculative, the two specimens may derive from an unimproved northern hill-sheep (PA1), as might be expected in Yorkshire in the 17th century, and a sheep derived from the ‘improved’ flocks that were spreading through England in the 18th century, predominantly from estates in the Midlands (PA2). Although selected for a proof-of-principle investigation, these two documents may have given us a snapshot of livestock Improvement in process. This is a controversial period in agricultural history, specifically around the issue of whether the livestock Improvers developed new strains *de novo*, or built on changes that were already in progress [48]. As a productive and inherently datable biomolecular source, parchments will enable a more nuanced analysis of livestock regionalisation, one that complements the documentary record and is more chronologically precise than the archaeological record alone.

CONCLUSIONS

The findings of this first next generation sequencing study of parchment have shown that parchment is a highly suitable substrate for large-scale aDNA analyses. We were able to retrieve 9% and 7% of the ovine genome from PA1 and PA2, respectively, at high quality (mapping quality ≥ 30) using just half a lane each of an Illumina HiSeq 2000. This result suggests that the production of whole historic domesticate genomes or targeted exon sequencing from parchment is a realistic possibility. We were able to provide a unique species assignment for both pieces using both proteomic and aDNA methods and estimate an external contamination rate comparable to, and probably lower than from other aDNA sources. Whole mitochondrial genomes produced from both samples allowed their placement within the most populous sheep haplogroup and SNP data allowed an estimation of the modern day sheep breeds that most closely resemble the historic samples to be identified. Further sequencing of parchment samples from a variety of time periods and locations should allow for genetic maps from a variety of domestic species to be built, providing important insights into the last 1,000 years of animal breeding history.

ACKNOWLEDGMENTS

The authors would like to thank Alison Fairburn and Catherine Dand for their assistance with sampling for this project and Frauke Stock and two anonymous reviewers for helpful comments. The authors would also like to thank the ovine HapMap project for

access to the 50K genotype data. This work was supported by ERC Investigator grant 295729-CodeX, EU Marie Curie Fellowship Palimpsest to SF and EU Marie Curie ETN LeCHE Fellowship to MDT. The Ultraflex III was used courtesy of the York Centre of Excellence in Mass Spectrometry. The York Centre of Excellence in Mass Spectrometry was created thanks to a major capital investment through Science City York, supported by Yorkshire Forward with funds from the Northern Way Initiative.

DATA ACCESS

Raw sequence data can be found at the European Nucleotide Archive, study accession PRJEB5933 <http://www.ebi.ac.uk/ena/data/view/PRJEB5933>

REFERENCES

1. Bower, M. A., Campana, M. G., Checkley □ Scott, C., Knight, B. & Howe, C. J. 2010 The potential for extraction and exploitation of DNA from parchment: a review of the opportunities and hurdles. *J. Amer. Inst. Conserv.* **33**, 1–11.
2. Campana, M. G. et al. 2010 A flock of sheep, goats and cattle: ancient DNA analysis reveals complexities of historical parchment manufacture. *J. Archaeol. Sci.* **37**, 1317–1325.
3. Apostolides, A., Broadberry, S., Campbell, B., Overton, M. & van Leeuwen, B. 2008 English agricultural output and labour productivity, 1250–1850: some preliminary estimates. <http://www.basvanleeuwen.net/Papers.htm>.
4. Wrigley, E. A. 2006 The transition to an advanced organic economy: half a millennium of English agriculture1. *Econ. Hist. Rev.* **59**, 435–480.
5. Brock, F. 2013 Radiocarbon Dating of Historical Parchments. *Radiocarbon* **55**, 353–363.
6. Kijas, J. W. et al. 2012 Genome-Wide Analysis of the World’s Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biol.* **10**, e1001258.

7. Bovine HapMap Consortium et al. 2009 Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* **324**, 528–532.
8. Bai, Y., Sartor, M. & Cavalcoli, J. 2012 Current status and future perspectives for sequencing livestock genomes. *J. Anim. Sci. Biotechnol.* **3**, 8.
9. Woodward, S., Kahila, G., Smith, P., Greenblatt, C. & Zias, J. 1996 Analysis of parchment fragments from the Judean Desert using DNA techniques. In *Current Research and Technological Developments of the Dead Sea Scrolls: Conference on the Texts from the Judean Desert, Jerusalem, 30 April, 1995* (eds D. W. Parry & S. D. Ricks), pp. 215–238. Leiden: Brill.
10. Bar-Gal, G., Greenblatt, C., Woodward, S., Broshi, M. & Smith, P. 2001 The genetic signature of the dead sea scrolls. In *Historical Perspectives: From the Hasmoneans to Bar Kokhba in the Light of the Dead Sea Scrolls : Proceedings of the Fourth International Symposium of the Orion Center for the Study of the Dead Sea Scrolls and Associated Literature, 27-31 January, 1999* (eds D. Goodblatt A. Pinnick & D. Schwartz), pp. 165–171. Leiden: Brill.
11. Poulakakis, N., Tselikas, A., Bitsakis, I., Mylonas, M. & Lymberakis, P. 2007 Ancient DNA and the genetic signature of ancient Greek manuscripts. *J. Archaeol. Sci.* **34**, 675–680.
12. Burger, J., Hummel, S. & Herrmann, B. 2000 Palaeogenetics and cultural heritage. Species determination and STR-genotyping from ancient DNA in art and artefacts. *Thermochim. Acta* **365**, 141–146.
13. Burger, J., Pfeiffer, I., Hummel, S., Fuchs, R., Brenig, B. & Hermann, B. 2001 Mitochondrial and nuclear DNA from (pre) historic hide-derived material. *Anc. Biomol.* **3**, 227–238.
14. Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M. & Pääbo, S. 2001 Ancient DNA. *Nat. Rev. Genet.* **2**, 353–359.
15. Gilbert, M. T. P., Bandelt, H.-J., Hofreiter, M. & Barnes, I. 2005 Assessing ancient DNA studies. *Trends Ecol. Evol.* **20**, 541–544.
16. Krause, J., Briggs, A. W., Kircher, M., Maricic, T., Zwyns, N., Derevianko, A. & Pääbo, S. 2010 A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr. Biol.* **20**, 231–236.
17. Raghavan, M. et al. 2013 Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91.
18. Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E., Gilbert, M. T. P., Götherström, A. & Jakobsson, M. 2012 Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336**, 466–469.
19. Green, R. E. et al. 2010 A draft sequence of the Neandertal genome. *Science* **328**, 710–722.
20. Reich, D. et al. 2010 Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060.
21. Orlando, L. et al. 2013 Recalibrating Equus evolution using the genome sequence of an early

Middle Pleistocene horse. *Nature* **499**, 74–78.

22. Knapp, M., Clarke, A. C., Horsburgh, K. A. & Matisoo-Smith, E. A. 2012 Setting the stage – Building and working in an ancient DNA laboratory - Special Issue: Ancient DNA. *Annals of Anatomy - Anatomischer Anzeiger* **194**, 3–6.
23. Kirsanow, K. & Burger, J. 2012 Ancient human DNA. *Ann. Anat.* **194**, 121–132.
24. Sánchez-Quinto, F. et al. 2012 Genomic Affinities of Two 7,000-Year-Old Iberian Hunter-Gatherers. *Curr. Biol.* **22**, 1494–1499.
25. Van Doorn, N. L., Hollund, H. & Collins, M. J. 2011 A novel and non-destructive approach for ZooMS analysis: ammonium bicarbonate buffer extraction. *Archaeol. Anthropol. Sci.* **3**, 281–289.
26. Buckley, M., Collins, M., Thomas-Oates, J. & Wilson, J. C. 2009 Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **23**, 3843–3854.
27. Buckley, M., Witcher Kansa, S., Howard, S., Campbell, S., Thomas-Oates, J. & Collins, M. 2010 Distinguishing between archaeological sheep and goat bones using a single collagen peptide. *J. Archaeol. Sci.* **37**, 13–20.
28. Campana, M. G., Robinson, T., Campos, P. F. & Tuross, N. 2013 Independent confirmation of a diagnostic sheep/goat peptide sequence through DNA analysis and further exploration of its taxonomic utility within the Bovidae. *J. Archaeol. Sci.* **40**, 1421–1424.
29. Edwards, C. J. et al. 2010 A complete mitochondrial genome sequence from a mesolithic wild aurochs (*Bos primigenius*). *PLoS One* **5**, e9255.
30. MacHugh, D. E., Edwards, C. J., Bailey, J. F., Bancroft, D. R. & Bradley, D. G. 2000 The extraction and analysis of ancient DNA from bone and teeth: a survey of current methodologies. *Anc. Biomol.*
31. Craig, D. W. et al. 2008 Identification of genetic variants using bar-coded multiplexed sequencing. *Nat. Methods* **5**, 887–893.
32. Martin, M. 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12.
33. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.
34. Li, H. & Durbin, R. 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760.
35. Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J. F., Al-Rasheid, K. A. S., Willerslev, E., Krogh, A. & Orlando, L. 2012 Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**, 178.
36. Li, H. et al. 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.

37. Purcell, S. et al. 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575.
38. Edgar, R. C. 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.
39. Gouy, M., Guindon, S. & Gascuel, O. 2010 SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224.
40. Olivieri, A. et al. 2013 Mitogenomes from two uncommon haplogroups mark late glacial/postglacial expansions from the near east and neolithic dispersals within Europe. *PLoS One* **8**, e70492.
41. Hiendleder, S., Mainz, K., Plante, Y. & Lewalski, H. 1998 Analysis of mitochondrial DNA indicates that domestic sheep are derived from two different ancestral maternal sources: no evidence for contributions from urial and argali sheep. *J. Hered.* **89**, 113–120.
42. Meadows, J. R. S., Hiendleder, S. & Kijas, J. W. 2011 Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel. *Heredity* **106**, 700–706.
43. Hiendleder, S., Kaupe, B., Wassmuth, R. & Janke, A. 2002 Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proc. Biol. Sci.* **269**, 893–904.
44. Pangallo, D., Chovanova, K. & Makova, A. 2010 Identification of animal skin of historical parchments by polymerase chain reaction (PCR)-based methods. *J. Archaeol. Sci.* **37**, 1202–1206.
45. Wells, R. A. E. 1984 Sheep-Rustling in Yorkshire in the Age of the Industrial and Agricultural Revolutions. *Northern history* **20**, 127–145.
46. Wykes, D. L. 2004 Robert Bakewell (1725-1795) of Dishley: Farmer and Livestock Improver. *Agric. Hist. Rev.* **52**, 38–55.
47. Bowles, D., Carson, A. & Isaac, P. 2014 Genetic Distinctiveness of the Herdwick Sheep Breed and Two Other Locally Adapted Hill Breeds of the UK. *PLoS One* **9**, e87823.
48. Thomas, R., Holmes, M. & Morris, J. 2013 “So bigge as bigge may be”: tracking size and shape change in domestic livestock in London (AD 1220–1900). *J. Archaeol. Sci.* **40**, 3309–3325.

FIGURE CAPTIONS

FIGURE 1. Histograms illustrating the relative frequency of raw sequence read alignments to the human, cow, sheep and goat genomes with a tolerance of zero mismatches. A=PA1 and B=PA2, red - one hit one genome, orange - multiple hits one genome, dark blue - one hit multiple genomes, blue - multiple hits multiple genomes. Notably, only the sheep genome shows a significant body of aligning reads that do not also align to other species. Cross alignment of other reads is expected due to the high homology, especially of repeated elements, among ruminant genomes.

FIGURE 2. Synthetic maps illustrating average allele sharing between the two parchment partial genome sequences and reference genotypes from selected modern sheep breeds. Higher sharing is denoted by warmer colours. A localization of genetic affinity for both to western Europe is clear, with the earlier PA1 showing more sharing with Northern Britain and PA2 with Southern Britain and Ireland. Approximate geographical origin of breeds from Kijas *et al.* [6].

SHORT TITLE

Paging through history

TABLES

<i>sample</i>	<i>raw reads</i>	<i>aligned reads raw(%)</i>	<i>aligned reads high quality¹ (%)</i>	<i>aligned reads mtDNA</i>	<i>ovine 50k panel SNPs called</i>
PA1(17 th C)	17,006,629	11,292,116 (66.4)	6,047,847 (35.5)	11,271	3,168
PA2 (18 th C)	31,493,502	14,403,079 (45.7)	5,256,723 (16.7)	14,043	2,291

Table:1 Summary of ancient sequence data from both parchment samples.

1. High quality reads consists of a mapping quality ≥ 30 , no reads which also align to the human genome hg19 and unique as described by the XT and X1 of the BWA mapping algorithm.



