Inferring the Characteristics of Ancient Populations using Bioinformatic Analysis of Genome-wide DNA Sequencing Data

Graham Gower

Australian Centre for Ancient DNA
School of Biological Sciences
Faculty of Sciences
University of Adelaide

Thesis submitted in fulfilment of the
requirements for the degree of
Doctor of Philosophy

February 2019

Contents

Abstract

In this thesis, I apply, evaluate, and develop methods for learning about past populations from genome-wide sequencing data. Specifically, I:

- apply methods based on random genetic drift between populations, to determine that pre-Holocene gene flow occurred between the ancestors of domestic cattle (*Bos primigenius*) and European bison (*Bison priscus*), and that the contribution of *Bos* genealogy to the bison lineage was less than 10 %;
- use simulations to assess the impact of short genomic scaffolds when inferring past populations sizes with the pairwise sequentially Markov coalescent, and show that population size inferences can be robust for scaffold lengths as short as 100 kb;
- perform genetic sex determination of ancient DNA specimens to show that bison (*Bison spp.*) and brown bear (*Ursus arctos*) specimens are approximately 75 % male, and that male-biased observations likely stem from the ecological and social structures of the populations;
- and develop a suite of software tools for processing hairpin bisulfite sequencing data, which can be used to investigate genome-wide DNA methylation levels in ancient DNA.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree. The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works. I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time. I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Graham Gower

19/10/2018

Date

Acknowledgements

This thesis would not have been possible without the ongoing support of my supervisors Alan Cooper, Bastien Llamas, Julien Soubrier, and Kieren Mitchell. You provided me with an opportunity, and the freedom, to learn a great deal about a broad range of interesting biological phenomena. Thank you for your time, energy, criticism, and many motivating ideas. I also wish to thank all the members of ACAD, for enjoyable journal club discussions, social events, and sporadic tomfoolery. And speaking of tomfoolery, it's been a pleasure to collaborate with our friends in the School of Mathematical Sciences—Nigel, Jono, and Ben, you've inspired me to spend my leisure time reading maths books. Finally, to my partner Saskia; thank you for being there for me no matter what, and thank you for listening to my long winded and overly-detailed explanations with far too many digressions before sometimes finally getting to the point.

Chapter 1

Introduction

1.1 Evolution and population dynamics

Evolutionary biologists wish to learn about organisms living now, and those that lived in the past. Motivations are diverse and include understanding evolutionary processes, conservation of species, or characterising the etiology of diseases. How are living things related to each other? How did their form and function come to be? What causes them to go extinct? What processes that are occurring now, have also occurred in the past? Can the past and the present tell us about the future? Such ideas have a long history in biology, starting with Darwin, who was immensely influenced by the uniformitarianism of Lyell (1835). By comparing living species with those that are extinct, we can derive hypotheses about the circumstances for extinction, and why similar species did not suffer the same fate. Using what we learn of the past we can try to predict, and hopefully pacify, current and future extinction risk.

Population parameters such as spatial distribution, age-specific mortality, sex ratios, migration rates between populations, and population size can all be informative regarding behavioural and ecological characteristics. Changes to these population parameters over time can further pinpoint responses to the environment, with regard to large-scale disruptions from geological and climatic events. Demographic parameters of past populations have often been discussed by comparing the morphology of extant species, in the context of morphological assessments of fossils (Robson & Wood, 2008). This is made possible by a wealth of reference material for hominid remains, for example, to morphologically determine a specimen's sex (Frayer & Wolpoff, 1985; Rehg & Leigh, 1999), and age at death (Dean & Liversidge, 2015; Dean, 2016).

Increasingly, genetic data are being used in addition to, or in place of, morphological data. This is particularly important where remains are fragmentary or rare, such as for Denisovans, a hominin group originally described from a finger bone and a tooth (Reich *et al.*, 2010). But genetics also offers to answer questions that were previously difficult or impossible to answer, such as determining average generation times (Moorjani *et al.*, 2016), inferring ancestral population sizes (Li & Durbin, 2011), quantifying past gene flow (Patterson *et al.*, 2012), or clarifying that distinct morphological forms are different sexes of a single species (Bunce *et al.*, 2003; Huynen *et al.*, 2003; Bover *et al.*, 2018). Like for morphological data, two complementary approaches exist for investigating past populations with genetic data: use the data from modern individuals to infer things about their ancestors; or observe the populations directly from the DNA of subfossil remains.

1.2 Inference based on modern data

1.2.1 Detecting gene flow

The ancestors of modern populations leave extensive signatures in the genomes of their descendents. Phylogenetic relationships can be used to infer population split times (Bouckaert *et al.*, 2014; Molak *et al.*, 2013) and evolutionary rates of change (Bouckaert *et al.*, 2014; Rabosky, 2014). But relationships between individuals within a population are rarely tree-like, and similarly, speciation may not be characterised by clean separation into reciprocally monophyletic groups (Mallet, 2005). To detect gene flow in past populations, a four-population test statistic (F_4) has been developed (Reich *et al.*, 2009; Patterson *et al.*, 2012), which is sensitive to even small quantities of gene flow. The principal idea of the test is that in a phylogeny, the components of random genetic drift along two distinct branches are uncorrelated, whereas if gene flow existed between the two branches then genetic drift will be correlated.

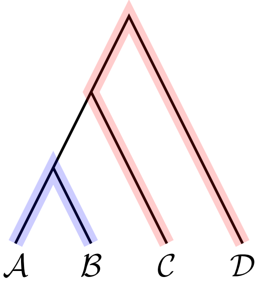


Figure 1.1: A four taxon phylogenetic tree with branches representing random genetic drift. The expected difference in allele frequencies between $\mathcal{A}$ and $\mathcal{B}$ ($\mathcal{C}$ and $\mathcal{D}$) is due to drift occurring along the path shown in blue (red).

Consider the four populations $\mathcal{A}, \mathcal{B}, \mathcal{C}$ and $\mathcal{D}$ in Figure 1.1, and suppose we wish to test for gene flow between $\mathcal{A}$ and $\mathcal{C}$ (or $\mathcal{B}$ and $\mathcal{C}$) that occurred after $\mathcal{A}$ and $\mathcal{B}$ split. If a, b, c, and d are the respective allele frequencies in populations $\mathcal{A}, \mathcal{B}, \mathcal{C}$, and $\mathcal{D}$, then Patterson's $F_4 = \mathrm{cov}(a - b, c - d)$. When the four populations have a strictly tree-like relationship, then the $\mathcal{A}$—$\mathcal{B}$ drift is uncorrelated with the $\mathcal{C}$—$\mathcal{D}$ drift, so $F_4 = 0$. In contrast, if gene flow has occurred between the ancestors of $\mathcal{A}$ and $\mathcal{C}$ (or the ancestors of $\mathcal{B}$ and $\mathcal{C}$), the excess in allele sharing between $\mathcal{A}$ and $\mathcal{C}$ (or $\mathcal{B}$ and $\mathcal{C}$) produces a correlation between the $\mathcal{A}$—$\mathcal{B}$ drift and the $\mathcal{C}$—$\mathcal{D}$ drift, so $F_4 \neq 0$. In the latter case, non-intersecting drift paths as shown in Figure 1.1 are not a complete representation of the relationships. The sign of the test statistic indicates whether gene flow occurred between $\mathcal{A}$ and $\mathcal{C}$ ($F_4 > 0$), or between $\mathcal{B}$ and $\mathcal{C}$ ($F_4 < 0$),

A related statistic, Patterson's D, can be used and interpreted in the same way as the F_4, but is normalised to have a scale between -1 and 1 (D is a correlation, while F_4 is a covariance). The D statistic is sometimes referred to as the "ABBA-BABA" test when calculated from single individuals in each of the four populations (Green *et al.*, 2010). This family of drift-based statistics are extensible, and have explicit ties to well established notions of allele frequency variation such as Wright's inbreeding coefficient (Bhatia *et al.*, 2013; Peter, 2016).

1.2.2 Coalescent theory and population size inference

The coalescent describes convergence of lineages backwards-in-time for a sample of homologous genes. For a population with fixed size N, the neutral Wright-Fisher model has N haploid individuals in the current generation choose their parents uniformly at random from the N individuals in the previous generation (Hein *et al.*, 2005, p. 13). Hence two individuals in the current generation have the same parent, *i.e.* they *coalesce* in the previous generation, with probability $1/N$. While this is a discrete-time process (distinct non-overlapping generations), Kingman (1982b) observed that it can be approximated by a continuous-time process with time scaled to have units of N generations. This continuous time approximation simplified proofs of many theoretical results, and was shown to be robust to departures from the neutral Wright-Fisher model, such as the Moran model for diploid organisms with overlapping generations (Kingman, 1982a,b). One immediate consequence of the coalescent is that the expected time to most recent common ancestor (TMRCA) is directly proportional to N (Tavaré, 1984), so small populations have a relatively recent TMRCA compared with that of large populations (see Figure 1.2). Changes in population size through time can be accounted for (Griffiths & Tavare, 1994; Donnelly & Tavare, 1995), and coalescent-based inference frameworks exist to estimate past population sizes from a collection of homologous non-recombining DNA sequences (Pybus *et al.*, 2000; Drummond *et al.*, 2005; Minin *et al.*, 2008).

Recombination produces distinct gene trees (marginal genealogies) at either sides of a recombination breakpoint (Hein *et al.*, 2005, pp. 138), and the *coalescent with recombination* was introduced to describe this gene-tree-generating process in a coalescent framework (Hudson, 1983, 1990). The algorithm is relatively straightforward, and is particularly useful for simulating recombining haplotypes (Hudson, 1990, 2002). However, the relationships between the marginal genealogies can be very complicated due to their spatial interdependence

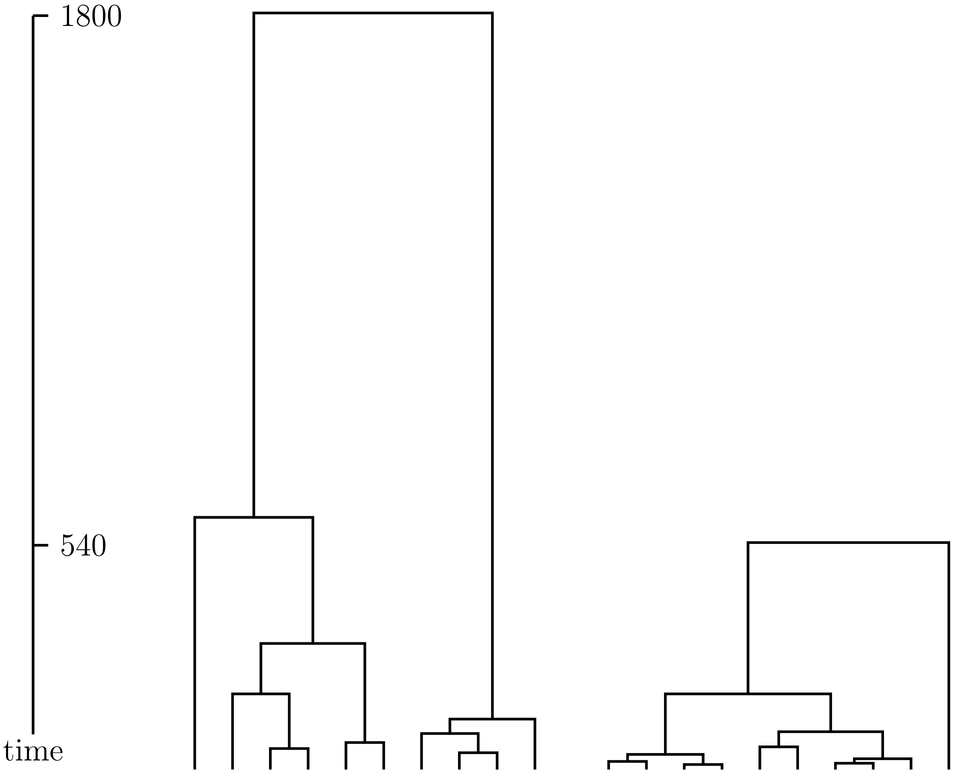


Figure 1.2: Expected coalescent times (in generations) for 10 samples from each of two Wright-Fisher populations. Left: $N = 1000$. Right: $N = 300$.

along chromosomes, termed an *ancestral recombination graph* (ARG) (Griffiths & Marjoram, 1997). Until recently, generating long sequences of genotypes using this exact process was computationally impractical, as the entire ARG must be simulated before mutations are placed onto the graph (for a memory efficient solution see Kelleher *et al.*, 2016). One workaround is the *sequentially Markov coalescent* (SMC), which generates marginal genealogies sequentially along a chromosome (Wiuf & Hein, 1999), and avoids the full ARG-generating process by making each marginal genealogy dependent only upon the last one generated (the Markov property) (McVean & Cardin, 2005; Marjoram & Wall, 2006).

While this approximation has been useful for rapidly simulating long haplotypes (Chen *et al.*, 2009; Staab *et al.*, 2015), an arguably more important application is in the generating process of a hidden Markov model (HMM). A HMM is a parameter inference framework that requires a Markovian model to describe how the data is generated (Rabiner, 1989; Durbin *et al.*, 1998), which can be used with the SMC to estimate population parameters from whole genome sequences (Dutheil *et al.*, 2009). The SMC/HMM approach was popularised by the pairwise SMC (PSMC) program, (Li & Durbin, 2011) which infers past population sizes from a single diploid genome. Population size inference has since been extended to simultaneously use up to 12 genomes with the multiple SMC (MSMC) program (Schiffels & Durbin, 2014) and hundreds of genomes in the SMC++ program (Terhorst *et al.*, 2017).

A population bottleneck in the recent past implies that almost all gene trees for modern haplotypes will coalesce at, or more recently than, the time of the bottleneck (Hein *et al.*, 2005, pp. 104-106). So while a post-bottleneck individual's genome is the direct result of ancestral processes, the signatures of coalescent and recombination events prior to the bottleneck have been largely erased. For example, comparative analyses between modern representatives of a domestic species, and a related wild species, can reveal changes that occurred after their split and prior to domestication, but it may not be possible to accurately determine the timing of such changes.

1.3 Direct observation using ancient DNA

1.3.1 What is ancient DNA?

Ancient DNA (aDNA) refers to the DNA of a dead organism, where the DNA has not been deliberately preserved, and may thus be partially degraded. DNA can remain preserved in and on bones, teeth, coprolites, hair, soft tissue, soil, ice, or elsewhere, long after the death of the organism from which it originated. Cold and dry conditions are the best for preserving DNA (Hofreiter *et al.*, 2015; Kistler *et al.*, 2017), and permafrost holds the record for the oldest sample from which a genome has been successfully obtained, a 700 thousand-year-old horse (Orlando *et al.*, 2013). While obtaining DNA from a specimen this old is atypical, successful DNA extraction is regularly reported from samples up to and beyond the 50 thousand-year limit of radiocarbon dating (*e.g.* see sample ages in Shapiro *et al.*, 2004). This time span encompasses a number of climatic fluctuations and megafaunal extinctions (Cooper *et al.*, 2015), making aDNA well suited to studying extinctions and responses to climate change. Methods for analysing modern data can generally be applied to aDNA datasets, and datasets compiled from both modern and ancient sources. But aDNA provides the ability to directly observe past populations, in a way that is not possible with sequencing data from modern individuals alone.

1.3.2 Challenges for ancient DNA studies

When compared to DNA extracted from a living organism, aDNA is considerably more fragmented, contains single nucleotide miscoding lesions, and is usually highly contaminated with non-target DNA (Pääbo *et al.*, 1989). Fragmentation of DNA following the death of an organism may be caused by enzymatic activity, or hydrolysis of phosphodiester bonds in the DNA backbone (Pääbo *et al.*, 2004; Briggs *et al.*, 2007; Overballe-Petersen *et al.*, 2012; Dabney *et al.*, 2013). Very few long fragments remain in aDNA, as the proportion of recoverable molecules decreases exponentially with molecule length (Glocke & Meyer, 2017). Almost all fragmentation occurs soon after the death of the organism, and loss of DNA over time is most likely due to bulk diffusion out of the tissue (e.g. porous bone) (Pääbo *et al.*, 1989; Kistler *et al.*, 2017). The fragmentation process leaves single-stranded DNA protruding from the ends of otherwise double-stranded DNA molecules. Within the single-stranded portion, cytosine residues may be converted into uracil (C→U) via spontaneous hydrolytic deamination (Hofreiter *et al.*, 2001; Brotherton *et al.*, 2007; Briggs *et al.*, 2007). While this process also operates within double-stranded DNA, it is far more frequent in single-stranded DNA (Frederico *et al.*, 1990; Lindahl, 1993). Cytosine deamination results in an excess of transition substitutions in sequencing data, which to a large extent can be mitigated by removing uracils with uracil-DNA-glycosylase (UDG) prior to library construction (Briggs *et al.*, 2010). However, not all deaminated cytosines are removed by this enzyme (see DNA methylation section below), so it is common to exclude transition substitutions from certain analyses, even for data from UDG-treated libraries.

Ancient remains contain DNA that derives from sources other than the target of interest (Pääbo *et al.*, 1989; Cooper & Poinar, 2000). The DNA recovered from a subfossil is regularly dominated by microbial and fungal contaminants that colonise the sample post mortem. Additional contamination may be introduced by handling the sample during collection, during subsequent laboratory procedures, or from the laboratory reagents themselves. Separating the target of interest from contaminating sources is thus essential, particularly in the extreme case that a contaminating sequence is closely related to the target organism (*e.g.* human contamination of a human sample). Because modern contaminants are far less likely to contain terminal C→U substitutions, it is common to use these as indicators of authentic aDNA molecules (Skoglund *et al.*, 2014; Meyer *et al.*, 2016).

The short and damaged DNA fragments obtained from degraded specimens are poorly suited to *de novo* assembly (Nagarajan & Pop, 2013; Ekblom & Wolf, 2014; Sohn & Nam, 2018). Repetitive parts of the genome longer than the length of a single fragment cannot be traversed, so complex mammalian genomes will only be assembled into very short contigs, with limited ability to order and orient contigs into longer scaffolds (*e.g.* see Feigin *et al.*, 2018). Hence studies of nuclear ancient DNA almost exclusively rely upon mapping reads to a reference assembly. For ancient specimens corresponding to an extant lineage, there may exist a genome reference assembled using sequences derived from a modern individual, but extinct lineages must rely on genomic resources from a less related taxon. A choice may be required between mapping reads to a low-quality reference assembly of a closely related non-model organism, or mapping to the more complete reference assembly of a distantly related model organism. Genome references for non-model organisms are typically assembled from short-read sequencing data only, and such assemblies are notoriously characterised by misassemblies and low contiguity (Earl *et al.*, 2011; Zhang *et al.*, 2012; Bradnam *et al.*, 2013; Denton *et al.*, 2014; Briskine & Shimizu, 2017). On the other hand, reads mapped to a distantly related reference tend to map uniquely only in regions with relatively high sequence homology between the target organism and the reference, such as conserved regions of the genome (Prüfer *et al.*, 2010). This produces a reference-homology bias that is exacerbated for samples with fewer or shorter endogenous reads.

1.3.3 Ancient DNA analyses

Ancient DNA can be used to detect and date gene flow from an extinct population into an extant lineage (Green *et al.*, 2010; Sankararaman *et al.*, 2012), or investigate the kinds of genetic deterioration that occur immediately prior to an extinction (Palkopoulou *et al.*, 2015; Rogers & Slatkin, 2017). For ancient specimens with modern descendents, genetic comparisons permit an investigation of how populations have changed over time, such as the progressive dilution of genetic material from archaic introgression (Fu *et al.*, 2016). Predomestication specimens have been used on numerous occasions to determine the timing and location of domestication events (Skoglund *et al.*, 2015; Scheu *et al.*, 2015; Caliebe *et al.*, 2017; Ottoni *et al.*, 2017; Dymova *et al.*, 2017; Daly *et al.*, 2018), which are relatively cryptic to analysis from modern genetic data due to very severe domestication bottlenecks. Another promising avenue of research is the identification of DNA methylation patterns in ancient individuals.

DNA methylation is an epigenetic mechanism that facilitates gene regulation, and has been implicated as a major contributor for cell differentiation, X-inactivation, parent-of-origin imprinting, transposable element silencing, stress response, and various diseases (Stewart *et al.*, 2016; Edwards *et al.*, 2017; Barros-Silva *et al.*, 2018; Zhang *et al.*, 2018). The identification of differential methylation between species can therefore be an indicator of functional differences (Hernando-Herraez *et al.*, 2015). The most common form of DNA methylation in mammals, 5-methylcytosine (5mC), was initially detected in aDNA by treating extracted DNA with UDG to remove uracil residues (Briggs *et al.*, 2010). The remaining miscoding lesions were almost entirely CpG→TpG substitutions in the nuclear genome. In mammals, 5mC occurs mostly in CpG contexts (Edwards *et al.*, 2017), and is not found in the mitochondria (Mechta *et al.*, 2017). Methylated cytosines deaminate directly to thymine (5mC→T), rather than to uracil (C→U, in unmethylated cytosines), and so are not removed from DNA by UDG (Lindahl, 1979). Although only deaminated 5mC sites could be inferred, such sites accumulate in a time-dependent manner (Ehrlich *et al.*, 1986; Shen *et al.*, 1994), which implies the long-term survival of 5mC after death. Using UDG-treated samples, it is possible to computationally assess regional methylation levels (Pedersen *et al.*, 2014; Gokhman *et al.*, 2014). But this approach requires deep sequencing of the aDNA sample, and is thus limited to exceptionally well preserved samples.

The first direct evidence of post-mortem 5mC preservation was obtained by applying bisulfite sequencing to aDNA (Llamas *et al.*, 2012). Bisulfite treatment converts unmethylated cytosines into uracils but leaves methylated cytosines intact (Frommer *et al.*, 1992; Clark *et al.*, 1994). The data are mapped to a reference using bisulfite-aware software, and the methylation status can be determined with base-level precision from positions where reads contain thymines but the reference has a cytosine (Krueger *et al.*, 2012). However, bisulfite treatment is also a harsh chemical agent—once double-stranded library molecules are denatured, strand breakages can occur (preferentially in longer molecules) which make DNA unamplifiable (Munson *et al.*, 2007). Thus bisulfite treatment applied to aDNA exacerbates the difficulty of uniquely mapping reads to a reference sequence, as reads have low complexity (comprised of mostly of A, G, and T bases), and have very short average length. Hence for the majority of aDNA specimens, new approaches will be required to profile DNA methylation.

1.4 Thesis overview

1.4.1 Motivation & Aims

Genome-wide DNA sequencing data from modern and ancient individuals both provide the opportunity to learn about populations that lived in the past. Thus, advances in how we obtain, process, and analyse these data are important for being able to answer detailed questions about a wide range of organisms, and evolutionary processes more generally. In this thesis, I aim to broaden our understanding and our ability to characterise past mammal populations. I aim to do this by: applying and extending existing methods to new datasets; evaluating the limitations of existing tools; and developing new software.

1.4.2 Early cave art and ancient DNA record the origin of European bison

It is uncontroversial that European bison (*Bison bonasus*) and American bison (*Bison bison*) are more closely related than either is to other bovids, and that they form a sister group to yak (*Bos grunniens*) (Groves & Grubb, 2011; Hassanin *et al.*, 2013). Yet European bison have a mitochondrial lineage more closely related to domestic cattle (*Bos taurus*) than to American bison (Bibi, 2013). Further, both European and American bison have fertile female offspring when crossbred with other *Bos* species such as domestic cattle. This has led to suggestions that the ancestor of European bison received female-biased gene flow from aurochs (*Bos primigenius*, the ancestor of domestic cattle), facilitating the capture of a *Bos*-like mitochondrial lineage (Verkaar *et al.*, 2004). In chapter 2, I use genome-wide nuclear SNP data from ancient steppe bison (*Bison priscus*, the ancestor of European bison) to determine the extent of gene flow, if any, between pre-Holocene *Bison* and *Bos* lineages. The ratio of two F_4 statistics (Patterson *et al.*, 2012) is used to obtain an upper bound on the quantity of gene flow, and an exact estimate is derived by simulating data and using approximate Bayesian computation (Beaumont *et al.*, 2002) to obtain a parameter estimate. Chapter 2 also identifies an extinct *Bos*-like mitochondrial lineage related to European bison, which is likewise assessed for gene flow. A hypergeometric test is then used to determine if putative introgressed *Bos* SNPs are common between the extinct and extant mitochondrial lineages.

1.4.3 Population size history from short scaffolds: how short is too short?

Two ubiquitous programs for inferring past population size fluctuations, from a single diploid genome, are based upon the SMC (Li & Durbin, 2011; Schiffels & Durbin, 2014). These programs apply an HMM to contiguous stretches of genomic information to identify changes in the local density of heterozygous sites along the genome, which indicate ancestral recombination breakpoints. Each local recombination block has its own TMRCA, which are used to infer the distribution of TMRCAs, and hence estimate past population sizes.

However, data for non-model organisms are often reliant upon mapping to a low-quality reference assembly, containing tens or hundreds of thousands of ultra-short contigs or scaffolds. In chapter 3, I use simulations and empirical data to assess the robustness of SMC-based population size inferences for scaffold lengths as short as 10 kb.

1.4.4 Widespread male sex bias in mammal fossil and museum collections

Sex determination from shotgun sequencing data can be performed by calculating the ratio of reads mapping to the X chromosome versus the autosomes (Skoglund *et al.*, 2013; Mittnik *et al.*, 2016). For males, this ratio is half that obtained for females, due to X chromosome copy number differences. Using this approach, a significant excess of males has recently been observed in mammoth fossil remains (Pečnerová *et al.*, 2017). The segregation of sexes in mammoth herds was proposed to explain this bias, as young adult males are largely solitary and thought to be more exposed to dying in taphonomically favourable locations. Pečnerová *et al.* (2017) further predicted that a male bias would be found for steppe bison remains, as bison have superficially similar sex-segregated herds for most of the year.

This prediction is tested in chapter 4, by calculating the sex ratio in genetically sexed subfossil bison (*Bison spp.*) remains. The sex ratio is also calculated for brown bear (*Ursus arctos*) subfossils, as they have a very different life history strategy and social structure to that of mammoths and bison, and could thus be considered a negative control. Genetic sexing from shotgun sequencing data is performed using explicit male and female binomial models, and model selection with a likelihood ratio test, which differs from previous methods that used an arbitrary threshold value for separating males and females (Skoglund *et al.*, 2013; Mittnik *et al.*, 2016). This approach is applied here to shotgun data aligned to a scaffold-level reference assembly for the first time. Alternative causes for male-biased observations are considered. Using logistic regression analysis, a variety of sample-associated metadata are tested to identify possible explanatory variables for the sex ratios. The lone male model implies that males and females die in different locations, so a multivariate two-sample kernel test is implemented to assess possible spatial differences between males and females. Finally, there is a possibility that a male bias could be introduced during sample collection or curation, and such a collection bias is investigated by compiling sex ratios from databases of four large mammal collections (which correspond to individuals that have been hunted or trapped in recent centuries).

1.4.5 PP5mC: preprocessing hairpin-ligated bisulfite-treated DNA sequences

Hairpin bisulfite sequencing (HBS-seq) (Laird *et al.*, 2004; Zhao *et al.*, 2014) has the potential to produce genome-wide DNA methylation data for a range of aDNA samples, as it overcomes the low mappability of short three-state DNA sequences from traditional bisulfite sequencing. For HBS-seq, double-stranded libraries are constructed by ligating a hairpin (stem-loop) structure to one end of the DNA molecule, and a forked Illumina TruSeq adapter to the other. When a library molecule is then denatured during bisulfite treatment, the top and bottom strands remain connected because of the hairpin, and paired-end sequencing yields data for both strands. The original four-state nucleotide sequences can thus be reconstructed from HBS-seq reads, then mapped with regular DNA-seq alignment software. Only one software package currently exists for this purpose (HBS-tools) (Sun *et al.*, 2015). HBS-tools has several deficiencies, such as: erroneously trimming adapter sequences from the start of reads; ignoring potentially valuable information at the ends of reads derived from short molecules; and apparently unnecessary steps are taken during sequence reconstruction, making the software needlessly slow. In chapter 5, I present PP5mC, a new software pipeline for processing HBS-seq data. The pipeline reconstructs the maximum likelihood four-state nucleotide sequence from all available sequencing information. Using simulated data, PP5mC is compared against HBS-tools with respect to computational performance, the accuracy of reconstructed sequences, and the accuracy of inferred methylation levels. As a proof of concept for aDNA, PP5mC is used to process HBS-seq data from a 50 thousand-year-old bison skull.

1.5 References

Barros-Silva D, Marques C, Henrique R, Jerónimo C, Barros-Silva D, Marques CJ, Henrique R, & Jerónimo C (2018). Profiling DNA methylation based on next-generation sequencing approaches: new insights and clinical applications. *Genes*, **9(9)**:429. http://dx.doi.org/10.3390/genes9090429

Beaumont MA, Zhang W, & Balding DJ (2002). Approximate Bayesian computation in population genetics. *Genetics*, **162(4)**:2025–2035

Bhatia G, Patterson N, Sankararaman S, & Price AL (2013). Estimating and interpreting F_{st}: the impact of rare variants. *Genome Res*, **23(9)**:1514–1521. http://dx.doi.org/10.1101/gr.154831.113

Bibi F (2013). A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. *BMC Evol Biol*, **13**:166. http://dx.doi.org/10.1186/1471-2148-13-166

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, & Drummond AJ (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*, **10(4)**:e1003537. http://dx.doi.org/10.1371/journal.pcbi.1003537

Bover P, Llamas B, Thomson VA, Pons J, Cooper A, & Mitchell KJ (2018). Molecular resolution to a morphological controversy: the case of North American fossil muskoxen *Bootherium* and *Symbos*. *Mol Phylogenet Evol*, **129**:70–76. http://dx.doi.org/10.1016/j.ympev.2018.08.008

Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, Boisvert S, Chapman JA, Chapuis G, Chikhi R, *et al.* (2013). Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *Gigascience*, **2(1)**:10. http://dx.doi.org/10.1186/2047-217X-2-10

Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prüfer K, Meyer M, Krause J, Ronan MT, Lachmann M, *et al.* (2007). Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A*, **104(37)**:14616–14621. http://dx.doi.org/10.1073/pnas.0704665104

Briggs AW, 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 Res*, **38(6)**:e87. http://dx.doi.org/10.1093/nar/gkp1163

Briskine RV & Shimizu KK (2017). Positional bias in variant calls against draft reference assemblies. *BMC Genomics*, **18**. http://dx.doi.org/10.1186/s12864-017-3637-2

Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, Austin J, & Cooper A (2007). Novel high-resolution characterization of ancient DNA reveals C $>$ U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res*, **35(17)**:5717–5728. http://dx.doi.org/10.1093/nar/gkm588

Bunce M, Worthy TH, Ford T, Hoppitt W, Willerslev E, Drummond A, & Cooper A (2003). Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature*, **425(6954)**:172–175. http://dx.doi.org/10.1038/nature01871

Caliebe A, Nebel A, Makarewicz C, Krawczak M, & Krause-Kyora B (2017). Insights into early pig domestication provided by ancient DNA analysis. *Sci Rep*, **7**. http://dx.doi.org/10.1038/srep44550

Chen GK, Marjoram P, & Wall JD (2009). Fast and flexible simulation of DNA sequence data. *Genome Res*, **19(1)**:136–142. http://dx.doi.org/10.1101/gr.083634.108

Clark SJ, Harrison J, Paul CL, & Frommer M (1994). High sensitivity mapping of methylated cytosines. *Nucleic Acids Res*, **22(15)**:2990–2997

Cooper A & Poinar HN (2000). Ancient DNA: do it right or not at all. *Science*, **289(5482)**:1139–1139. http://dx.doi.org/10.1126/science.289.5482.1139b

Cooper A, Turney C, Hughen KA, Brook BW, McDonald HG, & Bradshaw CJA (2015). Abrupt warming events drove Late Pleistocene Holarctic megafaunal turnover. *Science*, **349(6248)**:602–606. http://dx.doi.org/10.1126/science.aac4315

Dabney J, Meyer M, & Pääbo S (2013). Ancient DNA damage. *Cold Spring Harb Perspect Biol*, **5(7)**:a012567. http://dx.doi.org/10.1101/cshperspect.a012567

Daly KG, Delser PM, Mullin VE, Scheu A, Mattiangeli V, Teasdale MD, Hare AJ, Burger J, Verdugo MP, Collins MJ, *et al.* (2018). Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science*, **361(6397)**:85–88. http://dx.doi.org/10.1126/science.aas9411

Dean MC (2016). Measures of maturation in early fossil hominins: events at the first transition from australopiths to early *Homo*. *Philos Trans R Soc Lond B Biol Sci*, **371(1698)**. http://dx.doi.org/10.1098/rstb.2015.0234

Dean MC & Liversidge HM (2015). Age estimation in fossil hominins: comparing dental development in early *Homo* with modern humans. *Ann Hum Biol*, **42(4)**:415–429. http://dx.doi.org/10.3109/03014460.2015.1046488

Denton JF, Lugo-Martinez J, Tucker AE, Schrider DR, Warren WC, & Hahn MW (2014). Extensive error in the number of genes inferred from draft genome assemblies. *PLoS Comput Biol*, **10(12)**:e1003998. http://dx.doi.org/10.1371/journal.pcbi.1003998

Donnelly P & Tavare S (1995). Coalescents and genealogical structure under neutrality. *Annu Rev Genet*, **29(1)**:401–421. http://dx.doi.org/10.1146/annurev.ge.29.120195.002153

Drummond AJ, Rambaut A, Shapiro B, & Pybus OG (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol*, **22(5)**:1185–1192. http://dx.doi.org/10.1093/molbev/msi103

Durbin R, Eddy SR, Krogh A, & Mitchison G (1998). *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. Cambridge University Press, Cambridge, UK : New York, 1 edition edition. ISBN 978-0-521-62971-3

Dutheil JY, Ganapathy G, Hobolth A, Mailund T, Uyenoyama MK, & Schierup MH (2009). Ancestral population genomics: the coalescent hidden Markov model approach. *Genetics*, **183(1)**:259–274. http://dx.doi.org/10.1534/genetics.109.103010

Dymova MA, Zadorozhny AV, Mishukova OV, Khrapov EA, Druzhkova AS, Trifonov VA, Kichigin IG, Tishkin AA, Grushin SP, & Filipenko ML (2017). Mitochondrial DNA analysis of ancient sheep from Altai. *Anim Genet*, **48(5)**:615–618. http://dx.doi.org/10.1111/age.12569

Earl D, Bradnam K, John JS, Darling A, Lin D, Fass J, Yu HOK, Buffalo V, Zerbino DR, Diekhans M, *et al.* (2011). Assemblathon 1: a competitive assessment of de novo short read assembly methods. *Genome Res*, **21(12)**:2224–2241. http://dx.doi.org/10.1101/gr.126599.111

Edwards JR, Yarychkivska O, Boulard M, & Bestor TH (2017). DNA methylation and DNA methyltransferases. *Epigenetics Chromatin*, **10(1)**:23. http://dx.doi.org/10.1186/s13072-017-0130-8

Ehrlich M, Norris KF, Wang RY, Kuo KC, & Gehrke CW (1986). DNA cytosine methylation and heat-induced deamination. *Biosci Rep*, **6(4)**:387–393. http://dx.doi.org/10.1007/BF01116426

Ekblom R & Wolf JBW (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evol Appl*, **7(9)**:1026–1042. http://dx.doi.org/10.1111/eva.12178

Feigin CY, Newton AH, Doronina L, Schmitz J, Hipsley CA, Mitchell KJ, Gower G, Llamas B, Soubrier J, Heider TN, *et al.* (2018). Genome of the Tasmanian tiger provides insights into the evolution and demography of an extinct marsupial carnivore. *Nat Ecol Evol*, **2(1)**:182–192. http://dx.doi.org/10.1038/s41559-017-0417-y

Frayer DW & Wolpoff MH (1985). Sexual dimorphism. *Annual Review of Anthropology*, **14(1)**:429–473. http://dx.doi.org/10.1146/annurev.an.14.100185.002241

Frederico LA, Kunkel TA, & Shaw BR (1990). A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry*, **29(10)**:2532–2537. http://dx.doi.org/10.1021/bi00462a015

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, & Paul CL (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A*, **89(5)**:1827–1831

Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, Furtwängler A, Haak W, Meyer M, Mittnik A, *et al.* (2016). The genetic history of Ice Age Europe. *Nature*, **534(7606)**:200–205. http://dx.doi.org/10.1038/nature17993

Glocke I & Meyer M (2017). Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. *Genome Res*, **27(7)**:1230–1237. http://dx.doi.org/10.1101/gr.219675.116

Gokhman D, Lavi E, Prufer K, Fraga MF, Riancho JA, Kelso J, Paabo S, Meshorer E, & Carmel L (2014). Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. *Science*, **344(6183)**:523–527. http://dx.doi.org/10.1126/science.1250368

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MHY, *et al.* (2010). A draft sequence of the Neandertal genome. *Science*, **328(5979)**:710–722. http://dx.doi.org/10.1126/science.1188021

Griffiths RC & Marjoram P (1997). An ancestral recombination graph. In *Progress in population genetics and human evolution*. Springer

Griffiths RC & Tavare S (1994). Sampling theory for neutral alleles in a varying environment. *Philos Trans R Soc Lond B Biol Sci*, **344(1310)**:403–410. http://dx.doi.org/10.1098/rstb.1994.0079

Groves C & Grubb P (2011). *Ungulate taxonomy*. Johns Hopkins University Press, Baltimore, Md. ISBN 978-1-4214-0093-8

Hassanin A, An J, Ropiquet A, Nguyen TT, & Couloux A (2013). Combining multiple autosomal introns for studying shallow phylogeny and taxonomy of Laurasiatherian mammals: application to the tribe Bovini (Cetartiodactyla, Bovidae). *Mol Phylogenet Evol*, **66(3)**:766–775. http://dx.doi.org/10.1016/j.ympev.2012.11.003

Hein J, Schierup MH, & Wiuf C (2005). *Gene genealogies, variation and evolution: a primer in coalescent theory*. Oxford University Press, Oxford ; New York, 1 edition edition. ISBN 978-0-19-852996-5

Hernando-Herraez I, Garcia-Perez R, Sharp AJ, & Marques-Bonet T (2015). DNA methylation: insights into human evolution. *PLoS Genet*, **11(12)**:e1005661. http://dx.doi.org/10.1371/journal.pgen.1005661

Hofreiter M, Jaenicke V, Serre D, Haeseler Av, & Pääbo S (2001). DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res*, **29(23)**:4793–4799. http://dx.doi.org/10.1093/nar/29.23.4793

Hofreiter M, Paijmans JLA, Goodchild H, Speller CF, Barlow A, Fortes GG, Thomas JA, Ludwig A, & Collins MJ (2015). The future of ancient DNA: technical advances and conceptual shifts. *BioEssays*, **37(3)**:284–293. http://dx.doi.org/10.1002/bies.201400160

Hudson RR (1983). Properties of a neutral allele model with intragenic recombination. *Theor Popul Biol*, **23(2)**:183–201. http://dx.doi.org/10.1016/0040-5809(83)90013-8

Hudson RR (1990). Gene geneologies and the coalescent process. In D Futuyma & J Antonovic, eds., *Oxford Surveys in Evolutionary Biology*, volume 7, pp. 1–44

Hudson RR (2002). Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics*, **18(2)**:337–338. http://dx.doi.org/10.1093/bioinformatics/18.2.337

Huynen L, Millar CD, Scofield RP, & Lambert DM (2003). Nuclear DNA sequences detect species limits in ancient moa. *Nature*, **425(6954)**:nature01838. http://dx.doi.org/10.1038/nature01838

Kelleher J, Etheridge AM, & McVean G (2016). Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Comput Biol*, **12(5)**:e1004842. http://dx.doi.org/10.1371/journal.pcbi.1004842

Kingman JFC (1982a). The coalescent. *Stochastic Processes and their Applications*, **13(3)**:235–248. http://dx.doi.org/10.1016/0304-4149(82)90011-4

Kingman JFC (1982b). On the genealogy of large populations. *Journal of Applied Probability*, **19**:27. http://dx.doi.org/10.2307/3213548

Kistler L, Ware R, Smith O, Collins M, & Allaby RG (2017). A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic Acids Res*, **45(11)**:6310–6320. http://dx.doi.org/10.1093/nar/gkx361

Krueger F, Kreck B, Franke A, & Andrews SR (2012). DNA methylome analysis using short bisulfite sequencing data. *Nat Meth*, **9(2)**:145–151. http://dx.doi.org/10.1038/nmeth.1828

Laird CD, Pleasant ND, Clark AD, Sneeden JL, Hassan KMA, Manley NC, Vary JC, Morgan T, Hansen RS, & Stöger R (2004). Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. *Proc Natl Acad Sci U S A*, **101(1)**:204–209. http://dx.doi.org/10.1073/pnas.2536758100

Li H & Durbin R (2011). Inference of human population history from individual whole-genome sequences. *Nature*, **475(7357)**:493–496. http://dx.doi.org/10.1038/nature10231

Lindahl T (1979). DNA glycosylases, endonucleases for apurinic/apyrimidinic sites, and base excision-repair. In WE Cohn, ed., *Progress in Nucleic Acid Research and Molecular Biology*, volume 22, pp. 135–192. Academic Press. http://dx.doi.org/10.1016/S0079-6603(08)60800-4

Lindahl T (1993). Instability and decay of the primary structure of DNA. *Nature*, **362(6422)**:709–715. http://dx.doi.org/10.1038/362709a0

Llamas B, Holland ML, Chen K, Cropley JE, Cooper A, & Suter CM (2012). High-resolution analysis of cytosine methylation in ancient DNA. *PLoS One*, **7(1)**:e30226. http://dx.doi.org/10.1371/journal.pone.0030226

Lyell C (1835). *Principles of Geology, Volume 1*. John Murray, Albemarle St, London, 4th edition. ISBN 978-0-226-49794-5

Mallet J (2005). Hybridization as an invasion of the genome. *Trends Ecol Evol*, **20(5)**:229–237. http://dx.doi.org/10.1016/j.tree.2005.02.010

Marjoram P & Wall JD (2006). Fast "coalescent" simulation. *BMC Genet*, **7**:16. http://dx.doi.org/10.1186/1471-2156-7-16

McVean GAT & Cardin NJ (2005). Approximating the coalescent with recombination. *Philos Trans R Soc Lond B Biol Sci*, **360(1459)**:1387–1393. http://dx.doi.org/10.1098/rstb.2005.1673

Mechta M, Ingerslev LR, Fabre O, Picard M, & Barrès R (2017). Evidence suggesting absence of mitochondrial DNA methylation. *Front Genet*, **8**. http://dx.doi.org/10.3389/fgene.2017.00166

Meyer M, Arsuaga JL, de Filippo C, Nagel S, Aximu-Petri A, Nickel B, Martínez I, Gracia A, de Castro JMB, Carbonell E, *et al.* (2016). Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature*. http://dx.doi.org/10.1038/nature17405

Minin VN, Bloomquist EW, & Suchard MA (2008). Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol*, **25(7)**:1459–1471. http://dx.doi.org/10.1093/molbev/msn090

Mittnik A, Wang CC, Svoboda J, & Krause J (2016). A molecular approach to the sexing of the triple burial at the Upper Paleolithic site of Dolní Věstonice. *PLoS One*, **11(10)**:e0163019. http://dx.doi.org/10.1371/journal.pone.0163019

Molak M, Lorenzen ED, Shapiro B, & Ho SYW (2013). Phylogenetic estimation of timescales using ancient DNA: the effects of temporal sampling scheme and uncertainty in sample ages. *Mol Biol Evol*, **30(2)**:253–262. http://dx.doi.org/10.1093/molbev/mss232

Moorjani P, Sankararaman S, Fu Q, Przeworski M, Patterson N, & Reich D (2016). A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. *Proc Natl Acad Sci U S A*, **113(20)**:5652–5657. http://dx.doi.org/10.1073/pnas.1514696113

Munson K, Clark J, Lamparska-Kupsik K, & Smith SS (2007). Recovery of bisulfite-converted genomic sequences in the methylation-sensitive QPCR. *Nucleic Acids Res*, **35(9)**:2893–2903. http://dx.doi.org/10.1093/nar/gkm055

Nagarajan N & Pop M (2013). Sequence assembly demystified. *Nat Rev Genet*, **14(3)**:157–167. http://dx.doi.org/10.1038/nrg3367

Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I, *et al.* (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, **499(7456)**:74–78. http://dx.doi.org/10.1038/nature12323

Ottoni C, Neer WV, Cupere BD, Daligault J, Guimaraes S, Peters J, Spassov N, Prendergast ME, Boivin N, Morales-Muñiz A, *et al.* (2017). The palaeogenetics of cat dispersal in the ancient world. *Nat Ecol Evol*, **1(7)**:0139. http://dx.doi.org/10.1038/s41559-017-0139

Overballe-Petersen S, Orlando L, & Willerslev E (2012). Next-generation sequencing offers new insights into DNA degradation. *Trends Biotechnol*, **30(7)**:364–368. http://dx.doi.org/10.1016/j.tibtech.2012.03.007

Pääbo S, Higuchi RG, & Wilson AC (1989). Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. *J Biol Chem*, **264(17)**:9709–9712

Pääbo S, Poinar H, Serre D, Jaenicke-Després V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, & Hofreiter M (2004). Genetic analyses from ancient DNA. *Annu Rev Genet*, **38(1)**:645–679. http://dx.doi.org/10.1146/annurev.genet.37.110801.143214

Palkopoulou E, Mallick S, Skoglund P, Enk J, Rohland N, Li H, Omrak A, Vartanyan S, Poinar H, Götherström A, *et al.* (2015). Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Current Biology*, **25(10)**:1395–1400. http://dx.doi.org/10.1016/j.cub.2015.04.007

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, & Reich D (2012). Ancient admixture in human history. *Genetics*, **192(3)**:1065–1093. http://dx.doi.org/10.1534/genetics.112.145037

Pečnerová P, Díez-del Molino D, Dussex N, Feuerborn T, von Seth J, van der Plicht J, Nikolskiy P, Tikhonov A, Vartanyan S, & Dalén L (2017). Genome-based sexing provides clues about behavior and social structure in the woolly mammoth. *Curr Biol*, **27(22)**:3505–3510.e3. http://dx.doi.org/10.1016/j.cub.2017.09.064

Pedersen JS, Valen E, Velazquez AMV, Parker BJ, Rasmussen M, Lindgreen S, Lilje B, Tobin DJ, Kelly TK, Vang S, *et al.* (2014). Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res*, **24(3)**:454–466. http://dx.doi.org/10.1101/gr.163592.113

Peter BM (2016). Admixture, population structure and F-statistics. *Genetics*. http://dx.doi.org/10.1534/genetics.115.183913

Prüfer K, Stenzel U, Hofreiter M, Pääbo S, Kelso J, & Green RE (2010). Computational challenges in the analysis of ancient DNA. *Genome Biol*, **11(5)**:R47. http://dx.doi.org/10.1186/gb-2010-11-5-r47

Pybus OG, Rambaut A, & Harvey PH (2000). An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics*, **155(3)**:1429–1437

Rabiner LR (1989). A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the Ieee*, pp. 257–286

Rabosky DL (2014). Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS One*, **9(2)**:e89543. http://dx.doi.org/10.1371/journal.pone.0089543

Rehg JA & Leigh SR (1999). Estimating sexual dimorphism and size differences in the fossil record: a test of methods. *Am J Phys Anthropol*, **110(1)**:95–104. http://dx.doi.org/10.1002/(SICI)1096-8644(199909)110:1<95::AID-AJPA8>3.0.CO;2-J

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PLF, *et al.* (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, **468(7327)**:1053–1060. http://dx.doi.org/10.1038/nature09710

Reich D, Thangaraj K, Patterson N, Price AL, & Singh L (2009). Reconstructing Indian population history. *Nature*, **461(7263)**:489–494. http://dx.doi.org/10.1038/nature08365

Robson SL & Wood B (2008). Hominin life history: reconstruction and evolution. *J Anat*, **212(4)**:394–425. http://dx.doi.org/10.1111/j.1469-7580.2008.00867.x

Rogers RL & Slatkin M (2017). Excess of genomic defects in a woolly mammoth on Wrangel island. *PLoS Genet*, **13(3)**:e1006601. http://dx.doi.org/10.1371/journal.pgen.1006601

Sankararaman S, Patterson N, Li H, Pääbo S, & Reich D (2012). The date of interbreeding between Neandertals and modern humans. *PLoS Genet*, **8(10)**:e1002947. http://dx.doi.org/10.1371/journal.pgen.1002947

Scheu A, Powell A, Bollongino R, Vigne JD, Tresset A, Çakırlar C, Benecke N, & Burger J (2015). The genetic prehistory of domesticated cattle from their origin to the spread across Europe. *BMC Genet*, **16(1)**:54. http://dx.doi.org/10.1186/s12863-015-0203-2

Schiffels S & Durbin R (2014). Inferring human population size and separation history from multiple genome sequences. *Nat Genet*, **46(8)**:919–925. http://dx.doi.org/10.1038/ng.3015

Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, Pybus OG, Gilbert MTP, Barnes I, Binladen J, *et al.* (2004). Rise and fall of the Beringian steppe bison. *Science*, **306(5701)**:1561–1565. http://dx.doi.org/10.1126/science.1101074

Shen JC, Rideout WM, & Jones PA (1994). The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucl Acids Res*, **22(6)**:972–976. http://dx.doi.org/10.1093/nar/22.6.972

Skoglund P, Ersmark E, Palkopoulou E, & Dalén L (2015). Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Current Biology*, **25(11)**:1515–1519. http://dx.doi.org/10.1016/j.cub.2015.04.019

Skoglund P, Northoff BH, Shunkov MV, Derevianko AP, Pääbo S, Krause J, & Jakobsson M (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc Natl Acad Sci U S A*, **111(6)**:2229–2234. http://dx.doi.org/10.1073/pnas.1318934111

Skoglund P, Storå J, Götherström A, & Jakobsson M (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J Archaeol Sci*, **40(12)**:4477–4482. http://dx.doi.org/10.1016/j.jas.2013.07.004

Sohn Ji & Nam JW (2018). The present and future of de novo whole-genome assembly. *Brief Bioinform*, **19(1)**:23–40. http://dx.doi.org/10.1093/bib/bbw096

Staab PR, Zhu S, Metzler D, & Lunter G (2015). scrm: efficiently simulating long sequences using the approximated coalescent with recombination. *Bioinformatics*, **31(10)**:1680–1682. http://dx.doi.org/10.1093/bioinformatics/btu861

Stewart KR, Veselovska L, & Kelsey G (2016). Establishment and functions of DNA methylation in the germline. *Epigenomics*. http://dx.doi.org/10.2217/epi-2016-0056

Sun Ma, Velmurugan KR, Keimig D, Xie H, Sun Ma, Velmurugan KR, Keimig D, & Xie H (2015). HBS-tools for hairpin bisulfite sequencing data processing and analysis. *Advances in Bioinformatics*, **2015**:e760423. http://dx.doi.org/10.1155/2015/760423

Tavaré S (1984). Line-of-descent and genealogical processes, and their applications in population genetics models. *Theor Popul Biol*, **26(2)**:119–164. http://dx.doi.org/10.1016/0040-5809(84)90027-3

Terhorst J, Kamm JA, & Song YS (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat Genet*, **49(2)**:303–309. http://dx.doi.org/10.1038/ng.3748

Verkaar ELC, Nijman IJ, Beeke M, Hanekamp E, & Lenstra JA (2004). Maternal and paternal lineages in cross-breeding bovine species. Has wisent a hybrid origin? *Mol Biol Evol*, **21(7)**:1165–1170. http://dx.doi.org/10.1093/molbev/msh064

Wiuf C & Hein J (1999). Recombination as a point process along sequences. *Theor Popul Biol*, **55(3)**:248–259. http://dx.doi.org/10.1006/tpbi.1998.1403

Zhang H, Lang Z, & Zhu JK (2018). Dynamics and function of DNA methylation in plants. *Nat Rev Mol Cell Biol*, **19(8)**:489–506. http://dx.doi.org/10.1038/s41580-018-0016-z

Zhang X, Goodsell J, & Norgren RB (2012). Limitations of the rhesus macaque draft genome assembly and annotation. *BMC Genomics*, **13**:206. http://dx.doi.org/10.1186/1471-2164-13-206

Zhao L, Sun Ma, Li Z, Bai X, Yu M, Wang M, Liang L, Shao X, Arnovitz S, Wang Q, *et al.* (2014). The dynamics of DNA methylation fidelity during mouse embryonic stem cell self-renewal and differentiation. *Genome Res*, p. gr.163147.113. http://dx.doi.org/10.1101/gr.163147.113

Chapter 2

Early cave art and ancient DNA record the origin of European bison

2.1 Authorship statement

Statement of Authorship

Title of Paper	Early cave art and ancient DNA record the origin of European bison
Publication Status	Published
Publication Details	Soubrier, J., et al. 2016. "Early Cave Art and Ancient DNA Record the Origin of European Bison." *Nature Communications* 7 (October): 13158. doi:10.1038/ncomms13158.

Principal Authors

Julien Soubrier			
Contribution to the Paper	Designed experiments. Performed bioinformatics analyses: processed and analysed NGS data, phylogenetics. Analysed and interpreted results. Wrote the paper with help from all co-authors.		
Overall percentage (%)	40		
Signature		Date	11.11.16

Graham Gower (Candidate)			
Contribution to the Paper	Designed experiments. Performed bioinformatics analyses: processed and analysed nuclear data (Paleomix, Principal Component Analysis, D and f statistics, Hypergeometric test, sensitivity analysis, co-contributor of ABC analysis). Analysed and interpreted results. Wrote the paper with help from all co-authors.		
Overall percentage (%)	40		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am one of two primary authors of this paper.		
Signature		Date	11.11.16

Co-Author Contributions

Ayla van Loenen (Candidate)			
Contribution to the Paper	Performed laboratory genetic analyses of mitochondrial and nuclear data that contributed towards the body of genetic data analysed in this paper, initial data processing steps of aforementioned genetic data, edited manuscript.		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis.		
Signature		Date	11.11.16

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Alan Cooper		
Contribution to the Paper	Designed experiments, provided samples, interpreted results. Wrote the paper with help from all co-authors.		
Signature		Date	11.11.16

Name of Co-Author	Bastien Llamas		
Contribution to the Paper	Designed experiments, laboratory work, analyses, interpreted results.		
Signature		Date	11.11.16

Name of Co-Author			
Contribution to the Paper			
Signature		Date	

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Johannes Krause		
Contribution to the Paper	Together with Frauke Langbein and Alexander Immel processed and provided sequence data from Bison from the Ukraine (provided to him by Marie-Anne Julien).		
Signature	[redacted]	Date	15.11.16

Name of Co-Author	Frauke Langbein		
Contribution to the Paper	Together with her supervisor Johannes Krause and Alexander Immel processed and provided sequence data from Bison from the Ukraine (provided by Marie-Anne Julien to Johannes Krause). Signed [redacted]half of her, since she has not been reachable.		
Signature	[redacted]	Date	15.11.16

Name of Co-Author	Alexander Immel		
Contribution to the Paper	Together with his supervisor Johannes Krause and Frauke Langbein processed and provided sequence data from Bison from the Ukraine (provided by Marie-Anne Julien to Johannes Krause).		
Signature	[redacted]	Date	15.11.16

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	David Chivall		
Contribution to the Paper	Radiocarbon dating of bison samples		
Signature		Date	18th November 2016

Name of Co-Author	Colin Groves		
Contribution to the Paper	Provided morphological and taxonomic background; suggested the link with cave art. Wrote the paper with help from all co-authors.		
Signature		Date	14/11/16

Name of Co-Author	Amelie Scheu		
Contribution to the Paper	Provided samples, background information and data.		
Signature		Date	18.11.2016

Name of Co-Author	Emilia Hofman-Kamińska		
Contribution to the Paper	Provided samples, interpretations of results and comments on the study.		
Signature		Date	18.11.2016

Name of Co-Author	Gennady Baryshnikov		
Contribution to the Paper	Bone material from field excavations.		
Signature		Date	14.11.2016

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate’s stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate’s stated contribution.

Name of Co-Author	Beth Shapiro		
Contribution to the Paper	Laboratory work, interpretation of results, comments on manuscript.		
Signature		Date	16 Nov 2016

Name of Co-Author	Federica Fontana		
Contribution to the Paper	Sample collecting. Data for sample contextualisation (Riparo Tagliente, IT)		
Signature		Date	19 November 2016

Name of Co-Author	Jared Decker		
Contribution to the Paper	Provided feedback on interpretation of the results. Along with Jeremy Taylor and Bob Schnabel, provide modern bison data.		
Signature		Date	14 November 2016

Name of Co-Author	Jerry Taylor		
Contribution to the Paper	Provided samples/data. Edited manuscript.		
Signature		Date	November 14, 2016.

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Johannes van der Plicht		
Contribution to the Paper	provided radiocarbon dates		
Signature		Date	14 november 2016

Name of Co-Author	Ludovic Orlando		
Contribution to the Paper	Provided feedback in data analyses and interpretation.		
Signature		Date	2016.12.01

Name of Co-Author	Małgorzata Tokarska		
Contribution to the Paper	Supplying samples, co-editing of the manuscript		
Signature		Date	14.11. 2016

Name of Co-Author	Michael Lee		
Contribution to the Paper	Assisted with phylogenetic analyses and data interpretation		
Signature		Date	14.11.16

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Pavel Kosintsev		
Contribution to the Paper	Provided samples, interpreted results		
Signature		Date	22.11.16

Name of Co-Author	Vladimir Doronichev		
Contribution to the Paper	Contributed samples and provided comments on this study		
Signature		Date	14.11.2016

Name of Co-Author	Liubov Golovanova		
Contribution to the Paper	Contributed samples and provided comments on this study		
Signature		Date	14.11.2016

Name of Co-Author	Ruth Bollongino		
Contribution to the Paper	Provided some mt-sequences and samples, discussed results and manuscript		
Signature		Date	16/11/2016

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Simon Ho		
Contribution to the Paper	Provided advice on phylogenetic analysis. Edited the draft manuscript.		
Signature		Date	14-Nov-16

Name of Co-Author	Tom Higham		
Contribution to the Paper	AMS radiocarbon dating of bone collagen extracts		
Signature		Date 14/11/2016	

Name of Co-Author	Wolfgang Haak		
Contribution to the Paper	Collected Bison samples from Russia, participated in early study design.		
Signature		Date	21/11/2016

2.2 Manuscript

nature COMMUNICATIONS

ARTICLE

Received 22 Apr 2016 | Accepted 9 Sep 2016 | Published 18 Oct 2016

DOI: 10.1038/ncomms13158

OPEN

Early cave art and ancient DNA record the origin of European bison

Julien Soubrier[1,*], Graham Gower[1,*], Kefei Chen[1], Stephen M. Richards[1], Bastien Llamas[1], Kieren J. Mitchell[1], Simon Y.W. Ho[2], Pavel Kosintsev[3], Michael S.Y. Lee[4,5], Gennady Baryshnikov[6], Ruth Bollongino[7], Pere Bover[1,8], Joachim Burger[7], David Chivall[9], Evelyne Crégut-Bonnoure[10,11], Jared E. Decker[12], Vladimir B. Doronichev[13], Katerina Douka[9], Damien A. Fordham[14], Federica Fontana[15], Carole Fritz[16], Jan Glimmerveen[17], Liubov V. Golovanova[13], Colin Groves[18], Antonio Guerreschi[15], Wolfgang Haak[1,19], Tom Higham[9], Emilia Hofman-Kamińska[20], Alexander Immel[19], Marie-Anne Julien[21,22], Johannes Krause[19], Oleksandra Krotova[23], Frauke Langbein[24], Greger Larson[25], Adam Rohrlach[26], Amelie Scheu[7], Robert D. Schnabel[12], Jeremy F. Taylor[12], Małgorzata Tokarska[20], Gilles Tosello[27], Johannes van der Plicht[28], Ayla van Loenen[1], Jean-Denis Vigne[29], Oliver Wooley[1], Ludovic Orlando[30,31], Rafał Kowalczyk[20], Beth Shapiro[32,33] & Alan Cooper[1]

The two living species of bison (European and American) are among the few terrestrial megafauna to have survived the late Pleistocene extinctions. Despite the extensive bovid fossil record in Eurasia, the evolutionary history of the European bison (or wisent, *Bison bonasus*) before the Holocene (<11.7 thousand years ago (kya)) remains a mystery. We use complete ancient mitochondrial genomes and genome-wide nuclear DNA surveys to reveal that the wisent is the product of hybridization between the extinct steppe bison (*Bison priscus*) and ancestors of modern cattle (aurochs, *Bos primigenius*) before 120 kya, and contains up to 10% aurochs genomic ancestry. Although undetected within the fossil record, ancestors of the wisent have alternated ecological dominance with steppe bison in association with major environmental shifts since at least 55 kya. Early cave artists recorded distinct morphological forms consistent with these replacement events, around the Last Glacial Maximum (LGM, ∼21–18 kya).

[1] Australian Centre for Ancient DNA, School of Biological Sciences, University of Adelaide, Adelaide, South Australia 5005, Australia. [2] School of Biological Sciences, University of Sydney, Sydney, New South Wales 2006, Australia. [3] Institute of Plant and Animal Ecology, Russian Academy of Sciences, 202 8 Marta Street, 620144 Ekaterinburg, Russia. [4] School of Biological Sciences, Flinders University, South Australia 5001, Australia. [5] Earth Sciences Section, South Australian Museum, North Terrace, Adelaide, South Australia 5000, Australia. [6] Zoological Institute RAS, Universitetskaya Naberezhnaya 1, 199034 St Petersburg, Russia. [7] Palaeogenetics Group, Institute of Anthropology, University of Mainz D-55128, Mainz, Germany. [8] Department of Biodiversity and Conservation, Institut Mediterrani d'Estudis Avançats (CSIC-UIB), Cr. Miquel Marquès 21, 07190 Esporles, Illes Balears. [9] Oxford Radiocarbon Accelerator Unit, Research Laboratory for Archaeology and the History of Art, University of Oxford, Oxford OX1 3QY, UK. [10] Museum Requien, 67 rue Joseph Vernet, 84000 Avignon, France. [11] Laboratoire TRACES UMR5608, Université Toulouse Jean Jaurès - Maison de la Recherche, 5 allée Antonio Machado, 31058 Toulouse, France. [12] Division of Animal Sciences, University of Missouri, Columbia, Missouri 65211, USA. [13] ANO Laboratory of Prehistory, 14 Linia 3e 11, 199034 St Petersburg, Russia. [14] Environment Institute and School of Biological Sciences, University of Adelaide, Adelaide, South Australia 5005, Australia. [15] Dipartimento di Studi Umanistici, Università degli Studi di Ferrara, 12 Via Paradiso, 44121 Ferrara, Italy. [16] CNRS, TRACES, UMR 5608 et CREAP, MSHS Toulouse, USR 3414, Maison de la Recherche, 5 allées Antonio Machado, 31058 Toulouse, France. [17] CERPOLEX/Mammuthus, Anna Paulownastraat 25A, NL-2518 BA Den Haag, The Netherlands. [18] School of Archaeology and Anthropology, Australian National University, Building 14, Canberra, Australian National University 0200, Australia. [19] Max Planck Institute for the Science of Human History, 07745 Jena, Germany. [20] Mammal Research Institute, Polish Academy of Sciences, Waszkiewicza 1c, 17-230 Białowieża, Poland. [21] Department of Archaeology, Centre for the Archaeology of Human Origins, University of Southampton, Avenue Campus, Southampton SO17 1BF, UK. [22] Unité Histoire naturelle de l'Homme préhistorique (UMR 7194), Sorbonne Universités, Muséum national d'Histoire narurelle, CNRS, 1 rue René Panhard, 75013 Paris, France. [23] Department of Stone Age, Institute of Archaeology, National Ukrainian Academy of Science, 04210 Kiev, Ukraine. [24] Institute for Archaeological Sciences, Archaeo and Palaeogenetics, University of Tübingen, 72070 Tübingen, Germany. [25] Palaeogenomics and Bio-Archaeology Research Network, Research Laboratory for Archaeology, Dyson Perrins Building, South Parks Road, Oxford OX1 3QY, UK. [26] School of Mathematical Sciences, The University of Adelaide, Adelaide, South Australia 5005, Australia. [27] Chercheur associé, CREAP, MSHS Toulouse, URS 3414, Maison de la Recherche, 5 allées Antonio Machado, 31058 Toulouse, France. [28] Centre for Isotope Research, Radiocarbon Laboratory, University of Groningen, Nijenborg 4, NI-9747 AG Groningen, The Netherlands. [29] Centre National de la Recherche Scientifique, Muséum National d'Histoire Naturelle, Sorbonne Universités, UMR7209, 'Archéozoologie, archéobotanique: sociétés, pratiques et environnements', CP56, 55 rue Buffon, 75005 Paris, France. [30] Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, ØsterVoldgade 5-7, Copenhagen 1350K, Denmark. [31] Université de Toulouse, University Paul Sabatier, Laboratoire AMIS, CNRS UMR 5288, 37 Allées Jules Guesde, Toulouse 31000, France. [32] Department of Ecology and Evolutionary Biology, University of California Santa Cruz, 1156 High Street, Santa Cruz, California 95064, USA. [33] UCSC Genomics Institute, University of California Santa Cruz, 1156 High Street, Santa Cruz, California 95064, USA. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to J.S. (email: julien.soubrier@adelaide.edu.au) or to A.C. (email: alan.cooper@adelaide.edu.au).

The extensive Late Pleistocene fossil record of bovids in Europe consists of two recognized forms: the aurochs (*Bos primigenius*), ancestor of modern cattle, and the mid/late Pleistocene 'steppe bison' (*Bison priscus*), which also ranged across Beringia as far as western Canada[1,2]. The European bison, or wisent (*Bison bonasus*), has no recognized Pleistocene fossil record and seems to suddenly appear in the early Holocene ($<$11.7 kya)[3,4], shortly after the disappearance of the steppe bison during the megafaunal extinctions of the Late Pleistocene[5–7]. The Holocene range of wisent included all lowlands of Europe, and several highland areas of eastern Europe (where it was termed the Caucasian form *B. bonasus caucasicus*) but range reduction and hunting by humans brought the species close to extinction, with modern populations descending from just 12 mostly Polish individuals that lived in the 1920s (refs 8,9). Nuclear DNA sequences and the morphology of the wisent show close similarities to American bison (*B. bison*), but wisent mitochondrial DNA (mtDNA) indicates a closer relationship with cattle. This suggests some form of introgression from cattle or a related *Bos* species[10–12], potentially associated with the recent extreme bottleneck event.

Both aurochs and bison feature heavily in Palaeolithic cave art, with 820 depictions displaying bison individuals ($\sim$21% of known cave ornamentation[13]). The diversity of bison representations has been explained as putative cultural and individual variations of style through time, since the steppe bison was assumed to be the only bison present in Late Paleolithic Europe[14–16]. However, two distinct morphological forms of bison (Fig. 1, Supplementary Information section) are clearly apparent in cave art: a long-horned form similar to modern American bison (which are thought to be descended from steppe bison), with very robust forequarters and oblique dorsal line, and a second form with thinner double-curved horns, smaller hump and more balanced body proportions, similar to wisent. The former is abundant in art older than the Last Glacial Maximum (LGM, $\sim$22–18 kya), while the latter dominates Magdalenian art ($\sim$17–12 kya, see Supplementary Information section). Similarly, two distinct morphological forms of Late Pleistocene bison have been reported from North Sea sediments[17].

To further examine the potential existence of a previously unrecognized fossil bison species within Europe, we sequenced ancient mtDNA and nuclear DNA from bones and teeth of 64 Late Pleistocene/Holocene bison specimens.

We reveal that the wisent lineage originated from hybridization between the aurochs and steppe bison, and this new form alternated ecologically with steppe bison throughout the Late Pleistocene and appears to have been recorded by early cave artists.

Results

New group of ancient European bison. The mtDNA sequences of 38 specimens, dated from $>$50 to 14 kya and ranging from the Caucasus, Urals, North Sea, France and Italy, formed a previously unrecognized genetic clade, hereafter referred to as CladeX, related to modern and historical wisent (including the Caucasian form; Fig. 2a,b). By using the radiocarbon-dated specimens to calibrate our phylogenetic estimate of the timescale, we inferred that the divergence between CladeX and modern wisent lineages occurred $\sim$120 (92–152) kya, likely during the last (Eemian) interglacial. Both these mitochondrial clades are more closely related to cattle than to bison, suggesting that they are descended from an ancient hybridization event that took place $>$120 kya (presumably between steppe bison and an ancestral form of aurochs, from which the mitochondrial lineage was acquired).

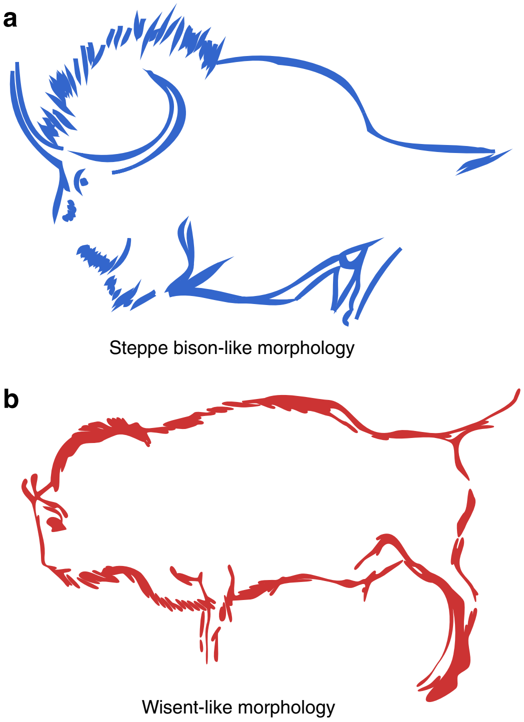


Figure 1 | Cave painting example of steppe bison-like and wisent-like morphs. (**a**) Reproduction from Lascaux cave (France), from the Solutrean or early Magdalenian period ($\sim$20,000 kya—picture adapted from ref. 53). (**b**) Reproduction from the Pergouset cave (France), from the Magdalenian period ($<$17,000 kya—picture adapted from ref. 54).

Hybrid origin of wisent and ancient European bison. To investigate the potential hybrid origins of wisent and CladeX, we used target enrichment and high-throughput methods to sequence $\sim$10,000 genome-wide bovine single-nucleotide polymorphisms (SNPs) from nine members of CladeX, an ancient ($>$55 kyr) and a historical (1911 AD) wisent specimen and two steppe bison (30 and $>$50 kyr). Principal Component Analysis (PCA) and phylogenetic analysis (Fig. 3 and Supplementary Fig. 10) of the nuclear data demonstrate that members of CladeX are closely related to the steppe bison. D-statistic[18] analyses confirm a closer affinity of both CladeX and the ancient wisent to steppe bison than to modern wisent (Fig. 3b), which is explicable because of rapid genetic drift during the severe bottleneck leading to modern wisent. Concordantly, our historical wisent sample (Caucasian, from 1911) displays a signal intermediate between modern wisent and both CladeX and steppe bison (Fig. 3b(3–5),c).

The nuclear and mitochondrial analyses together suggest that the common ancestor of the wisent and CladeX mitochondrial lineages originated from asymmetrical hybridization (or sustained introgression) between male steppe bison and female aurochs (see Supplementary Fig. 20). This scenario is consistent with the heavily polygynous mating system of most large bovids[19], and the observation that hybridization between either extant bison species and cattle usually results in F1 male infertility, consistent with Haldane's Rule of heterogametic crosses[20–22]. However, it is unclear whether hybridization took place only once or multiple times, and how and at what point after the initial hybridization event(s) the wisent–CladeX forms became distinct from the steppe bison.

To examine the extent of genetic isolation maintained through time by the hybrid forms (wisent and CladeX) from steppe bison, we characterized the genomic signals originating from either steppe bison or aurochs in the wisent and CladeX lineages.

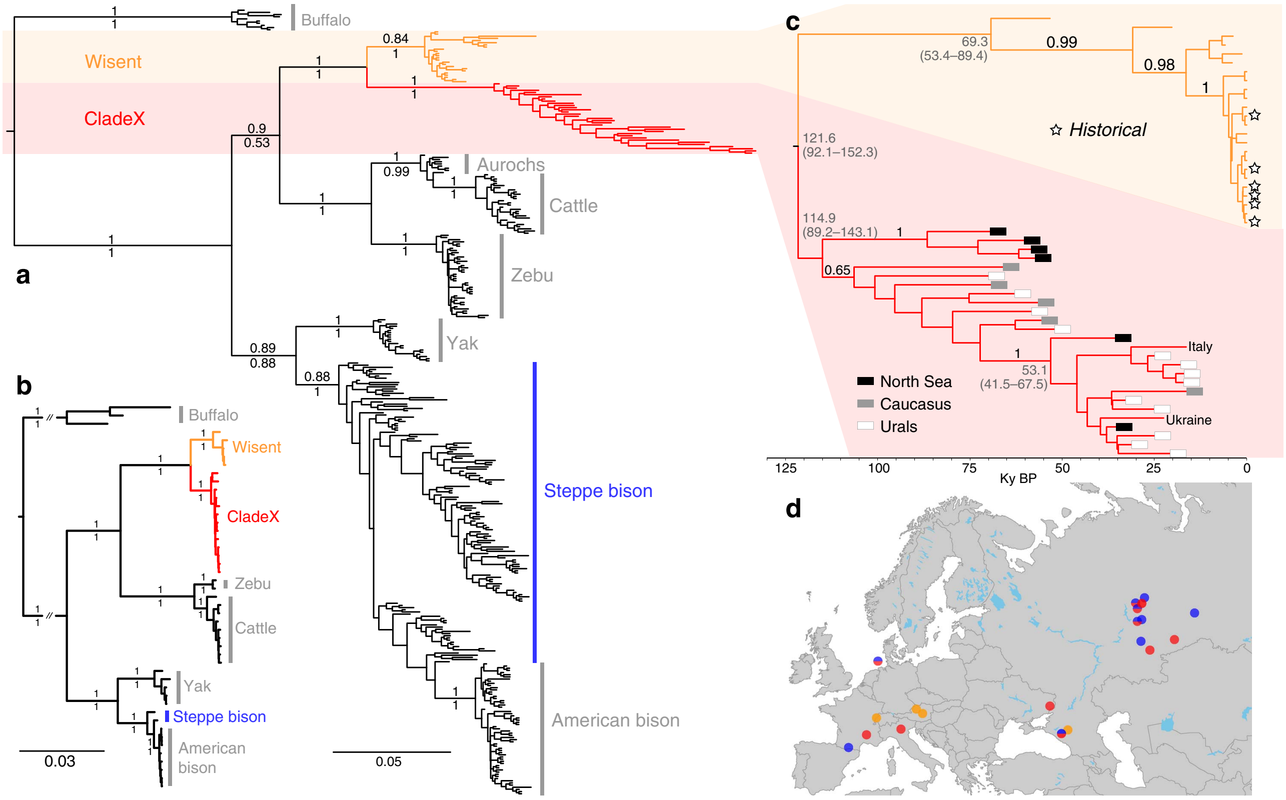


Figure 2 | Identification of CladeX. (**a**) Phylogenetic tree inferred from bovine mitochondrial control region sequences, showing the new clade of bison individuals. The positions of the newly sequenced individuals are marked in red for CladeX. (**b**) Bovine phylogeny estimated from whole-mitochondrial genome sequences, showing strong support for the grouping of wisent and CladeX with cattle (cow) and zebu. For both trees (**a**,**b**) numbers above branches represent the posterior probabilities from Bayesian inference, numbers below branches represent approximate likelihood ratio test support values from maximum-likelihood analysis and scale bars represent nucleotide substitutions per site from the Bayesian analysis. (**c**) Maximum-clade-credibility tree of CladeX and wisent estimated using Bayesian analysis and calibrated with radiocarbon dates associated with the sequenced bones. Dates of samples older than 50 kyr were estimated in the phylogenetic reconstruction. (**d**) Map showing all sampling locations, using the same colour code (red for CladeX, orange for wisent and blue for steppe bison).

Calculations of f_4 ratios[23] show the same high proportion of nuclear signal from steppe bison ($\geq$89.1%) and low proportion from aurochs ($\leq$10.9%) in both wisent and CladeX (Fig. 3d and Supplementary Table 6). Independent calculation of hybridization levels from ABC comparisons with simulated data also shows clear evidence of hybridization, with similar proportions of nuclear signal (97.2% probability that there is at least 1% aurochs ancestry and a 87.6% probability that there is at least 5% aurochs ancestry; see Supplementary Note 2 and Supplementary Tables 10 and 11). The agreement between these two methods is compelling evidence of hybridization. In addition, a greater number of derived alleles are common to both wisent and CladeX lineages (either from the imprint of steppe bison ancestry, aurochs ancestry, or from post-hybridization drift) than expected from multiple hybridization events (see Supplementary Note 2 and Supplementary Tables 8 and 9), implying that CladeX represents part of the Late Pleistocene wisent diversity. The age of the oldest genotyped specimens of CladeX (23 kyr) and wisent (>55 kyr) confirm that the initial hybridization event (or ultimate significant introgression of steppe bison) occurred before 55 kya. Together, the long-term stability of the nuclear and mitochondrial signal in wisent and CladeX indicates that the hybrid bison lineage maintained a marked degree of genetic isolation throughout the Late Pleistocene, consistent with the different morphologies observed in the North Sea specimens[17].

Hybrid and steppe Bison represent different ecological forms. The temporal distribution of genotyped individuals reveals that wisent mitochondrial lineages (including CladeX) are only observed before 50 kya and after 34 kya, when steppe bison appears to be largely absent from the European landscape (Fig. 4). The detailed records of the southern Ural sites allow the timing of the population replacements between steppe bison and wisent to be correlated with major palaeoenvironmental shifts, revealing that the wisent was associated with colder, more tundra-like landscapes and absence of a warm summer (Supplementary Fig. 22). Stable isotope data (∂^{13}C/∂^{15}N; Supplementary Fig. 23) and environment reconstructions show that wisent were present in a more diverse environment than steppe bison, with a more variable diet, suggesting that these two taxa occupied separate ecological niches.

Discussions

Contrary to previous palaeontological interpretations, the ancestors of modern wisent were present in Europe throughout the Late Pleistocene, and the two different bison morphs depicted in Paleolithic art suggest that early artists recorded the replacement of the steppe bison by the hybrid form (including CladeX) in Western Europe around the LGM. Two bison individuals have been genotyped from European caves during this period: a 19-kyr-old steppe bison from Southern France[24] and a

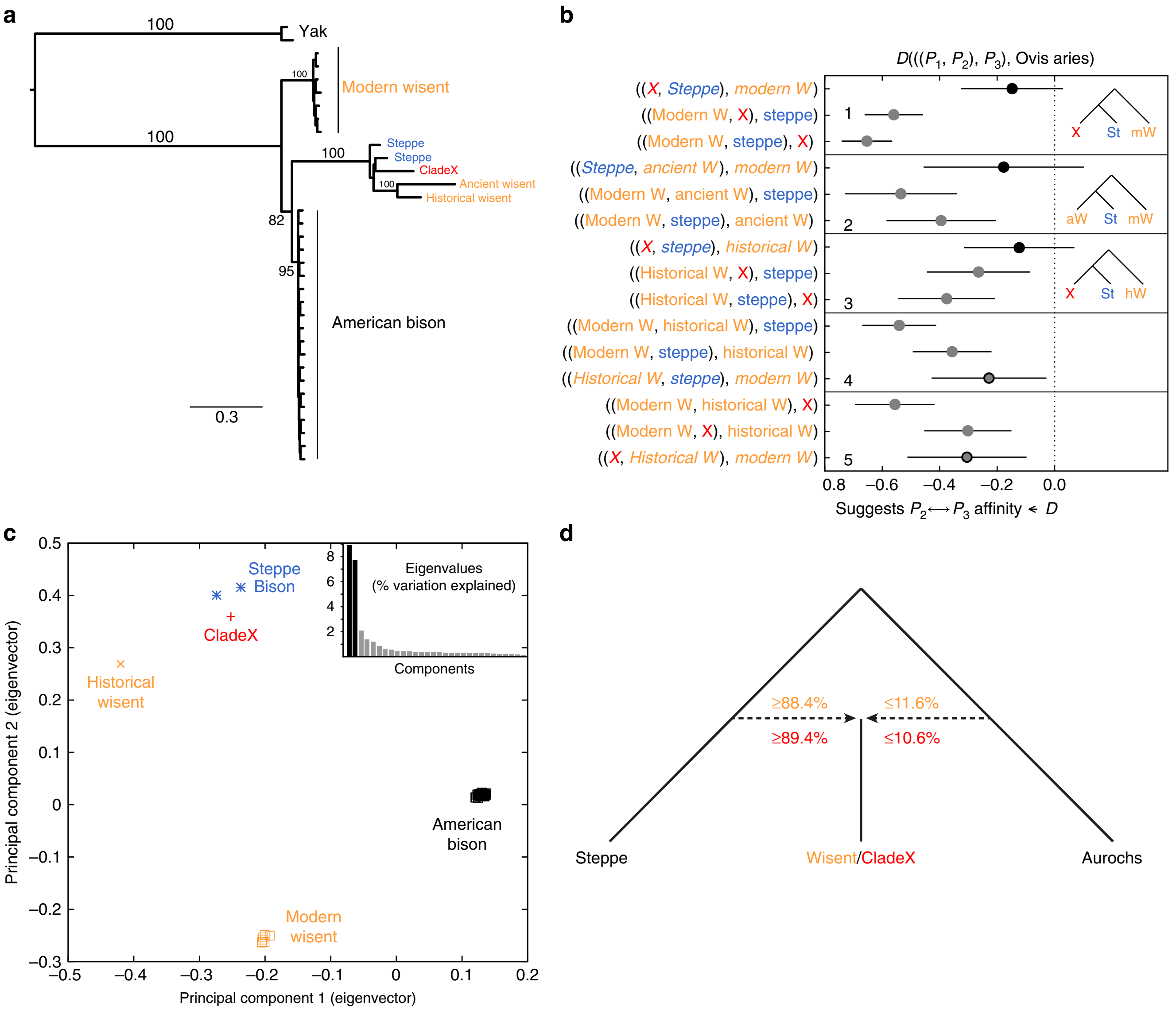


Figure 3 | Genome-wide data comparison of bison. (**a**) Maximum-likelihood phylogeny of modern and ancient bison from ~10,000 genome-wide nuclear sites, showing the close relationship between CladeX and steppe bison. However, a bifurcating phylogeny is not capable of displaying the complex relationships between these taxa (see Supplementary Fig. 8). Numbers above branches represent bootstrap values. (**b**) D-statistics from the same ~10,000 nuclear sites, using sheep as outgroup. For three bison populations, assuming two bifurcations and no hybridizations, three possible phylogenetic topologies can be evaluated using D-statistics, with the value closest to 0, indicating which topology is the most parsimonious. The topology being tested is shown on the vertical axis. Error bars are three s.e.'s (from block jackknife) either side of the data point. Data points that are significantly different from zero are shown in grey. The data point representing the topology in **a**, among a set of three possible topologies, is shown with a black outline. (**c**) Principal Component Analysis of ~10,000 genome-wide nuclear sites (ancient wisent not included due to the sensitivity of PCA to missing data, see Supplementary Fig. 10). (**d**) Proportion of steppe bison and aurochs ancestry in both wisent and CladeX lineages, calculated with f_4 ratios.

16-kyr-old wisent (CladeX) from Northern Italy (present study), corresponding to the timing of the morphological transition from steppe bison-like to wisent-like morphotypes apparent in cave art.

Combined evidence from genomic data, paleoenvironmental reconstructions and cave paintings strongly suggest that the hybridization of steppe bison with an ancient aurochs lineage during the late Pleistocene led to a morphologically and ecologically distinct form, which maintained its integrity and survived environmental changes on the European landscape until modern times. Although further analyses of deeper ancient genome sequencing will be necessary to characterize the phenotypic consequences of such hybridization, this adds to recent evidence of the importance of hybridization as a mechanism for speciation and adaptation of mammals[25–29] as is already accepted for plants. Lastly, the paraphyly of *Bos* with respect to *Bison*, and the evidence of meaningful hybridization between aurochs and bison, support the argument that both groups should be combined under the genus *Bos*[12,19,30].

Methods

Ancient DNA samples description and processing. Samples from a total of 87 putative bison bones were collected from three regions across Europe: Urals, Caucasus and Western Europe (Supplementary Data 1).

Dating of 45 samples that yielded DNA was performed at the Oxford Radiocarbon Accelerator Unit of the University of Oxford (OxA numbers), and the Ångström Laboratory of the University of Uppsala, Sweden, for the Swiss sample (Ua-42583). The calibration of radiocarbon dates was performed using OxCal v4.1 with the IntCal13 curve[31] (Supplementary Data 1).

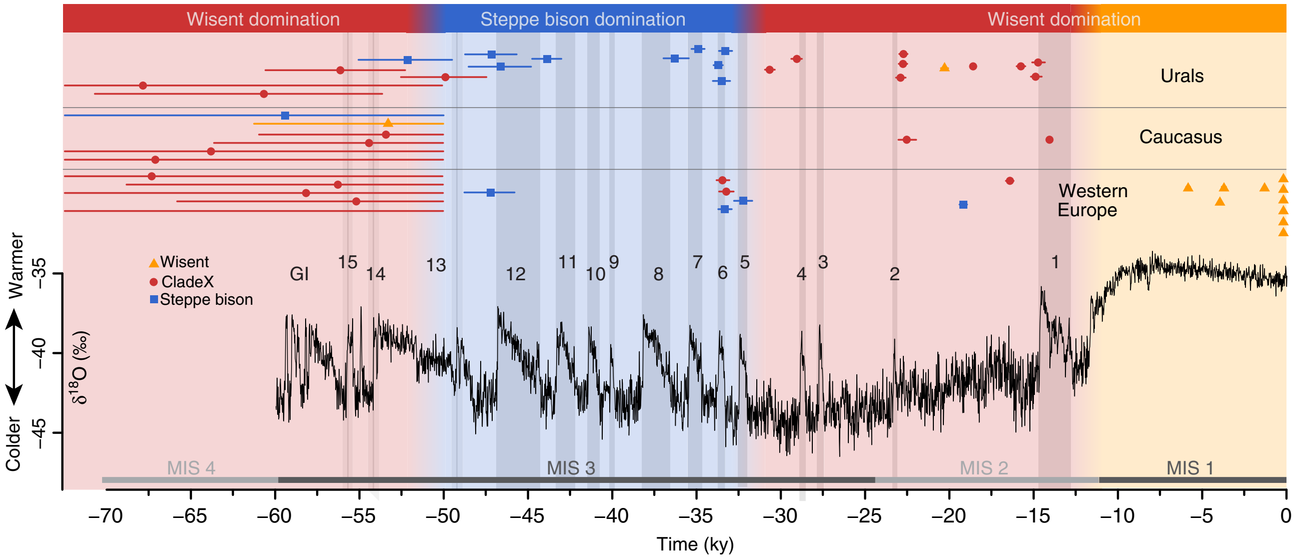


Figure 4 | Temporal and geographical distribution of bison in Europe. Individual calibrated AMS dates from the present study and published data are plotted on top of the NGRIP δ^{18}O record[55]. Age ranges for infinite AMS dates are from molecular clock estimates (Fig. 2c). Greenland interstadials (GIs) are numbered in black and marine isotope stages (MIS) in grey.

All ancient DNA work was conducted in clean-room facilities at the University of Adelaide's Australian Centre for Ancient DNA, Australia (ACAD), and at the University of Tuebingen, Germany (UT) following the published guidelines[32].

Samples were extracted using either phenol–chloroform[33] or silica-based methods[34,35] (see Supplementary Data 1).

Mitochondrial control region sequences (>400 bp) were successfully amplified from 65 out of 87 analysed samples in one or up to four overlapping fragments, depending on DNA preservation[33]. To provide deeper phylogenetic resolution and further examine the apparent close relationship between *Bos* and wisent mitochondria, whole-mitogenome sequences of 13 CladeX specimens, as well as one ancient wisent, one historical wisent and one steppe bison were generated using hybridization capture with either custom-made[36,37] (see Supplementary Note 1 for details).

In addition, genome-wide nuclear locus capture was attempted on DNA extracts from 13 bison samples (see Supplementary Table 2), using either an ~40,000 or an ~10,000 set of probes (as described in Supplementary Note 1). All targeted loci were part of the BovineSNP50 v2 BeadChip (Illumina) bovine SNP loci used in a previous phylogenetic study[38]. Ultimately, only the 9,908 loci common to both sets were used for comparative analysis.

Genetic data analysis. *Data processing*. Next-generation sequencing data were obtained from enriched libraries using paired-end reactions on Illumina HiSeq or MiSeq machines, and processed using the pipeline Paleomix v1.0.1 (ref. 39). AdapterRemoval v2 (ref. 40) was used to trim adapter sequences, merge the paired reads and eliminate all reads shorter than 25 bp. BWA v0.6.2 was then used to map the processed reads to either the reference mitochondrial genome of the wisent (NC_014044), American bison (NC_012346—only for the steppe bison A3133) or the *Bos taurus* genome reference UMD 3.1 (ref. 41). Minimum mapping quality was set at 25, seeding was disabled and the maximum number of gap opens was set to 2 (see Supplementary Tables 2 and 3).

MapDamage v2 (ref. 42) was used to check that the expected contextual mapping and damage patterns were observed for each library, depending on the enzymatic treatment used during library preparation (see Supplementary Table 3 and Supplementary Figs 1–3 for examples), and to rescale base qualities accordingly.

Phylogenetic analyses. The 60 newly sequenced bovine mitochondrial regions (Supplementary Data 1) were aligned with 302 published sequences (Supplementary Table 4), and a phylogenetic tree was inferred using both maximum-likelihood (PhyML v3 (ref. 43)) and Bayesian (MrBayes v3.2.3 (ref. 44)) methods (Fig. 2a and Supplementary Fig. 4). The same methods were used to obtain the whole-mitogenome phylogeny of 16 newly sequenced bison (Supplementary Data 1) aligned with 31 published sequences (Fig. 2b and Supplementary Fig. 5). To estimate the evolutionary timescale, we used the programme BEAST v1.8.1 (ref. 45) to conduct a Bayesian phylogenetic analysis of all radiocarbon-dated samples from CladeX and wisent (Fig. 1c), using the mean calibrated radiocarbon dates as calibration points. All parameters showed sufficient sampling after 5,000,000 steps, and a date-randomization test supported that the temporal signal from the radiocarbon dates associated with the ancient sequences was sufficient to calibrate the analysis[46] (Supplementary Fig. 6).

Finally, phylogenetic trees were inferred from nuclear loci data using RAxML v8.1.21 (ref. 47), first from published data of modern bovine representatives[38] (using sheep as an outgroup; Supplementary Fig. 7) and then including five ancient samples (two ancient steppe bison, an ancient wisent, a historical wisent and a CladeX bison; Fig. 2a), which had the highest number of nuclear loci successfully called among the ~10 k nuclear bovine SNPs targeted with hybridization capture (see Supplementary Fig. 8).

Principal Component Analysis. PCA (Fig. 3a and Supplementary Fig. 10) was performed using EIGENSOFT version 6.0.1 (ref. 48). In Fig. 3a, CladeX sample A006 was used as the representative of CladeX, as this sample contained the most complete set of nuclear loci called at the bovine SNP loci (see Supplementary Table 2). Other CladeX individuals, as well as ancient wisent, cluster towards coordinates 0.0, 0.0 (see Supplementary Fig. 10), because of missing data.

D and f statistics. Support for the bifurcating nuclear tree (Fig. 2a) was further tested using D-statistics calculated using ADMIXTOOLS version 3.0, git~3065acc5 (ref. 23). Sensitivity to factors like sampling bias, depth of coverage, choice of outgroup, heterozygosity (by haploidization) and missing data did not have notable influences on the outcome (Supplementary Figs 12–15).

The proportion of the wisent's ancestry differentially attributable to the steppe bison, and the aurochs was estimated with AdmixTools using an f_4 ratio[23] with sheep (*Ovis aries*) as the outgroup (Supplementary Figs S16, S17 and 3D). Again, the test was shown to be robust to haploidization.

Finally, to test whether the wisent lineages (including CladeX) have a common hybrid ancestry, or whether multiple independent hybridization events gave rise to distinct wisent lineages (Supplementary Fig. 18), we identify nuclear loci that have an ancestral state in the aurochs lineage, but a derived state in the steppe bison lineage (see Supplementary Note 2 section 'Identification of Derived Alleles'). Hypergeometric tests (Supplementary Tables 8 and 9) showed strong support for an ancestral hybridization event occurring before the divergence of the wisent lineages.

Testing admixture using ABC and simulated data. Admixture proportions were also independently tested using simulated data and an ABC approach. Nuclear genetic count data were simulated for two species trees (as described in Supplementary Fig. 19 and Supplementary Note 2 section) by drawing samples from two Multinomial distributions, where for tree topology X_1, $n^{X_1} \sim \mathrm{Mult}(N, p^{T,X_1})$, and for tree topology $X_2, n^{X_2} \sim \mathrm{Mult}(N, p^{T,X_2})$. The linear combination of these counts was then considered.

ABC was performed using the R package 'abc', with a ridge regression correction for comparison of the simulated and observed data using the 'abc' function[49]. The distance between the observed and simulated data sets is calculated as the Euclidean distance in a three-dimensional space, corrected for the within dimension variability. A tolerance $\epsilon = 0.005$ was chosen so that the closest $\ell \times \epsilon$ simulated data sets are retained. For each analysis we had $\ell = 100,000$, resulting in 500 posterior samples.

We performed leave-one-out cross-validation using the function 'cv4abc' on $\ell' = 250$ randomly selected simulations, and report the prediction error, calculated as

$$E_{\mathrm{pred}} = \frac{\sum_{i=1}^{\ell'} (\hat{\gamma}_i - \gamma_i)^2}{\mathrm{Var}(\gamma_i)}$$

for each analysis. At most, the prediction error was 0.5111 s.d.'s away from zero, and so we observe that the analysis has performed well (see Supplementary Table 10).

Palaeoenvironment reconstruction and stable isotope analyses. The Urals material has the most complete sampling through time (Fig. 4 and Supplementary Fig. 22), allowing us to contrast reconstructed paleoenvironmental proxies for the region (see Supplementary Note 3). Paleovegetation types were inferred for a convex hull of the Ural study region based on geo-referenced site locations for all genotyped ancient samples (Supplementary Fig. 21). Global maps of BIOME4 plant functional types[50] were accessed for 2,000-year time steps throughout the period from 70,000 years ago to the present day, with a 1° × 1° latitude/longitude grid cell resolution. We also generated estimates of the annual mean daily temperature and Köppen–Geiger climate classification[51] using the Hadley Centre Climate model (HadCM3)[52]. Finally, stable isotope values (δ13C and δ15N) obtained for all the genotyped bison individuals from the Ural region were compared between steppe bison and wisent (Supplementary Fig. 23).

Cave paintings. Two consistent morphological types can be distinguished within the diversity of bison representations (see Fig. 1 and Supplementary Figs 24–27). The first type, abundant before the LGM, is characterized by long horns (with one curve), a very oblique dorsal line and a very robust front part of the body (solid shoulders versus hindquarters), all traits similar to the modern American bison. The second type, dominating the more recent paintings between 18 and 15 kya, displays thinner sinuous horns (often with a double curve), a smaller hump and more balanced dimensions between the front and rear of the body, similar to modern wisent and to some extent aurochsen (see also Supplementary Note 4). The coincident morphological and genetic replacement indicate that variation in bison representations in Paleolithic art does not simply represent stylistic evolution, but actually reflects the different forms of bison genotyped in this study (that is, pre and post-hybridization) through time.

Data Availability. All newly sequenced mitochondrial control regions are deposited at the European Nucleotide Archive under the following accession numbers (LT599586–645) and all complete mitochondrial genomes at GenBank (KX592174–89). The BEAST input file (XML) is available as Supplementary Data set 2, the MrBayes input file (Nexus), including all whole-mitochondrial genomes, as Supplementary Data set 3 and the nuclear SNPs as Supplementary Data set 4 (VCF format). All other data are included in the Supplementary Material or available upon request to the corresponding authors.

References

1. Kurtén, B. *Pleistocene Mammals of Europe* (1968).
2. Geist, V. The relation of social evolution and dispersal in ungulates during the Pleistocene, with emphasis on the old world deer and the genus Bison. *Quat. Res.* **1,** 285–315 (1971).
3. Benecke, N. The holocene distribution of European bison: the archaeozoological record. *Munibe Antropol. Arkeol.* **57,** 421–428 (2005).
4. Bocherens, H., Hofman-Kamińska, E., Drucker, D. G., Schmölcke, U. & Kowalczyk, R. European Bison as a refugee species? Evidence from isotopic data on early holocene bison and other large herbivores in Northern Europe. *PLoS ONE* **10,** e0115090 (2015).
5. Stuart, A. J. Mammalian extinctions in the late Pleistocene of Northern Eurasia and North America. *Biol. Rev.* **66,** 453–562 (1991).
6. Lorenzen, E. D. *et al.* Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature* **479,** 359–364 (2011).
7. Cooper, A. *et al.* Abrupt warming events drove Late Pleistocene Holarctic megafaunal turnover. *Science* **349,** 602–606 (2015).
8. Slatis, H. M. An analysis of inbreeding in the European Bison. *Genetics* **45,** 275–287 (1960).
9. Tokarska, M., Pertoldi, C., Kowalczyk, R. & Perzanowski, K. Genetic status of the European bison Bison bonasus after extinction in the wild and subsequent recovery. *Mammal Rev.* **41,** 151–162 (2011).
10. Verkaar, E. L. C., Nijman, I. J., Beeke, M., Hanekamp, E. & Lenstra, J. A. Maternal and paternal lineages in cross-breeding bovine species. Has wisent a hybrid origin? *Mol. Biol. Evol.* **21,** 1165–1170 (2004).
11. Hassanin, A. *et al.* Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *C. R. Biol.* **335,** 32–50 (2012).
12. Bibi, F. A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. *BMC Evol. Biol.* **13,** 166 (2013).
13. Sauvet, G. & Wlodarczyk, L'art Pariétal, miroir des sociétés paléolithiques. *Zephyrus Rev. Prehist. Arqueol.* **53,** 217–240 (2000).
14. Breuil, H. *Quatre Cents Siècles d'art Pariétal; Les Cavernes Ornées de l'âge du Renne* (Centre d'études et de documentation préhistoriques, 1952).
15. Leroi-Gourhan, A. *Préhistoire de l'art Occidental* (1965).
16. Petrognani, S. *De Chauvet à Lascaux: l'art des Cavernes, Reflet de sociétés Préhistoriques en Mutation* (Editions Errance, 2013).
17. Drees, M. & Post, K. Bison bonasus from the North Sea, the Netherlands. *Cranium* **24,** 48–52 (2007).
18. Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28,** 2239–2252 (2011).
19. Groves, C. Current taxonomy and diversity of crown ruminants above the species level. *Zitteliana B* **32,** 5–14 (2014).
20. Haldane, J. B. S. Sex ratio and unisexual sterility in hybrid animals. *J. Genet.* **12,** 101–109 (1922).
21. Hedrick, P. W. Conservation genetics and North American bison (*Bison bison*). *J. Hered.* **100,** 411–420 (2009).
22. Derr, J. N. *et al.* Phenotypic effects of cattle mitochondrial DNA in American bison. *Conserv. Biol.* **26,** 1130–1136 (2012).
23. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192,** 1065–1093 (2012).
24. Marsolier-Kergoat, M.-C. *et al.* Hunting the extinct steppe bison (*Bison priscus*) mitochondrial genome in the Trois-Frères Paleolithic Painted Cave. *PLoS ONE* **10,** e0128267 (2015).
25. Ropiquet, A. & Hassanin, A. Hybrid origin of the Pliocene ancestor of wild goats. *Mol. Phylogenet. Evol.* **41,** 395–404 (2006).
26. Larsen, P. A., Marchán-Rivadeneira, M. R. & Baker, R. J. Natural hybridization generates mammalian lineage with species characteristics. *Proc. Natl Acad. Sci. USA.* **107,** 11447–11452 (2010).
27. Song, Y. *et al.* Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Curr. Biol.* **21,** 1296–1301 (2011).
28. Amaral, A. R., Lovewell, G., Coelho, M. M., Amato, G. & Rosenbaum, H. C. Hybrid speciation in a marine mammal: the clymene dolphin (*Stenella clymene*). *PLoS ONE* **9,** e83645 (2014).
29. Lister, A. M. & Sher, A. V. Evolution and dispersal of mammoths across the Northern Hemisphere. *Science* **350,** 805–809 (2015).
30. Groves, C. & Grubb, P. *Ungulate Taxonomy* (Johns Hopkins University Press, 2011).
31. Reimer, P. J. *et al.* IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55,** 1869–1887 (2013).
32. Willerslev, E. & Cooper, A. Ancient DNA. *Proc. R Soc. B Biol. Sci.* **272,** 3–16 (2005).
33. Shapiro, B. *et al.* Rise and fall of the Beringian steppe bison. *Science* **306,** 1561–1565 (2004).
34. Brotherton, P. *et al.* Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat. Commun.* **4,** 1764 (2013).
35. Rohland, N. & Hofreiter, M. Ancient DNA extraction from bones and teeth. *Nat. Protoc.* **2,** 1756–1762 (2007).
36. Llamas, B. *et al.* Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Sci. Adv.* **2,** e1501385 (2016).
37. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. *PLoS ONE* **5,** e14004 (2010).
38. Decker, J. E. *et al.* Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *Proc. Natl Acad. Sci. USA* **106,** 18644–18649 (2009).
39. Schubert, M. *et al.* Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Protoc.* **9,** 1056–1082 (2014).
40. Lindgreen, S. AdapterRemoval easy cleaning of next generation sequencing reads. *BMC Res. Notes* **5,** 337 (2012).
41. Zimin, A. V. *et al.* A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* **10,** R42 (2009).
42. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29,** 1682–1684 (2013).
43. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59,** 307–321 (2010).
44. Ronquist, F. *et al.* MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61,** 539–542 (2012).
45. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7,** 214 (2007).
46. Ho, S. Y. W. *et al.* Bayesian estimation of substitution rates from ancient DNA sequences with low information content. *Syst. Biol.* **60,** 366–375 (2011).
47. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22,** 2688–2690 (2006).
48. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2,** e190 (2006).
49. Csilléry, K., François, O. & Blum, M. G.B. abc: an R package for approximate Bayesian computation (ABC). *Methods Ecol. Evol.* **3,** 475–479 (2012).
50. Kaplan, J. O. *Geophysical Applications of Vegetation Modeling* (Lund University, 2001).
51. Peel, M. C., Finlayson, B. L. & McMahon, T. A. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* **11,** 1633–1644 (2007).

52. Singarayer, J. S. & Valdes, P. J. High-latitude climate sensitivity to ice-sheet forcing over the last 120 kyr. *Quat. Sci. Rev.* **29,** 43–55 (2010).
53. Leroi-Gourhan, A. & Allain, J. *Lascaux Inconnu* (CNRS, 1979).
54. Lorblanchet, M. *La Grotte Ornée de Pergouset (Saint-Géry, Lot). Un Sanctuaire Secret Paléolithique* (Maison des Sciences de l'Homme, 2001).
55. Wolff, E. W., Chappellaz, J., Blunier, T., Rasmussen, S. O. & Svensson, A. Millennial-scale variability during the last glacial: the ice core record. *Quat. Sci. Rev.* **29,** 2828–2838 (2010).

Acknowledgements

We are grateful to A.A. Krotova (Institute of Archeology Ukrainian Academy of Sciences), K. Wysocka (Vinnytsia Regional Local History Museum), M. Blant (Swiss Institute for Speleology and Karst Studies), G. Zazula and E. Hall (Yukon Palaeontology Program), C. Lefèvre (Muséum National d'Histoire Naturelle), M. Leonardi (Natural History Museum of Denmark), J.P. Brugal (Laboratoire méditerranéen de préhistoire Europe Afrique and Musée d'Orgnac), the Natural History Museum of Vienna and the Paleontological Institute of Moscow for providing access to samples.We thank A. Lister, K. Helgen and J. Tuke for their comments on the study, as well as A. Vorobiev, Y. Clément and M.E.H. Jones for their help in the project. This research was supported by the Australian Research Council, the European Commission (PIRSES-GA-2009-247652—BIOGEAST), the Polish National Science Centre (N N304 301940 and 2013/11/B/NZ8/00914), the Danish National Research Foundation (DNRF94), the Marie Curie International Outgoing Fellowship (7th European Community Framework Program—MEDITADNA, PIOF-GA-2011-300854, FP7-PEOPLE) and the Russian Foundation for Basic Research (N 15-04-03882).

Author contributions

J.S., G.G., K.C., S.M.R., B.L., K.J.M., S.Y.W.H., M.S.Y.L., B.S., A.R. and A.C. designed experiments; P.K., G.B., R.B., J.B., E.C.-B., V.B.D., F.F., J.G., L.V.G., A.G., W.H., M.-A.J., E.H.-K., O.K., F.L., G.L., A.S., M.T., J.v.d.P., J.-D.V., L.O. and R.K. provided samples, interpretations of results and comments on the study; K.C., S.M.R., B.L., P.B., W.H., J.K., A.I., A.v.L. and B.S. performed laboratory genetic analyses; D.C., K.D., T.H. and J.v.d.P. performed radiocarbon-dating analyses; J.S., G.G., S.Y.W.H., M.S.Y.L., J.E.D., R.D.S., A.R. and O.W. performed bioinformatic analyses; P.K. and D.A.F. performed palaeoenvironmental analyses; C.F. and G.T. provided data and interpretation of cave art; J.S., G.G., B.L., K.J.M., M.S.Y.L., J.E.D., C.G., W.H., J.F.T., L.O., R.K. and A.C. analysed the results; and A.C. and J.S. wrote the paper with help from all co-authors.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

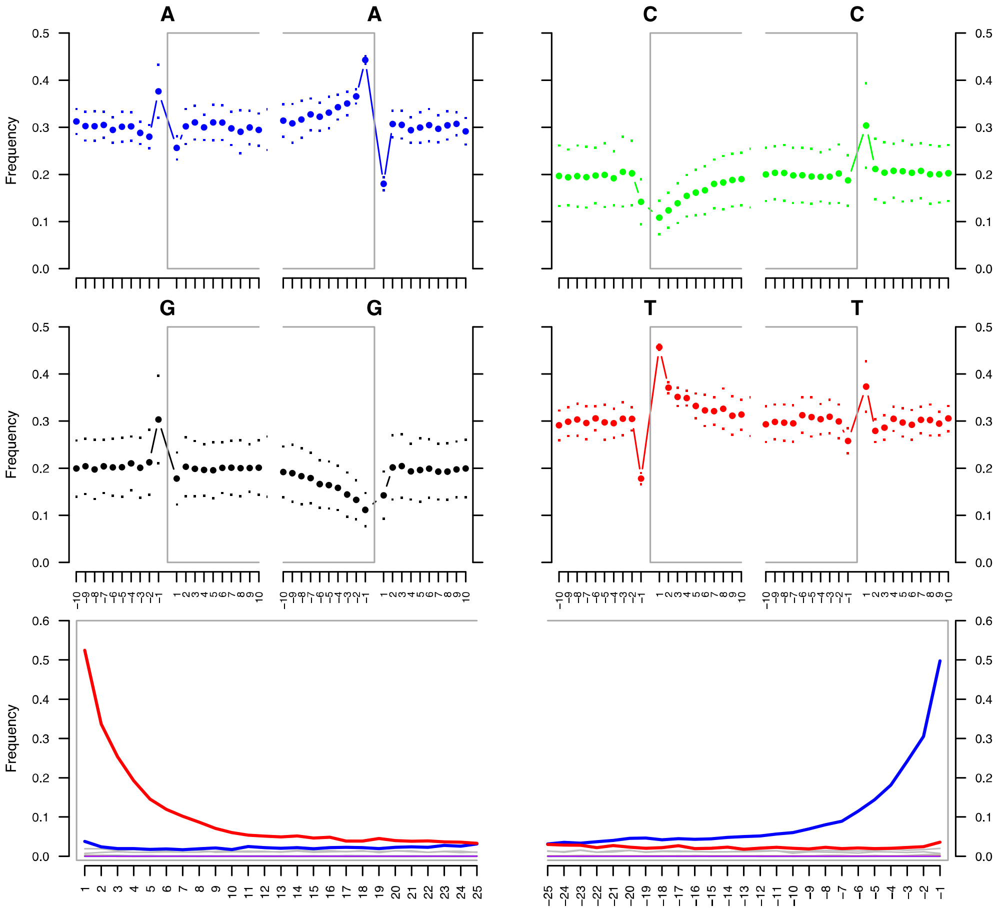
Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

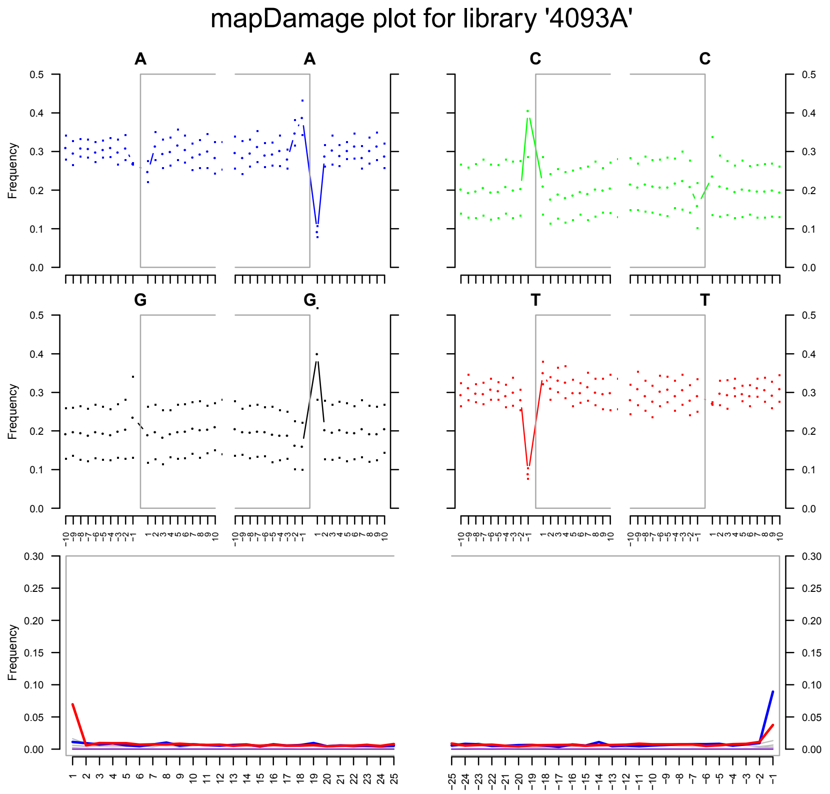
How to cite this article: Soubrier, J. *et al.* Early cave art and ancient DNA record the origin of European bison. *Nat. Commun.* **7**, 13158 doi: 10.1038/ncomms13158 (2016).

2.3 Supplementary Information

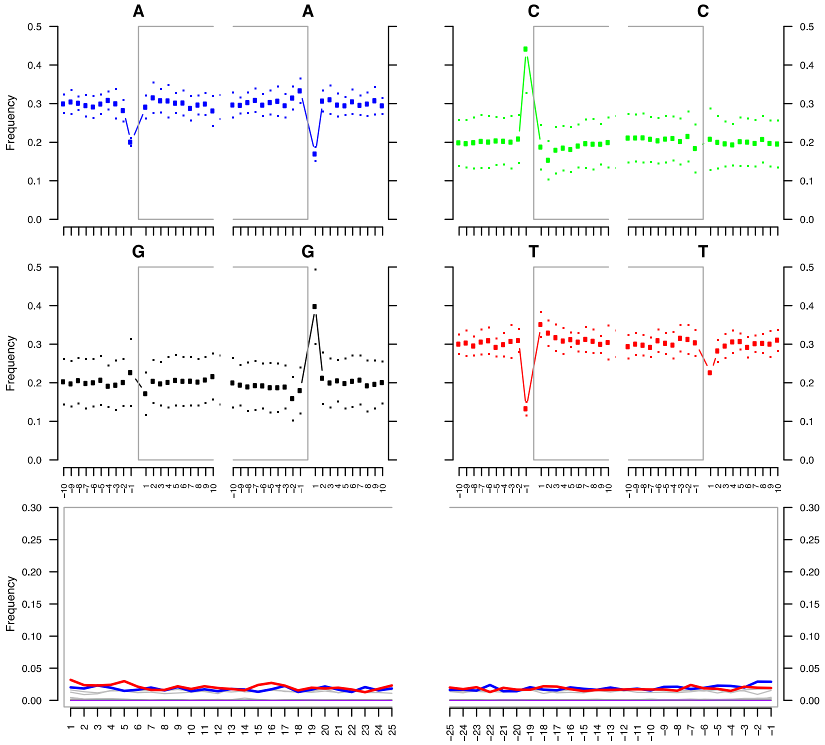
Supplementary Figures



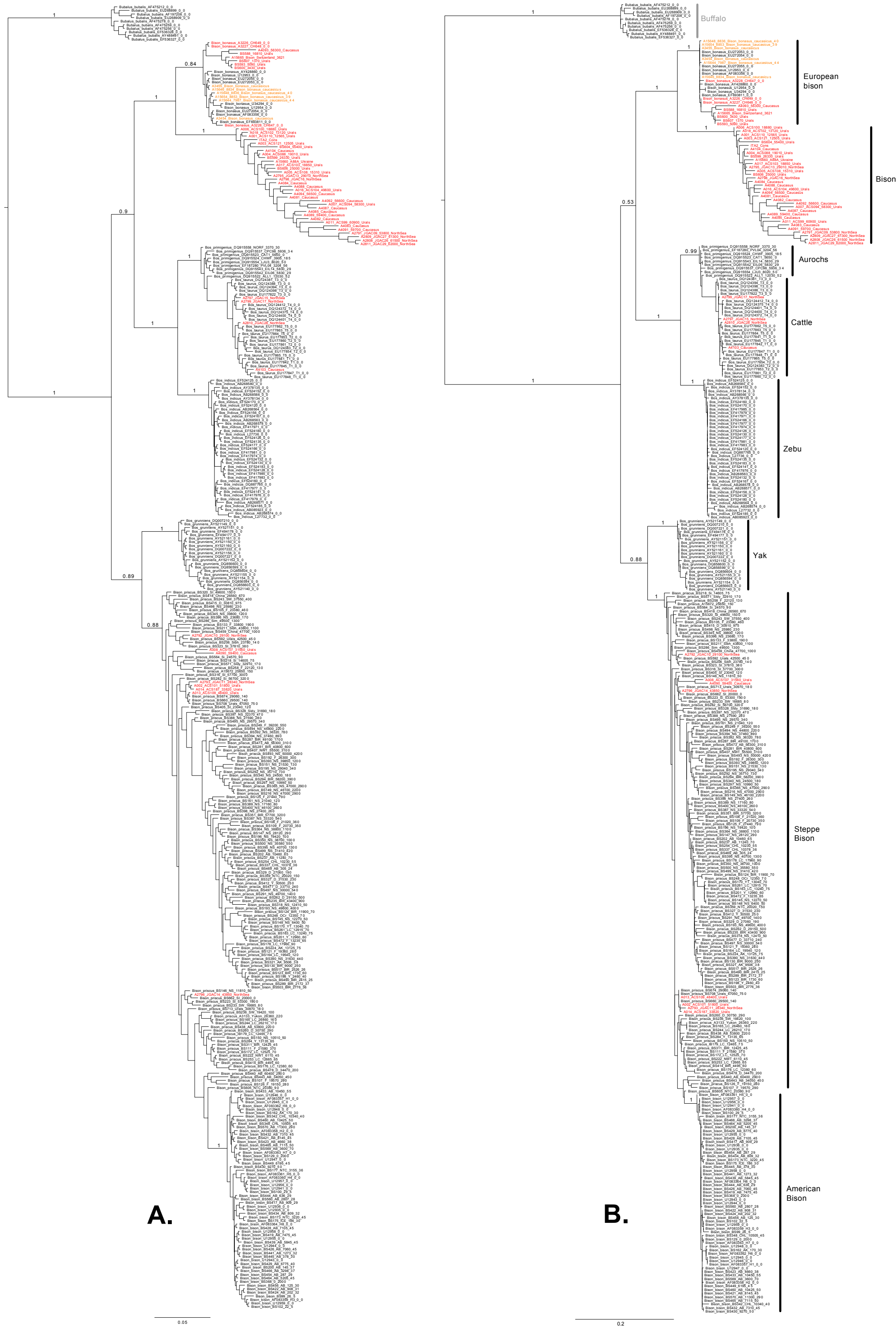
Supplementary Fig 1. Example of damage profile (sample LE257) obtained after sequencing of the whole mitochondrial genome using no treatment for the library preparation. As expected, there is an excess of purines found at the genomic position preceding the mapped reads, and an excess of C>T transitions at the first few positions of the reads.



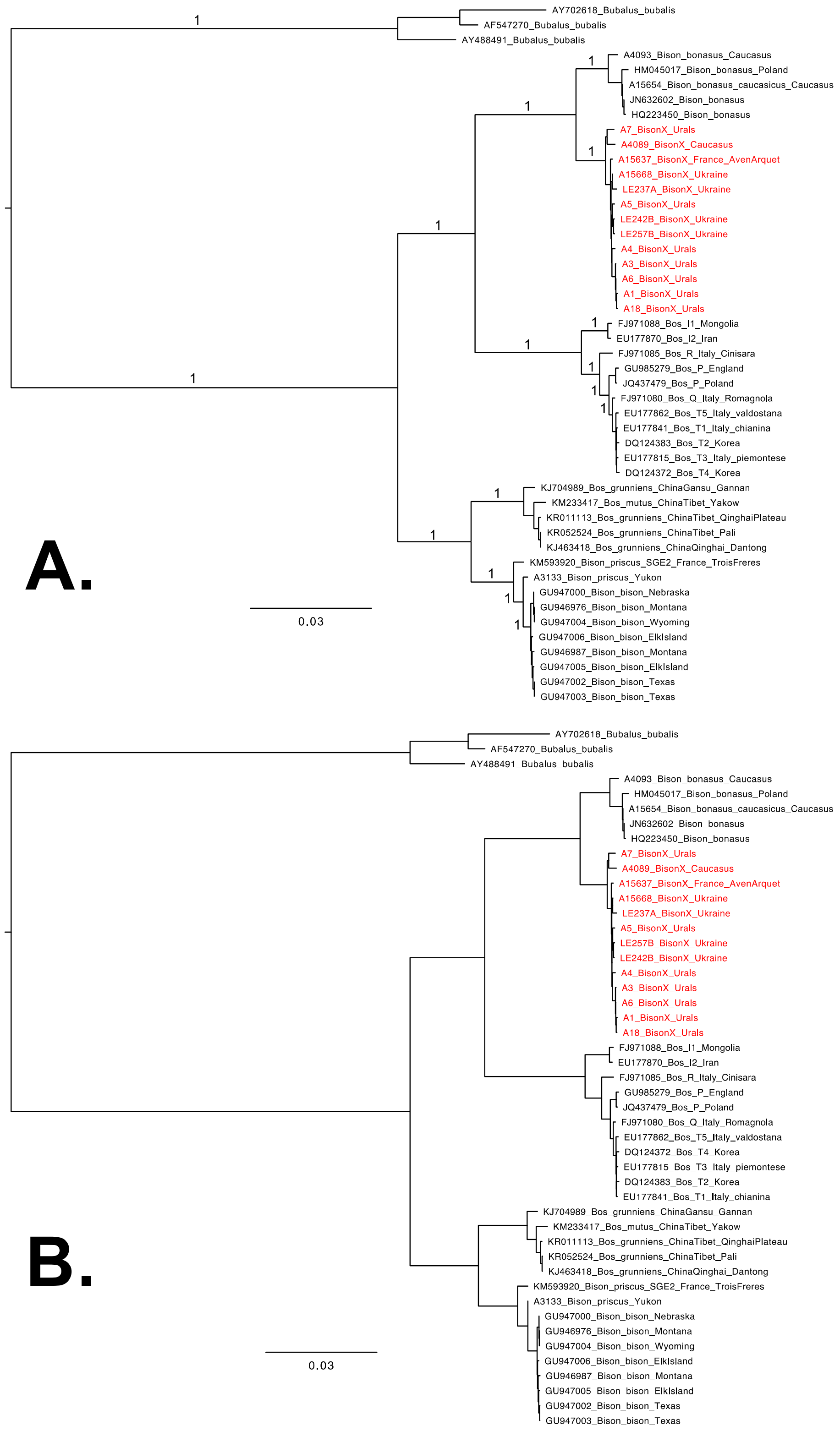
Supplementary Fig 2. Example of damage profile (sample A4093) obtained after sequencing of the whole mitochondrial genome using UDG-half treatment for the library preparation. As expected, there is an excess of cytosine found at the genomic position preceding the mapped reads, and an excess of C>T (and complementary G>A) transitions at the first (last) position of the reads.



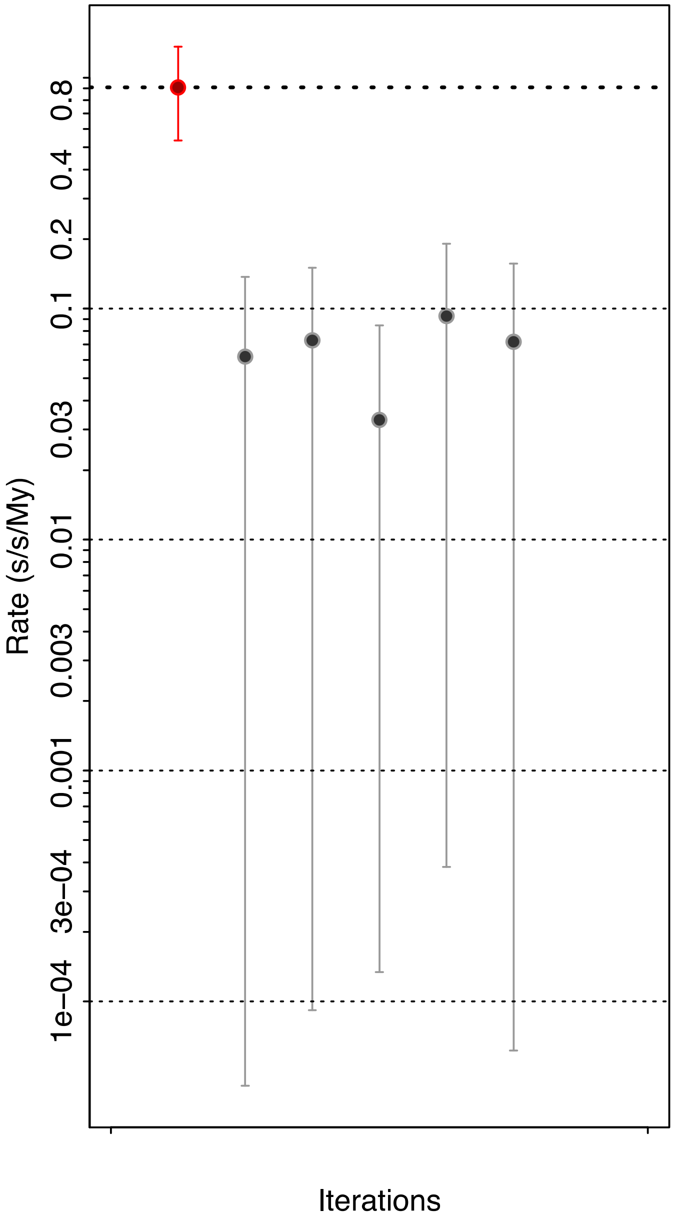
Supplementary Fig 3. Example of damage profile (sample A18) obtained after sequencing of the whole mitochondrial genome using full USER treatment for the library preparation. As expected, there is an excess of cytosine found at the genomic position preceding the mapped reads, and no excess of C>T transitions at the start of the reads.



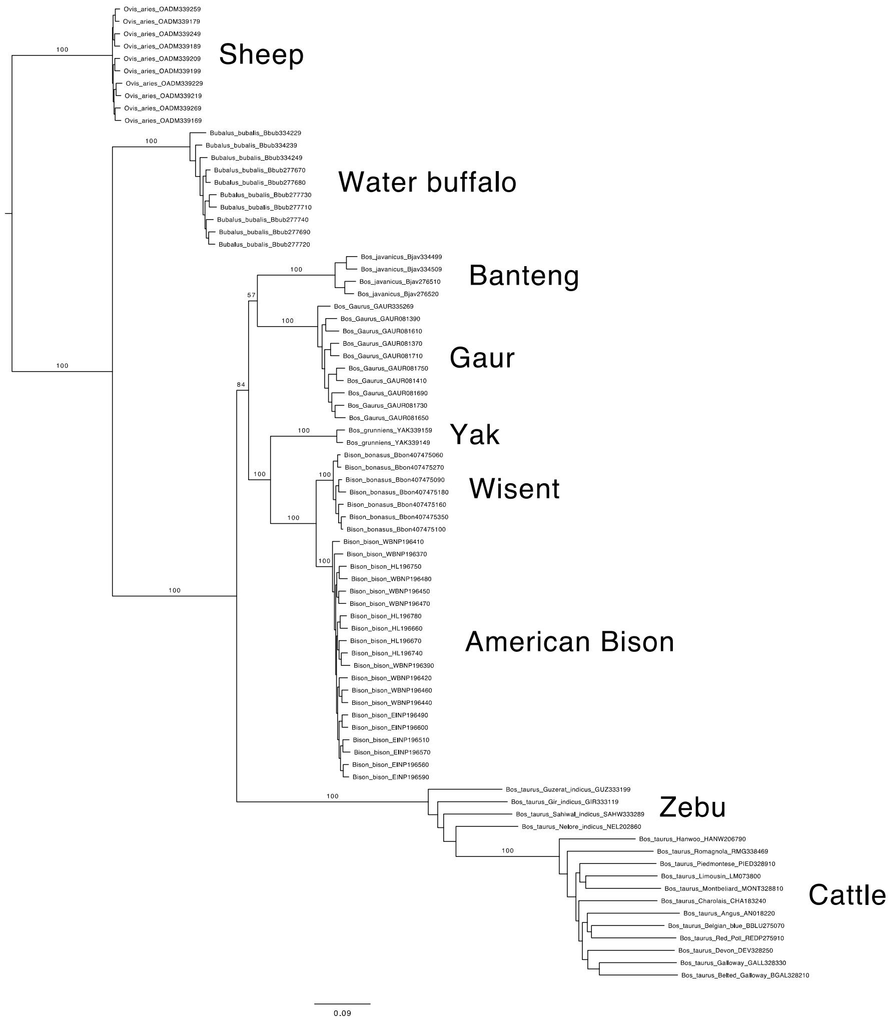
Supplementary Fig 4. Phylogenetic trees of mitochondrial control region sequences from 362 bovid samples. **A.** Majority-rule consensus tree from MrBayes. **B.** Maximum-likelihood tree from PhyML. The 60 newly sequenced individuals are in red font, with the Caucasian bison (*B. bonasus caucasicus*) in orange. Scale bars are given in substitutions per site.


1
AY702618_Bubalus_bubalis
AF547270_Bubalus_bubalis
AY488491_Bubalus_bubalis
1
A4093_Bison_bonasus_Caucasus
HM045017_Bison_bonasus_Poland
1
A15654_Bison_bonasus_caucasicus_Caucasus
JN632602_Bison_bonasus
HQ223450_Bison_bonasus
A7_BisonX_Urals
1
A4089_BisonX_Caucasus
A15637_BisonX_France_AvenArquet
A15668_BisonX_Ukraine
1
LE237A_BisonX_Ukraine
A5_BisonX_Urals
LE242B_BisonX_Ukraine
LE257B_BisonX_Ukraine
A4_BisonX_Urals
A3_BisonX_Urals
A6_BisonX_Urals
A1_BisonX_Urals
A18_BisonX_Urals
1
FJ971088_Bos_I1_Mongolia
EU177870_Bos_I2_Iran
1
FJ971085_Bos_R_Italy_Cinisara
1
1
GU985279_Bos_P_England
JQ437479_Bos_P_Poland
1
1
FJ971080_Bos_Q_Italy_Romagnola
EU177862_Bos_T5_Italy_valdostana
EU177841_Bos_T1_Italy_chianina
DQ124383_Bos_T2_Korea
EU177815_Bos_T3_Italy_piemontese
DQ124372_Bos_T4_Korea
1
KJ704989_Bos_grunniens_ChinaGansu_Gannan
KM233417_Bos_mutus_ChinaTibet_Yakow
KR011113_Bos_grunniens_ChinaTibet_QinghaiPlateau
KR052524_Bos_grunniens_ChinaTibet_Pali
KJ463418_Bos_grunniens_ChinaQinghai_Dantong
1
A.
KM593920_Bison_priscus_SGE2_France_TroisFreres
1
A3133_Bison_priscus_Yukon
1
GU947000_Bison_bison_Nebraska
GU946976_Bison_bison_Montana
1
GU947004_Bison_bison_Wyoming
GU947006_Bison_bison_ElkIsland
GU946987_Bison_bison_Montana
GU947005_Bison_bison_ElkIsland
GU947002_Bison_bison_Texas
GU947003_Bison_bison_Texas
0.03
AY702618_Bubalus_bubalis
AF547270_Bubalus_bubalis
AY488491_Bubalus_bubalis
A4093_Bison_bonasus_Caucasus
HM045017_Bison_bonasus_Poland
A15654_Bison_bonasus_caucasicus_Caucasus
JN632602_Bison_bonasus
HQ223450_Bison_bonasus
A7_BisonX_Urals
A4089_BisonX_Caucasus
A15637_BisonX_France_AvenArquet
A15668_BisonX_Ukraine
LE237A_BisonX_Ukraine
A5_BisonX_Urals
LE257B_BisonX_Ukraine
LE242B_BisonX_Ukraine
A4_BisonX_Urals
A3_BisonX_Urals
A6_BisonX_Urals
A1_BisonX_Urals
A18_BisonX_Urals
FJ971088_Bos_I1_Mongolia
EU177870_Bos_I2_Iran
FJ971085_Bos_R_Italy_Cinisara
GU985279_Bos_P_England
JQ437479_Bos_P_Poland
FJ971080_Bos_Q_Italy_Romagnola
EU177862_Bos_T5_Italy_valdostana
DQ124372_Bos_T4_Korea
EU177815_Bos_T3_Italy_piemontese
DQ124383_Bos_T2_Korea
EU177841_Bos_T1_Italy_chianina
KJ704989_Bos_grunniens_ChinaGansu_Gannan
KM233417_Bos_mutus_ChinaTibet_Yakow
KR011113_Bos_grunniens_ChinaTibet_QinghaiPlateau
KR052524_Bos_grunniens_ChinaTibet_Pali
KJ463418_Bos_grunniens_ChinaQinghai_Dantong
B.
KM593920_Bison_priscus_SGE2_France_TroisFreres
A3133_Bison_priscus_Yukon
GU947000_Bison_bison_Nebraska
GU946976_Bison_bison_Montana
GU947004_Bison_bison_Wyoming
GU947006_Bison_bison_ElkIsland
GU946987_Bison_bison_Montana
GU947005_Bison_bison_ElkIsland
GU947002_Bison_bison_Texas
GU947003_Bison_bison_Texas
0.03

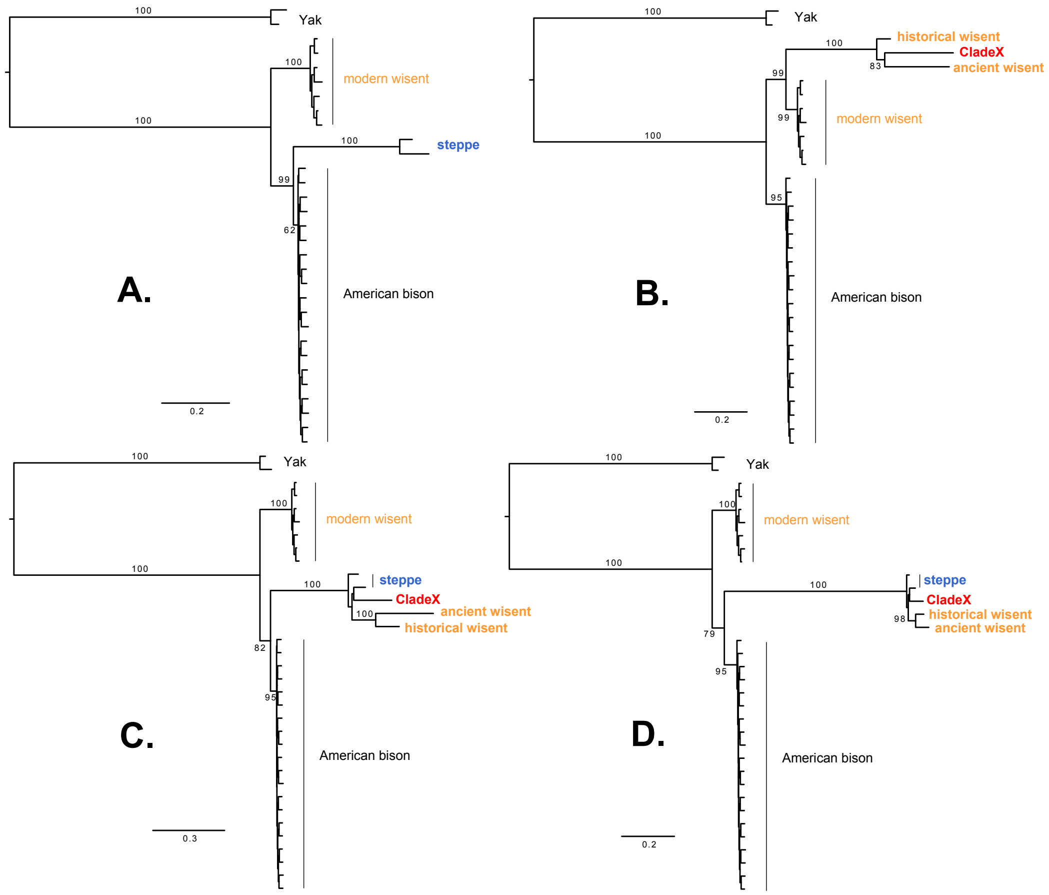
Supplementary Fig 5. Phylogenetic trees inferred from whole mitochondrial genomes. **A.** Majority-rule consensus tree from MrBayes. **B.** Maximum-likelihood tree from PhyML. CladeX bison individuals are colored in red. Scale bars are given in substitutions per site.



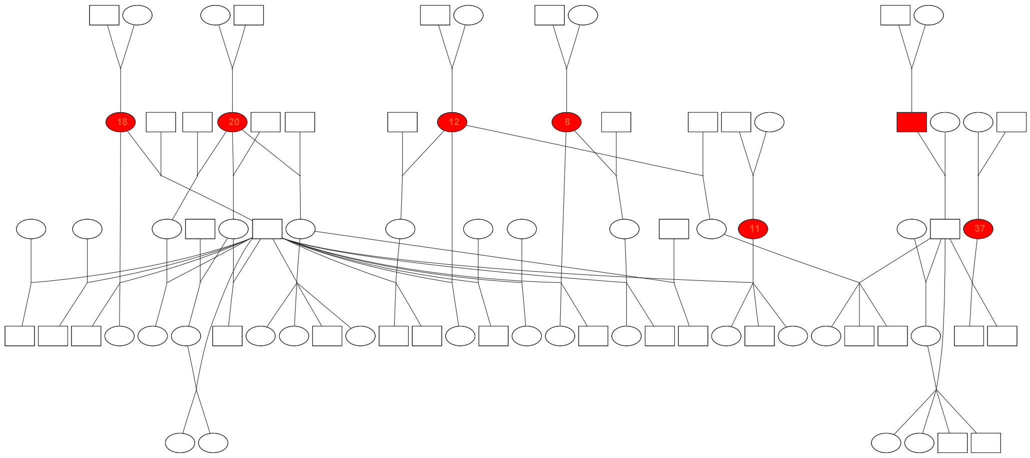
Supplementary Fig 6. Date-randomization test. The red circle and dotted line represent the mean estimate of the molecular rate obtained in the phylogenetic analysis of wisent and CladeX, calibrated using the radiocarbon dates associated with the ancient sequences. The grey lines represent the 95% HPD intervals of rates estimated with randomized dates. None of these margins overlap with the mean rate estimate from the original data set, demonstrating that the radiocarbon dates used for this study contain sufficient temporal information for calibrating the molecular clock.



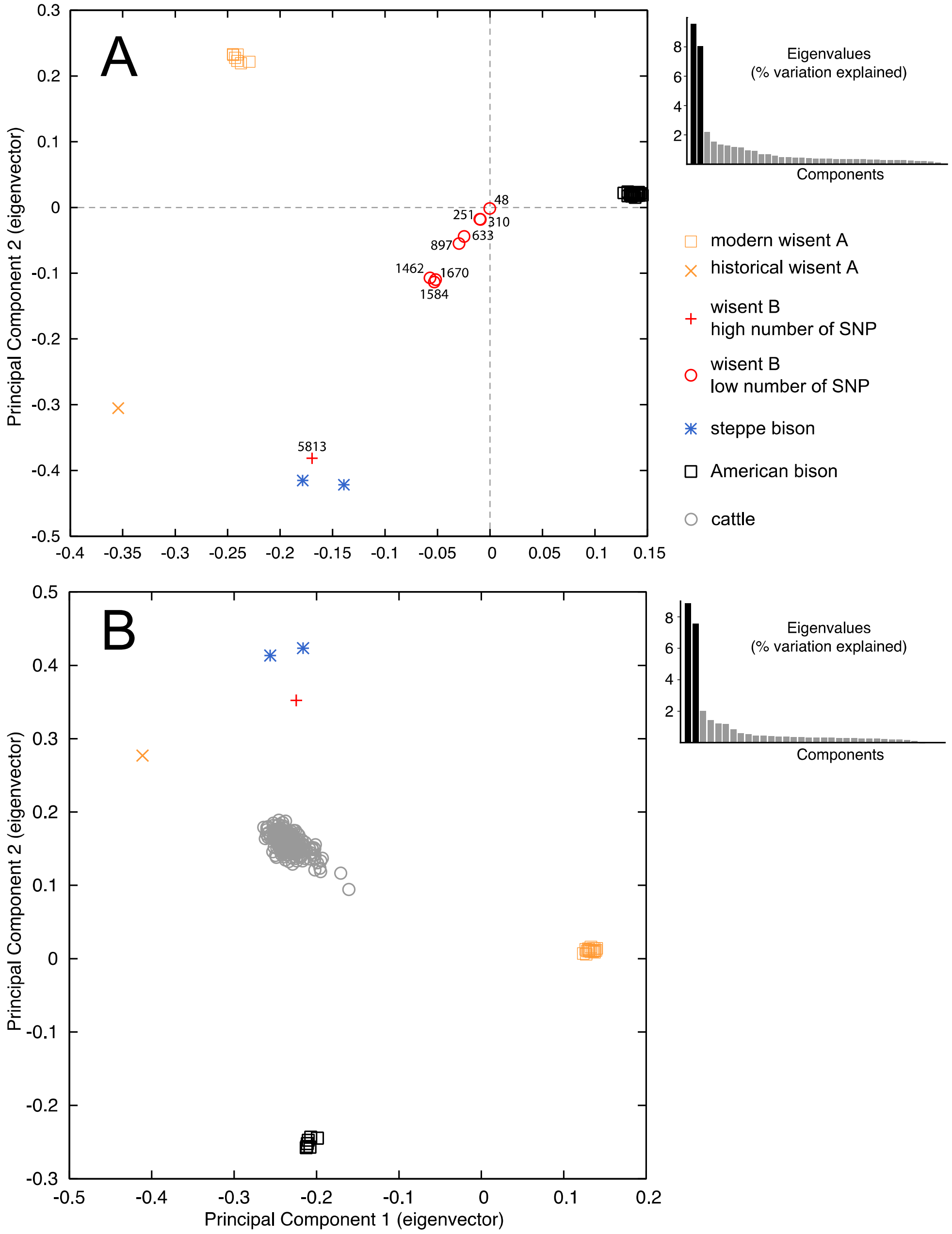
Supplementary Fig 7. Maximum-likelihood phylogeny of modern bovid species (and sheep as outgroup) from ~40k nuclear loci.



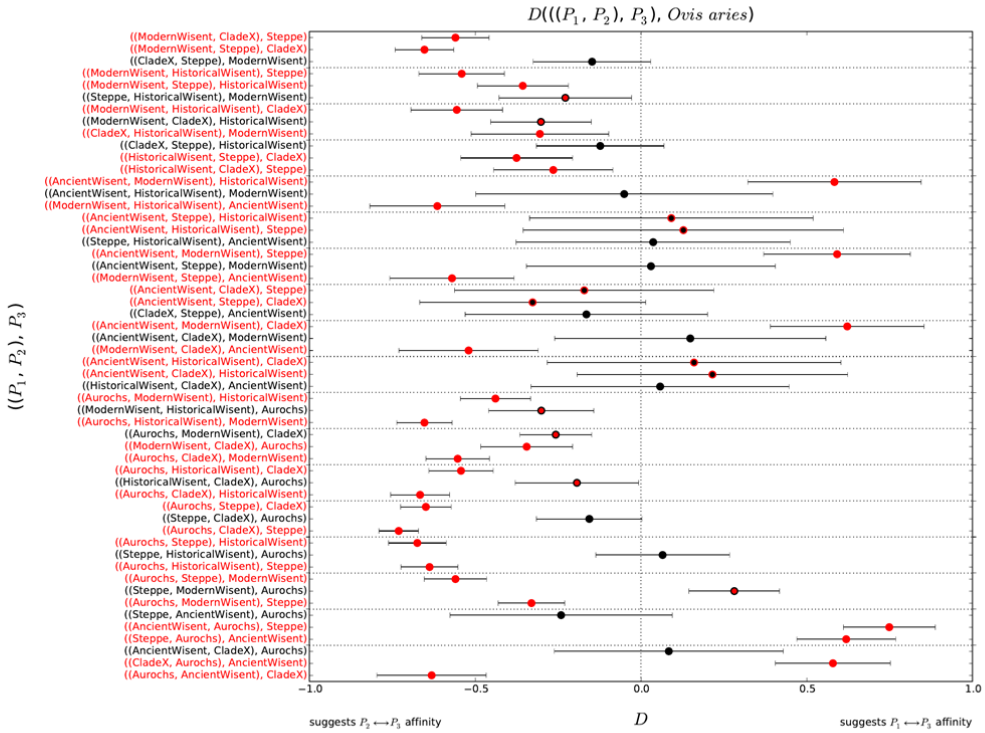
Supplementary Fig 8. Maximum-likelihood phylogenies of modern and ancient bison (and yak as outgroup), from ~10k nuclear loci. **A.** Phylogeny including the two ancient steppe bison. **B.** Phylogeny including the three pre-modern wisent. **C.** Phylogeny including the two steppe bison and three pre-modern wisent (ancient, historical and CladX). **D.** Replicate of **C.** but only using transversions for the non-modern samples.



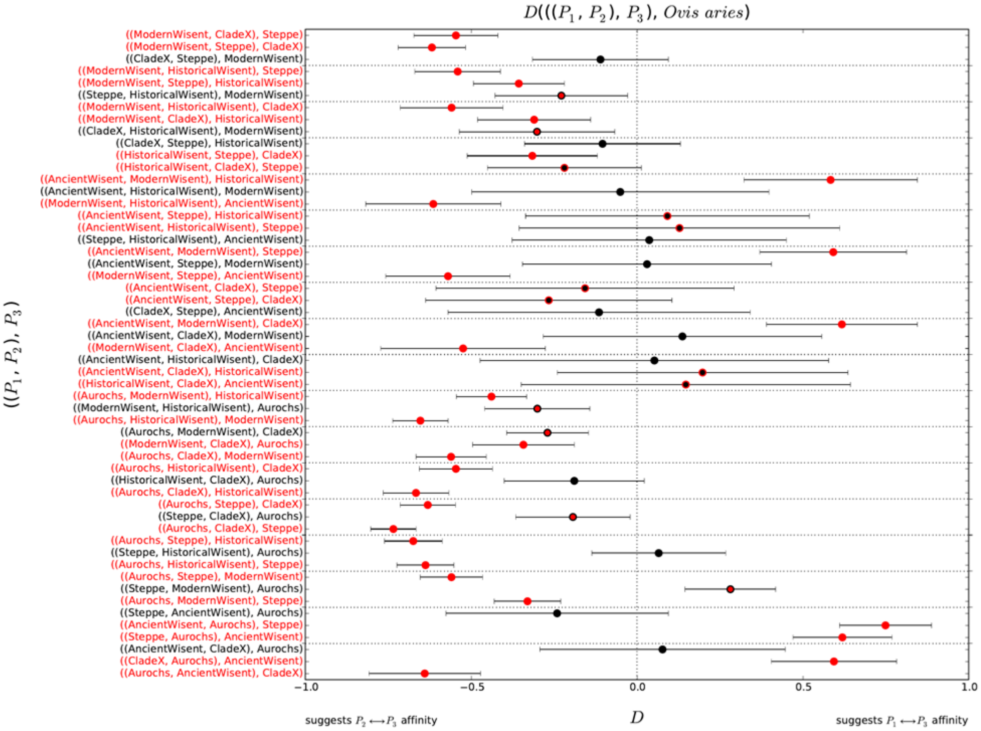
Supplementary Fig 9. Pedigree of wisent from the Białowieża Forest (Poland), from which seven genotyped individuals (in red) were included in the present study.



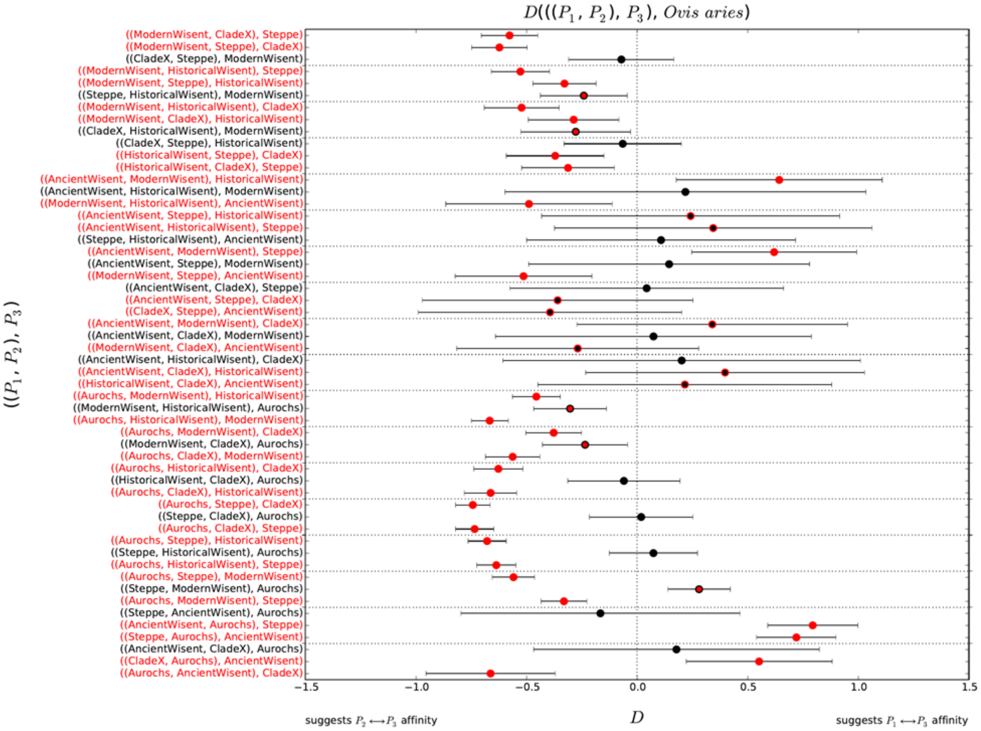
Supplementary Fig 10: A) Principal Component Analysis for nine CladeX individuals (including sample A006), one historical wisent, one ancient wisent, two steppe bison, seven modern wisent and 20 American bison. The numbers on the plot report the number of loci called for the individuals clustering towards zero coordinates (from Supplementary Table 2). Eigenvector 1 explains 9.58% of the variation, while eigenvector 2 explains 7.96% of the variation. B) Same Principal Component Analysis as Figure 3C with cattle individuals from Decker et al. (2009) projected onto original components.



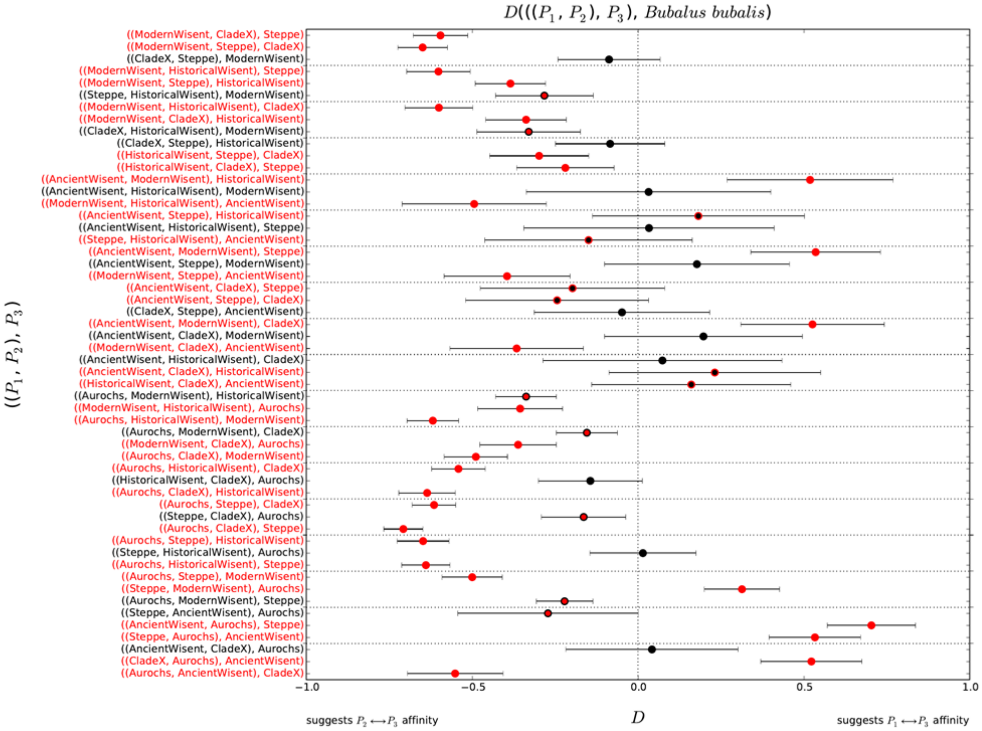
Supplementary Fig 11: Topology testing using D statistics, with sheep as outgroup. The topology being tested is shown on the vertical axis, with the most parsimonious of three possible topologies written in black. Data points that are significantly different (more than three standard errors) from zero are shown in red. The data point representing the topology closest to zero, amongst a set of three possible topologies, is shown with a black outline. Error bars are three standard errors either side of the data point, where the standard error was calculated using a block jackknife.



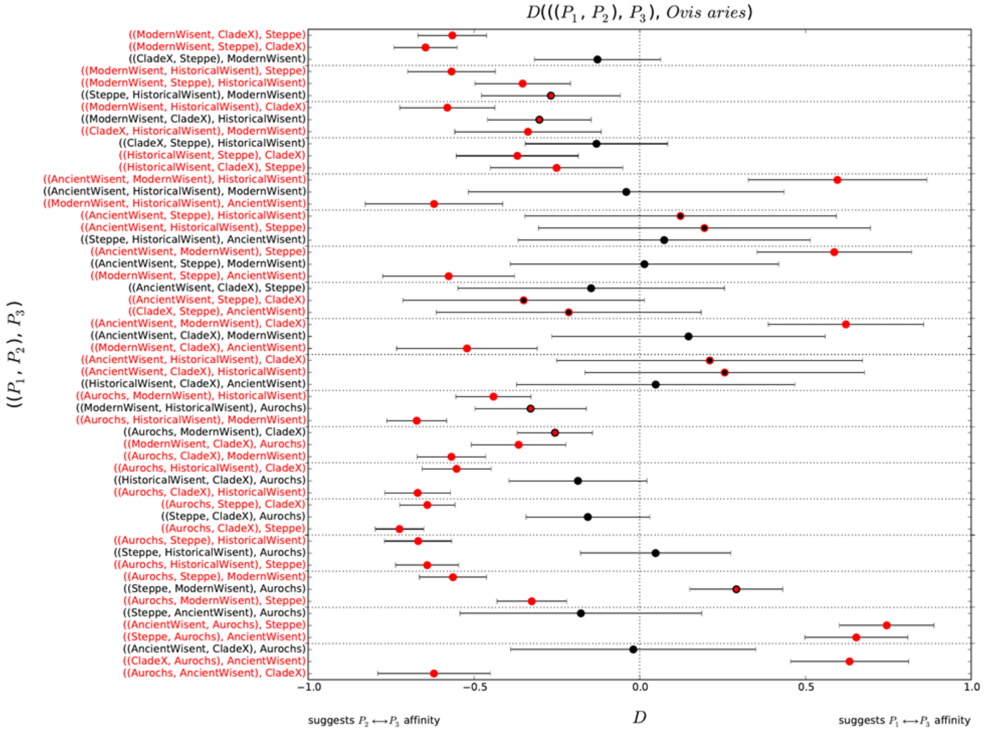
Supplementary Fig 12: Topology testing using D statistics, with sheep as outgroup. As in Supplementary Figure 11, except that sample A006 has been omitted from the CladeX group.



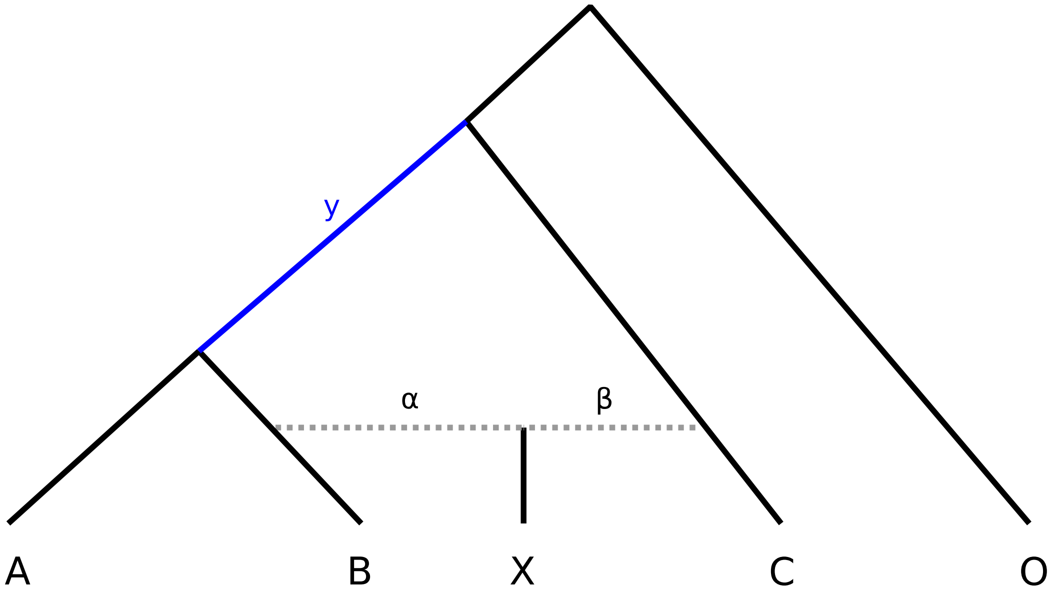
Supplementary Fig 13: Topology testing using D statistics, with sheep as outgroup. As in Supplementary Figure 11, except that genotypes called from read depths <2 have been omitted for extinct individuals.



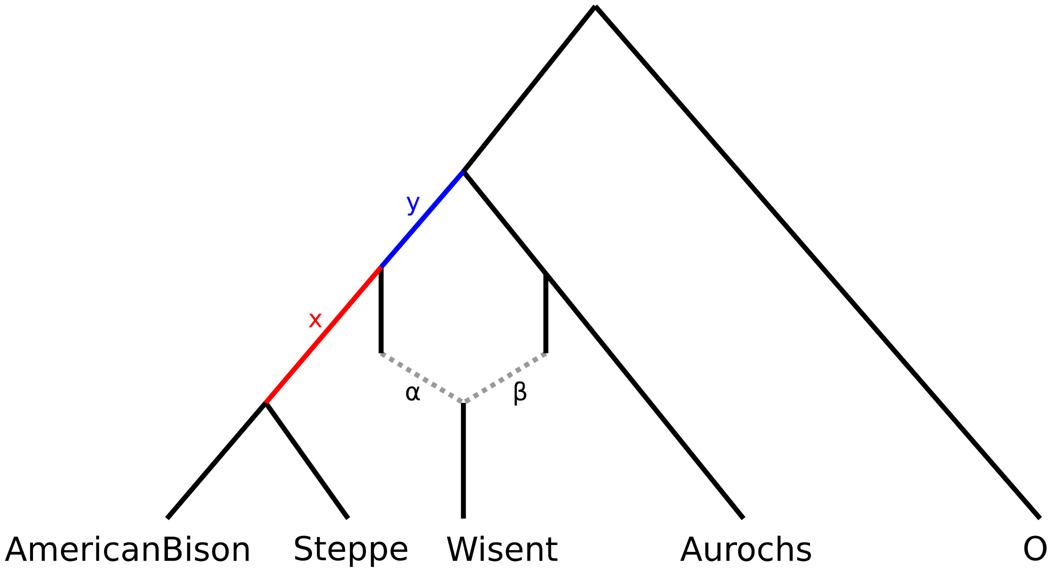
Supplementary Fig 14: Topology testing using D statistics, with Asian water buffalo as outgroup. As in Supplementary Figure 11, except the outgroup has been changed.



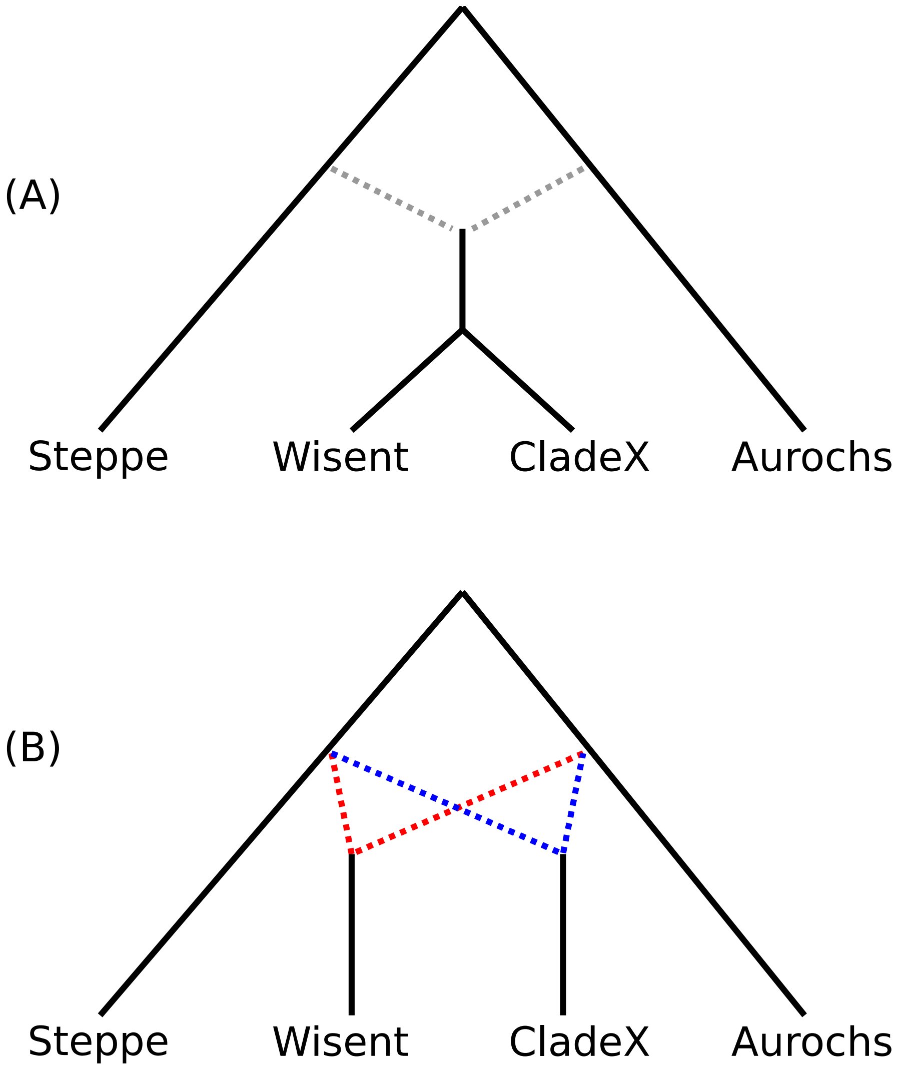
Supplementary Fig 15: Topology testing using D statistics, with sheep as outgroup. As in Supplementary Figure 11, except in extinct individuals, alleles have been randomly sampled from sites called as heterozygotes to simulate haploid sampling.



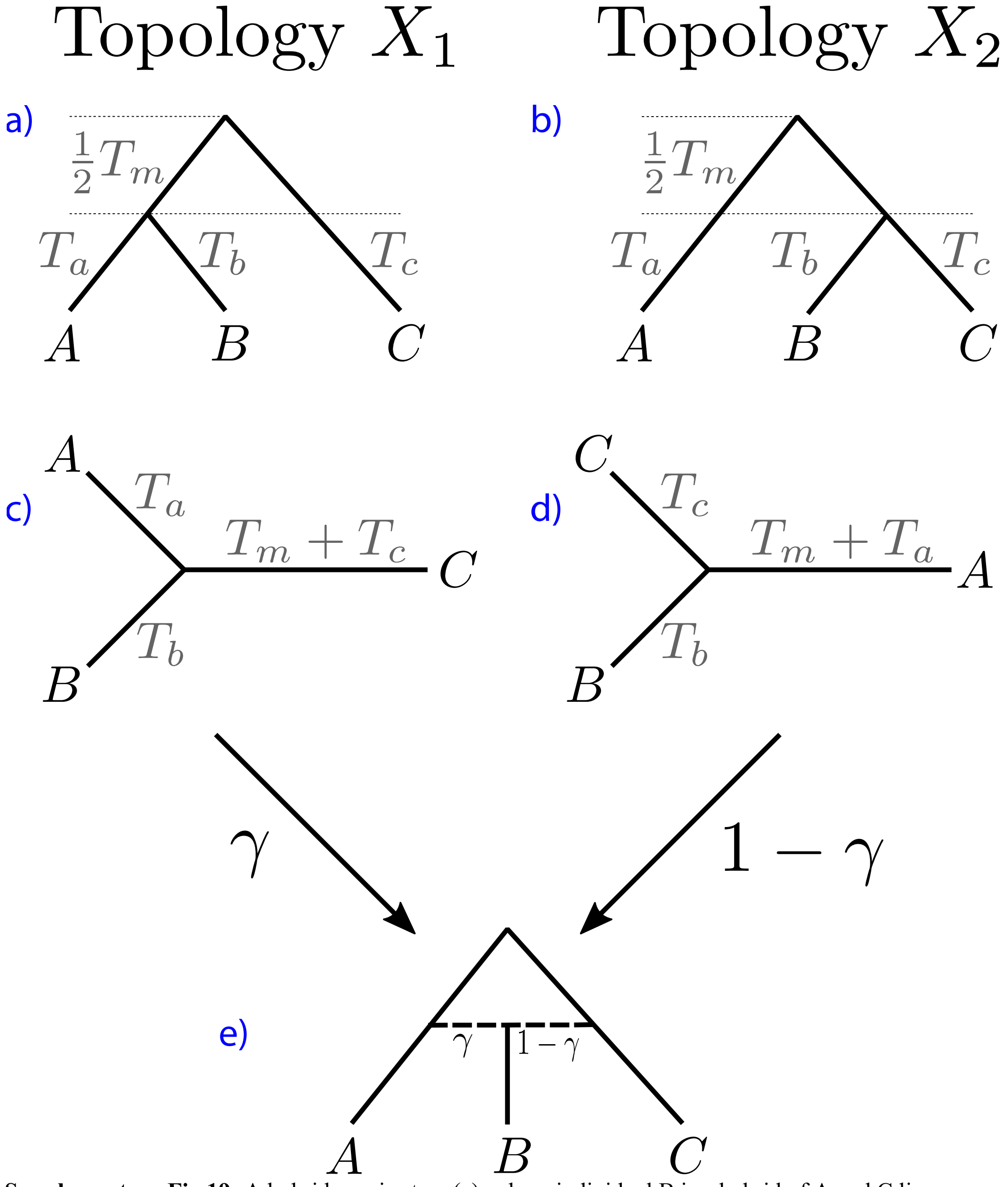
Supplementary Fig 16: An admixture graph showing the ancestry of X, where α is the proportion of ancestry from B and β=1-α is the proportion of ancestry from C.



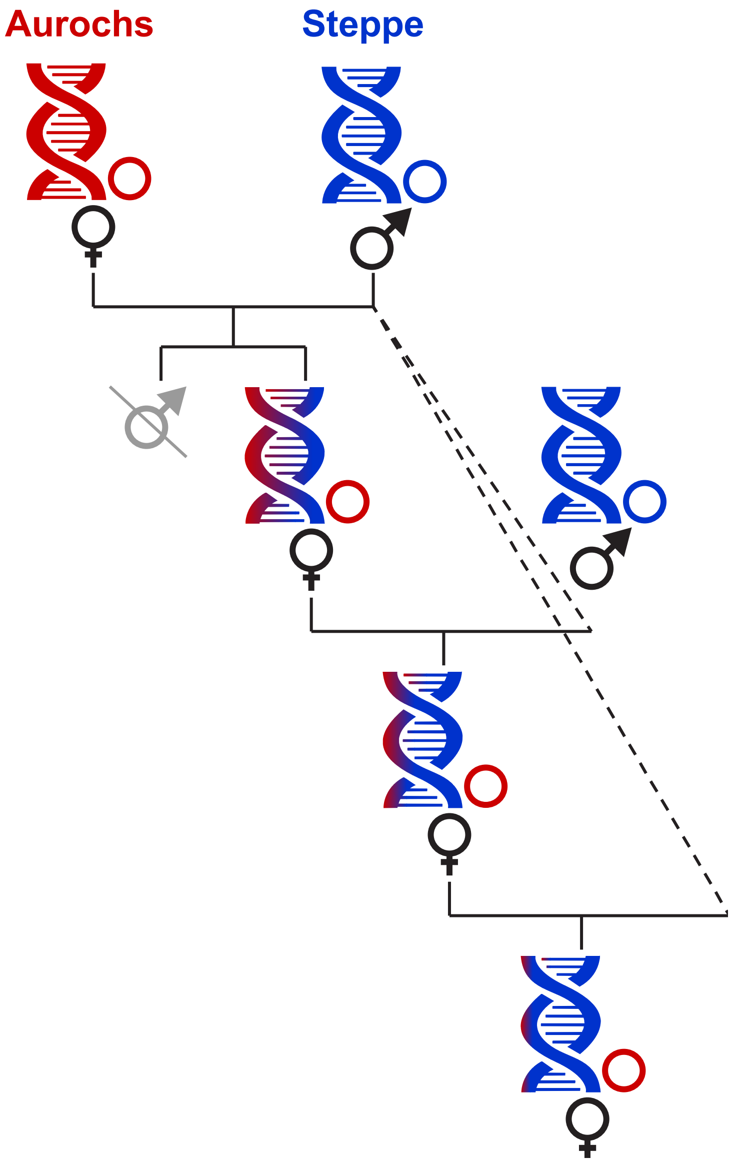
Supplementary Fig 17: An admixture graph showing the ancestry of the wisent, where α is the proportion of ancestry from steppe and β=1-α is the proportion of ancestry from aurochs.



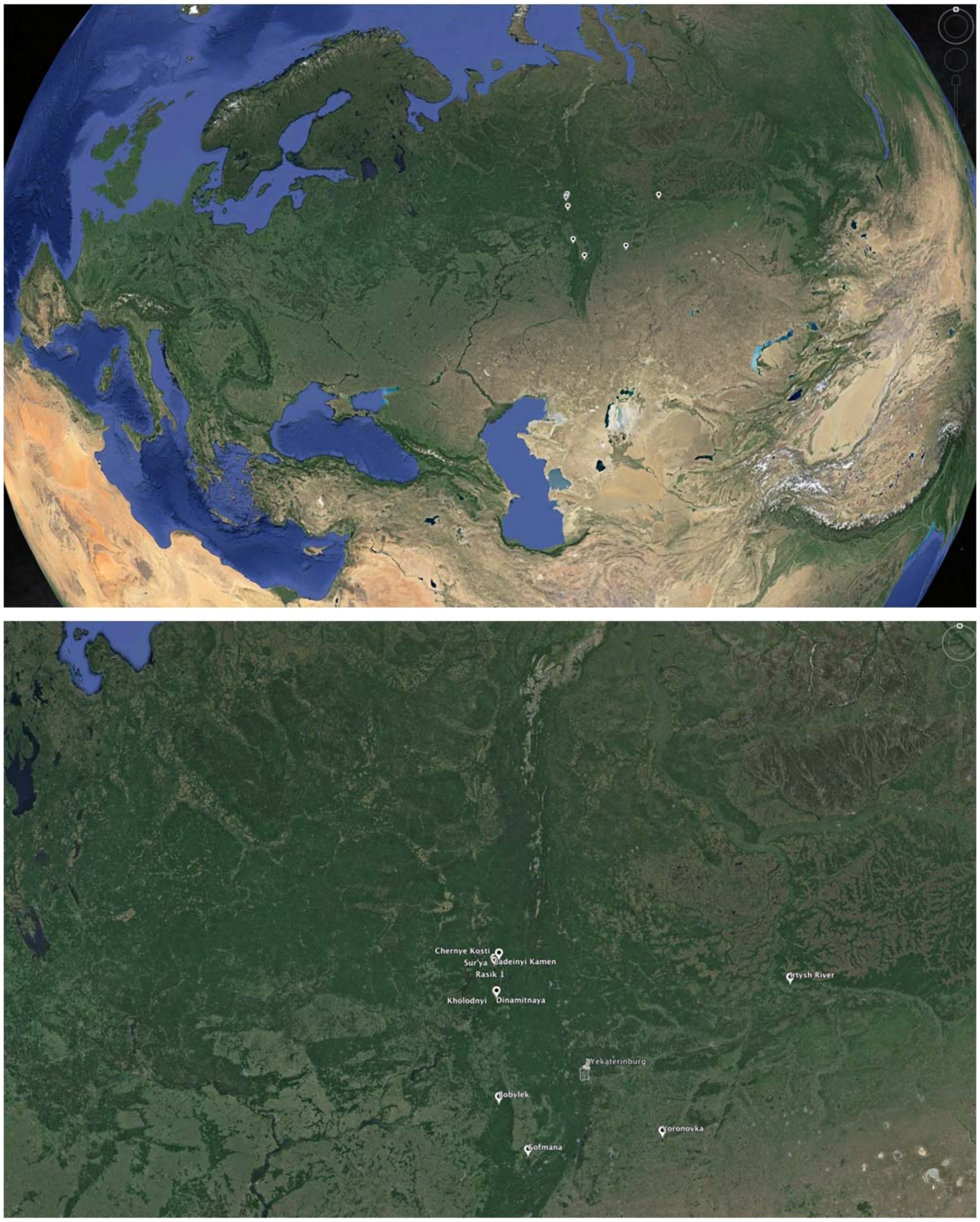
Supplementary Fig 18: Admixture graphs representing (A) a single hybridisation event prior to the divergence of the wisent, and (B) two independent hybridisation events leading to a wisent clade and a CladeX.



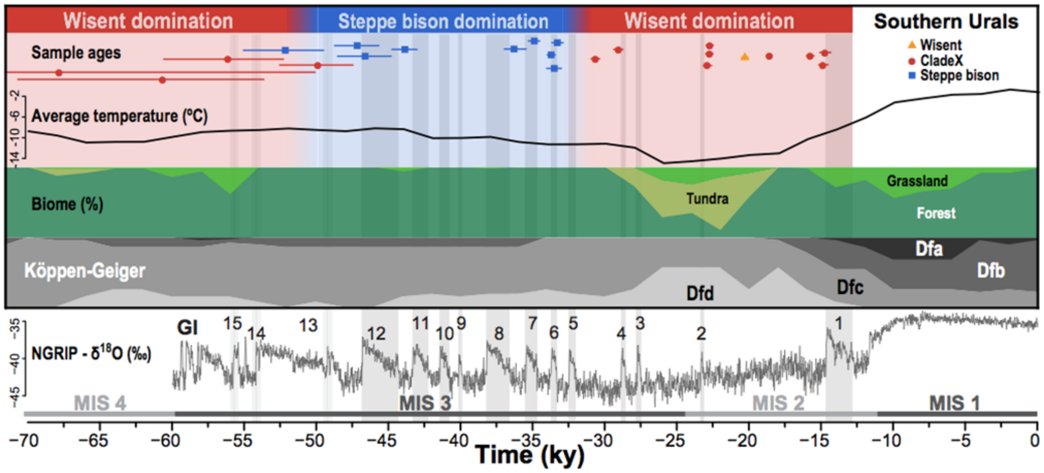
Supplementary Fig 19: A hybrid species tree (e), where individual B is a hybrid of A and C lineages, has two contributing species trees, (a) topology X_1, and (b) topology X_2, with proportion γ from topology X_1 and proportion $1-\gamma$ from topology X_2. The unrooted gene trees are shown for (c) topology X_1, and (d) topology X_2. Branch lengths T_a, T_b, T_c and $T_m$have units $2N_e\mu$ generations.



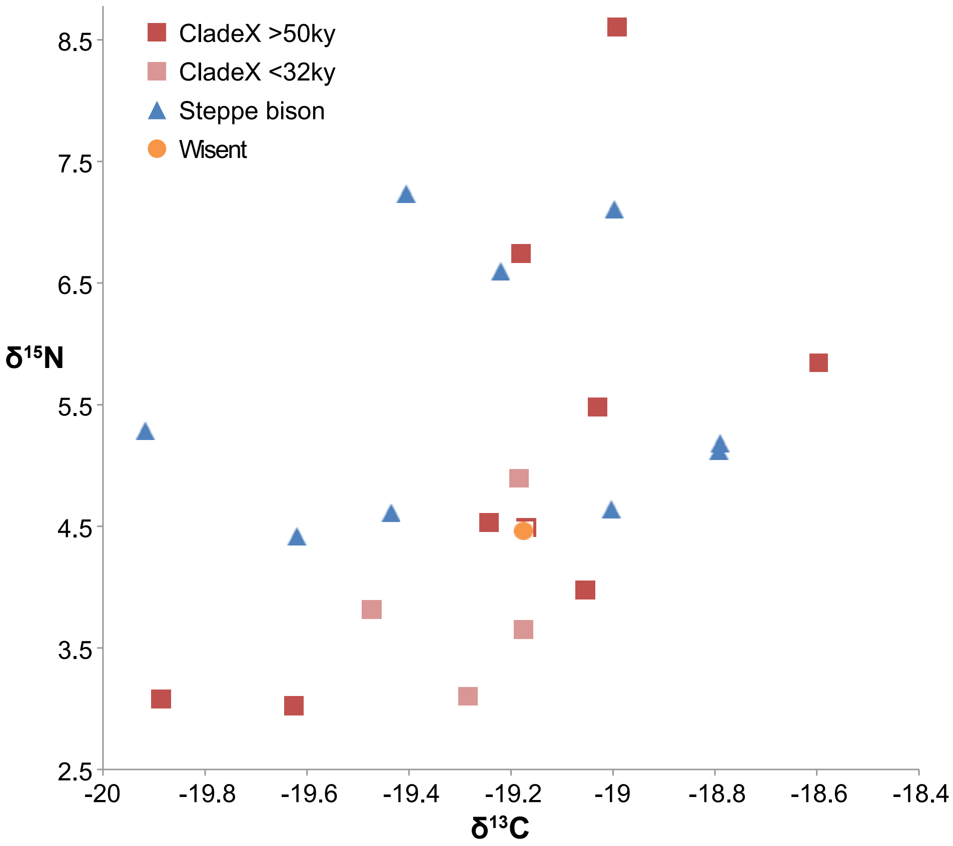
Supplementary Fig 20. Schematic representation of asymmetrical hybridisation between female aurochs and male steppe bison, and its genetic imprint on both nuclear and mitochondrial genomes after a few generations. The coloured double helix represents the nuclear genome, while the circles represent the strictly maternally inherited mitochondrial genome.



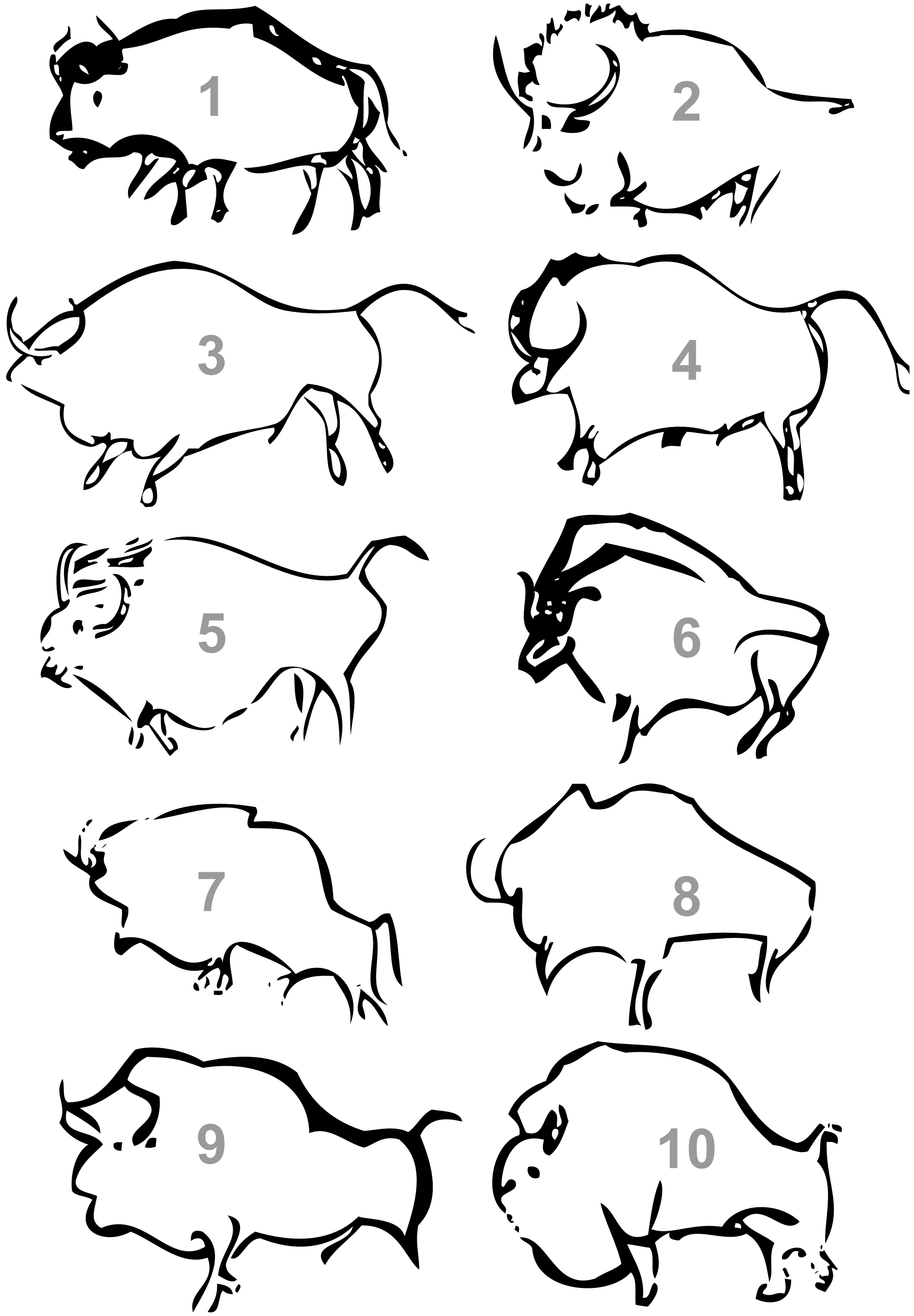
Supplementary Fig 21. Location of all cave sites from which bison samples have been genotyped in the Ural region.



Supplementary Fig 22. Chronology of the Urals samples showing a series of replacement patterns that correlate with climate events. Individual calibrated AMS dates are plotted on top of the NGRIP δO^{18} record [1]. Greenland Interstadials (GI) are numbered in black, and Marine Isotope Stages (MIS) in grey. Inferred average temperature, biome reconstruction and proportion of the area for different Koppen climate classes are shown for the exact region where bison were sampled in southern Urals (Koppen classes: D for 'snow', f for 'fully humid', then a=hot summer; b=warm summer; c=cool summer; d=extremely continental). The most recent population replacement between wisent and steppe bison occurs around 32-33 ky, when major environmental transitions are also observed: 1) Globally, as shown on the NGRIP record with the last major interglacial event (GI 5) before a long period of cold climate; but also 2) Locally, as shown on both the average temperature and biome reconstructions. In this situation, wisent are associated with a cooler climate and the presence of tundra-like vegetation. Although dating resolution is degrading for deeper time, a similar shift is apparent around 50-52 kya. Steppe bison occupied this environment in MIS 3, but have not been detected after this stage and indeed were in a severe population decline by GI 1[2].



Supplementary Fig 23. Stable δ13C and δ15N isotope values for all genotyped bison sampled from the Ural region.



Supplementary Fig 24. Steppe-like morphologies. In European Palaeolithic art, some bison depictions show morphological traits and anatomical details compatible with the morphology of steppe bison (or American bison ancestry). Dates are given as indication based on archaeological occupation determined for each site, or, in the absence of such dating, based on stylistic comparison with other depictions:

1. Grotte Chauvet-Pont d'Arc (Ardèche, France). Blurred black charcoal drawing. Aurignacian period (~35,100 ± 175 calBP. (from C. Fritz and G. Tosello)

2. Grotte de Lascaux (Dordogne, France). Carving. Solutrean (~22,200 ± 380 calBP) or early Magdalenian period (between ~19,300 ± 561 and ~20,597 ± 375 calBP). (adapted from A. Glory[3])

3. Grotte de Lascaux (Dordogne, France). Carving. Solutrean (~22,200 ± 380 calBP) or early Magdalenian period (between ~19,300 ± 561 and ~20,597 ± 375 calBP). (adapted from A. Glory[3])

4. Grotte de Lascaux (Dordogne, France). Carving. Solutrean (~22,200 ± 380 calBP) or early Magdalenian period (between ~19,300 ± 561 and ~20,597 ± 375 calBP). (adapted from A. Glory[3])

5. Grotte du Gabillou (Dordogne, France). Carving. Early Magdalenian period (~20,597 ± 375 calBP). (adapted from J. Gaussen)

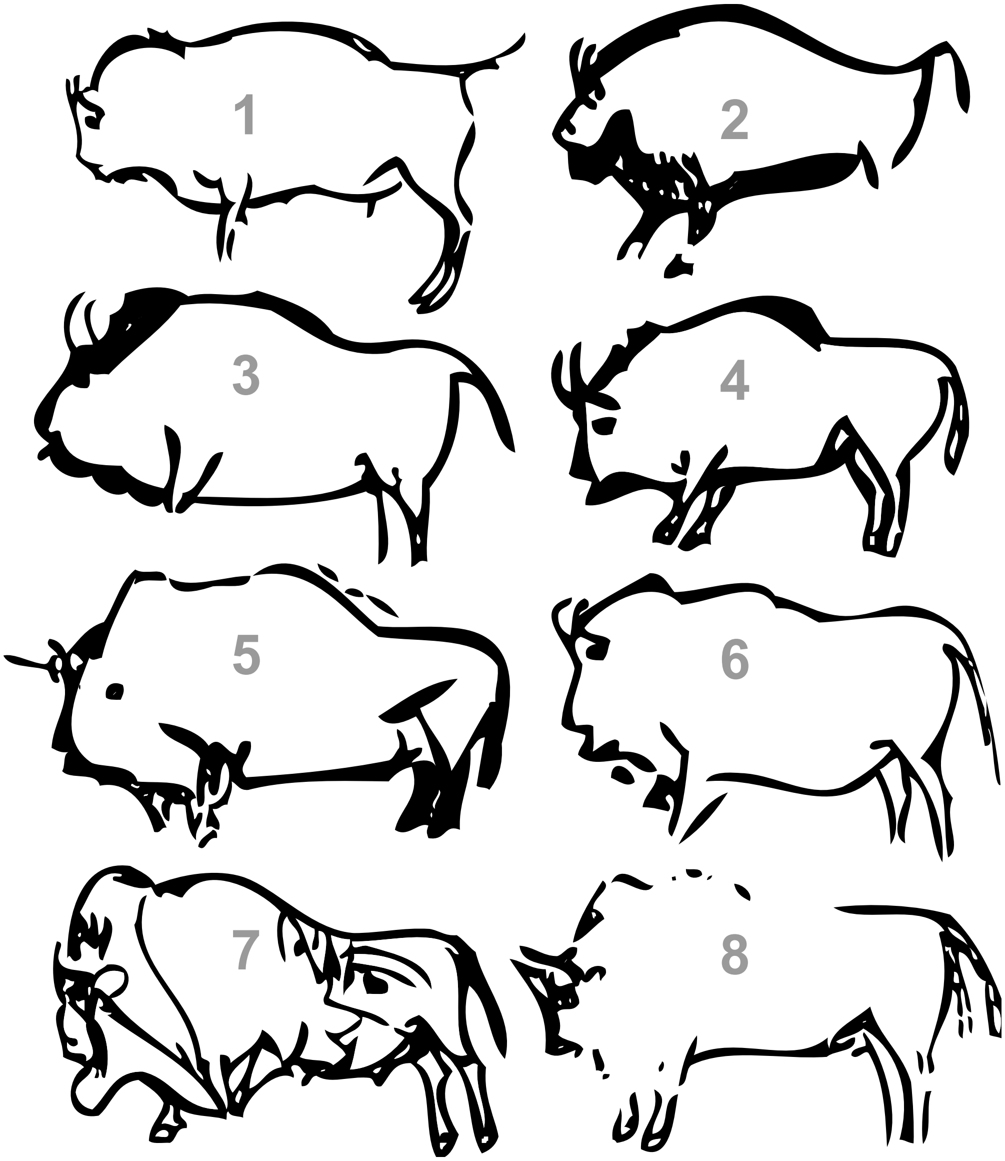
6. Grotte des Trois Frères (Ariège, France). Carving. Gravettian period (dating estimated based on stylistic analysis). (adapted from H. Breuil[4])

7. Grotte du Pech Merle (Lot, France). Painting (manganese). Gravettian period (~29,447 ± 443 calBP). (adapted from M. Lorblanchet[5])

8. Grotte du Pech Merle (Lot, France). Painting (manganese). Gravettian period (~29,447 ± 443 calBP). (adapted from M. Lorblanchet[5])

9. Grotte de La Pasiega (Cantabria, Spain). Black and red painting. Gravettian or Solutrean period (dating estimated based on stylistic analysis). (adapted from H. Breuil[4])

10. Abri du Roc de Sers (Charente, France). Carving on limestone. Solutrean period (< 20,442 ± 409 calBP). (adapted from L. Henri-Martin)



Supplementary Fig 25. Wisent-like morphologies. In European Palaeolithic art, some bison depictions show morphological traits and anatomical details compatible with identification of wisent ancestry. Dates are given as indication based on archaeological occupation determined for each site, or, in the absence of such dating, based on stylistic comparison with other depictions:

1. Grotte de Pergouset (Ardèche, France). Carving. Magdalenian period (dating estimated based on stylistic analysis). (adapted from M. Lorblanchet[5])

2. Grotte du Portel (Ariège, France). Painting. Magdalenian period (~14,250 ± 295 calBP). (adapted from H. Breuil[4])

3. Grotte de Niaux (Ariège, France). Painting. Magdalenian period (~17,000 ± 260 calBP). (adapted from H. Breuil[4])

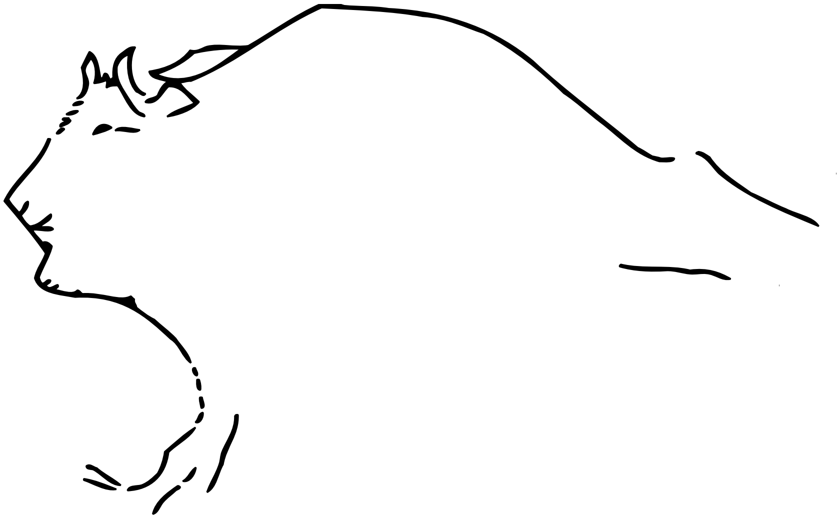
4. Grotte de Niaux (Ariège, France). Painting. Magdalenian period (~17,000 ± 260 calBP). (adapted from H. Breuil[4])

5. Grotte de Fontanet (Ariège, France). Carving. Magdalenian period (between ~14250 ± 295 calBP and ~16,600 ± 1000 calBP). (adapted from A. Glory[3])

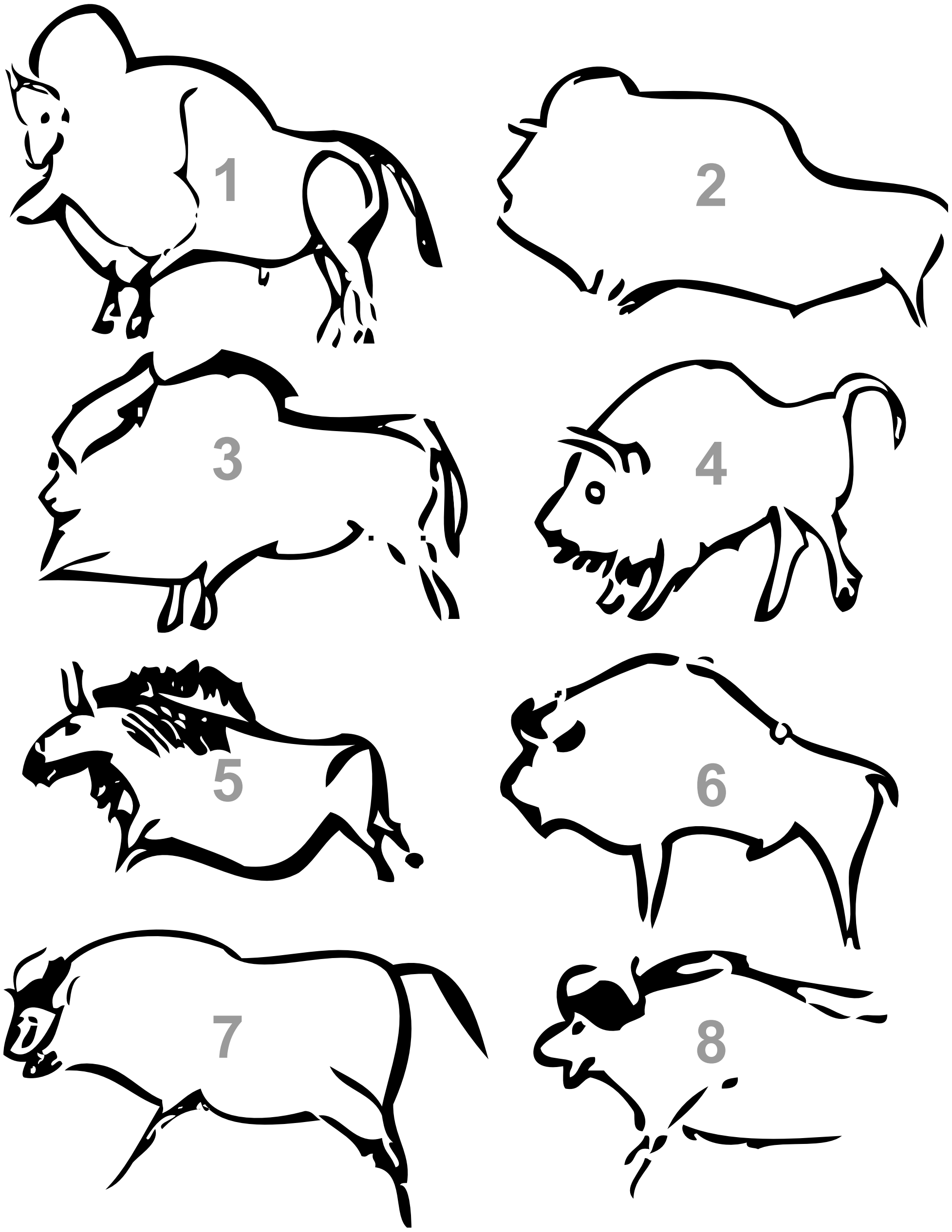
6. Grotte de Rouffignac (Dordogne, France). Painting. Magdalenian period (dating estimated based on stylistic analysis). (adapted from C. Barrière[6])

7. Grotte des Combarelles (Dordogne, France). Carving. Magdalenian period (between ~17,000 and ~14,300 calBP). (adapted from H. Breuil[4])

8. Grotte de Marsoulas (Haute-Garonne, France). Carving. Magdalenian period (dating estimated based on stylistic analysis). (from C. Fritz et G. Tosello)



Supplementary Fig 26. Bison carved on round stone from the Riparo di Tagliente site in Italy



Supplementary Fig 27. Undetermined morphologies. In European Palaeolithic art, some bison depictions show morphological traits and anatomical details that could be compatible with either bison form. These pictures illustrate the limits of cave art analyses for morphological assessment of bison forms, due to varying graphical conventions between cultures. Dates are given as indication based on archaeological occupation determined for each site, or, in the absence of such dating, based on stylistic comparison with other depictions:

1 Grotte de Font-de-Gaume (Dordogne, France). Black and red painting, and carving. Magdalenian period (dating estimated based on stylistic analysis). (adapted from H. Breuil[4])

2 Grotte de Niaux (Ariège, France). Painting. Magdalenian period (~17,000 ± 260 calBP). (adapted from H. Breuil[4])

3 Grotte des Trois Frères (Ariège, France). Carving. Magdalenian period (dating estimated based on stylistic analysis). (adapted from H. Breuil[4])

4 Grotte des Trois Frères (Ariège, France). Carving. Magdalenian period (dating estimated based on stylistic analysis). (adapted from H. Breuil[4])

5 Grotte des Trois Frères (Ariège, France). Carving. Gravettian period (dating estimated based on stylistic analysis). (adapted from H. Breuil[4])

6 Grotte de La Grèze (Dordogne, France). Carving. Gravettian period (dating estimated based on stylistic analysis) (adapted from N. Aujoulat)

7 Grotte Chauvet-Pont d'Arc (Ardèche, France). Blured black charcoal drawing. Aurignacian period (~35100 ± 175 calBP). (from C. Fritz-G. Tosello)

8 Grotte Chauvet-Pont d'Arc (Ardèche, France). Blured black charcoal drawing. Aurignacian period (~35100 ± 175 calBP). (from C. Fritz-G. Tosello)

Supplementary Tables

Supplementary Table 1. Primers and adapters used in this study

	Primer	Primer Sequence (5' - 3')	Length (a)
Set_ A1	BovCR-16351F	CAACCCCCAAAGCTGAAG	~96bp
	BovCR-16457R	TGGTTRGGGTACAAAGTCTGTG	
Set_ B1	BovCR-16420F	CCATAAATGCAAAGAGCCTCAYCAG	~172bp
	BovCR-16642R	TGCATGGGGCATATAATTTAATGTA	
Set_ A2	BovCR-16507F	AATGCATTACCCAAACRGGG	~184bp
	BovCR-16755R	ATTAAGCTCGTGATCTARTGG	
Set_ B2	BovCR-16633F(b)	GCCCCATGCATATAAGCAAG	~132bp
	BovCR-16810R(b)	GCCTAGCGGGTTGCTGGTTTCACGC	
Set_ A3	BovCR-16765F(b)	GAGCTTAAYTACCATGCCG	~125bp
	BovCR-16998R	CGAGATGTCTTATTTAAGAGGAAAGAATGG	
Set_ B3	BovCR-16960F	CATCTGGTTCTTTCTTCAGGGCC	~110bp
	BovCR-80R(b)	CAAGCATCCCCCAAAATAAA	
Frag1	BovCR_16738MF(c,d)	***CACGACGTTGTAAAACGAC***ATYGTACATAGYACATTATGTCAA	~67bp
	BovCR_16810TR(c,d)	***TACGACTCACTATAGGGCGA***GCCTAGCGGGTTGCTGGTTTCACGC	
Frag2	Mamm_12SE(d)	CTATAATCGATAAACCCCGATA	~96bp
	Mamm_12SH(d)	GCTACACCTTGACCTAAC	
	GAII_Indexing_x	CAAGCAGAAGACGGCATACGAGATNNNNNNNGAGTGACTGGAGTTCAGACGTGT	n/a
	IS4_indPCR.P5(e)	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT	n/a
	IS7_short_amp.P5(e)	ACACTCTTTCCCTACACGAC	n/a
	IS8_short_amp.P7(e)	GTGACTGGAGTTCAGACGTGT	n/a
	P5_short_RNAblock	ACACUCUUUCCCUACACGAC	n/a
	P7_short_RNAblock	GUGACUGGAGUUCAGACGUGU	n/a
	Bison_mt1_forward (f)	ACCGCGGTCATACGATTAAC	
	Bison_mt1_reverse (f)	AATTGCGAAGTGGATTTTGG	
	Bison_mt2_forward (f)	ATGAGCCAAAATCCACTTCG	
	Bison_mt2_reverse (f)	TGTATTTGCGTCTGCTCGTC	
	Bison_mt3_forward (f)	CGAATCCACAGCCGAACTAT	
	Bison_mt3_reverse (f)	TATAAAGCACCGCCAAGTCC	

(a): Primers are excluded from the length of PCR amplicon.

(b):[2].

(c): M13 (CAC GAC GTT GTA AAA CGA C) and T7 (TAC GAC TCA CTA TAG GGC GA) sequences were used as tags for primers BovCR_16738F and BovCR_16810R, respectively. This was done to obtain good quality Sanger sequences from short amplicons.

(d): One-step simplex PCRs.

(e): (Meyer and Kircher, "Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing.")

(f): Primer pairs for use to generate DNA baits for mitochondrial DNA capture.

Supplementary Table 2. Summary of nuclear alleles detected at bovine SNP loci: NGS results and locus counts for ancient samples; locus counts for modern samples

		Mapping results for the 9908 SNP positions						Number of SNP called out of the 9908 targeted for each ancient individuals							
									Coverage depth >=1				Coverage depth >=2		
Sample ID	**Method**	**Retained_reads**	**hits_raw**	**hits_unique**	**hits_raw_frac**	**hits_clonality**	**Mean coverage**	**Total**	REF/REF	REF/ALT	ALT/ALT	**Total**	REF/REF	REF/ALT	ALT/ALT
A15526		7045	1821	99	0.26	0.95	0.01	**49**	49	0	0	**1**	1	0	0
A017		1280556	3893	1289	0.00	0.67	0.13	**630**	591	0	39	**88**	49	0	39
A018		967346	3116	538	0.00	0.83	0.05	**253**	241	0	12	**28**	16	0	12
A001		656008	392937	3486	0.60	0.99	0.35	**1484**	1268	2	214	**523**	307	2	214
A003		1706985	12957	3423	0.01	0.74	0.35	**1569**	1363	5	201	**470**	264	5	201
A004	10k capture	240370	132883	645	0.55	1.00	0.07	**315**	287	0	28	**64**	36	0	28
A005		1736500	25788	3519	0.01	0.86	0.35	**1643**	1438	7	198	**464**	259	7	198
A006		10413909	99392	22312	0.01	0.78	2.25	**5690**	3468	104	2118	**4755**	2533	104	2118
A007		3583539	23832	2841	0.01	0.88	0.29	**1307**	1084	1	222	**509**	286	1	222
A15654		1700840	1227601	220913	0.72	0.82	22.28	**8738**	4532	230	3976	**8488**	4282	230	3976
A4093		9400283	62631	4478	0.01	0.93	0.45	**1946**	1480	2	464	**1031**	565	2	464
A3133	Shotgun / 10k	299829433	9812523	465082	0.03	0.95	46.87	**8898**	4579	321	3998	**8680**	4361	321	3998
A875	and 40k capture	3908972	291640	234493	0.07	0.20	23.65	**8433**	4341	342	3750	**8144**	4052	342	3750
CPC98_Aurochs	From published genome							**8882**	4770	1808	2304	**8810**	4698	1808	2304

Supplementary Table 3. Summary statistics for NGS of whole mitochondrial genomes

Sample ID	Retained_reads	hits_raw	hits_unique	hits_raw_frac	hits_clonality	AVG_Depth	STD_Depth	AVE_Length	STD_Length	5pC>T	3pG>A	Library repair
A001	4822143	1618364	86944	0.34	0.95	432.09	224.83	80.82	37.60	0.03	0.02	USER
A004	5150804	2314449	220697	0.45	0.90	1152.17	541.88	84.88	36.11	0.02	0.02	
A018	3790161	1021750	24699	0.27	0.98	130.53	60.04	85.32	34.05	0.03	0.03	
A4089	8618722	5380606	44044	0.62	0.99	237.83	155.46	87.18	33.56	0.02	0.02	
A3133	66864927	1958	1949	0.00	0.00	11.41	6.77	93.92	29.66	0.00	0.01	
A003	985033	371605	64372	0.38	0.83	334.44	112.68	84.31	34.07	0.08	0.07	Partial UDG
A005	521428	262622	39121	0.50	0.85	196.95	65.76	81.59	30.96	0.05	0.09	
A006	456078	120668	44541	0.26	0.63	208.39	93.86	75.86	25.87	0.13	0.17	
A007	431113	175432	43269	0.41	0.75	192.35	85.93	71.74	24.13	0.11	0.08	
A4093	212315	106221	16923	0.50	0.84	73.23	31.26	70.48	24.60	0.07	0.09	
A15637	469884	4401	2621	0.01	0.40	8.85	7.22	50.41	12.17	0.41	0.35	
A15654	294965	29628	28329	0.10	0.04	170.48	89.68	98.23	34.91	0.05	0.02	
A15668	230709	3603	2842	0.02	0.21	11.07	7.80	59.61	15.06	0.07	0.06	
LE237	507023	4271	2677	0.01	0.37	9.84	5.70	58.98	23.99	0.55	0.51	None
LE242	6912671	48793	35418	0.01	0.27	120.46	67.86	55.09	18.68	0.61	0.60	
LE257	4156307	184236	28788	0.04	0.84	94.38	38.34	53.17	20.00	0.52	0.50	

Supplementary Table 4. List of published mitochondrial control region sequences used for phylogenetic analysis. The Urals steppe bison are highlighted in red.

American bison

Bison_bison_AF083357_H1_0_0
Bison_bison_AF083358_H2_0_0
Bison_bison_AF083359_H3_0_0
Bison_bison_AF083360_H4_0_0
Bison_bison_AF083361_H5_0_0
Bison_bison_AF083362_H6_0_0
Bison_bison_AF083363_H7_0_0
Bison_bison_AF083364_H8_0_0
Bison_bison_BS100_29_5
Bison_bison_BS102_22_5
Bison_bison_BS129_0_2000
Bison_bison_BS162_AK_170_30
Bison_bison_BS173_NTC_3220_45
Bison_bison_BS175_ICE_186_30
Bison_bison_BS177_NTC_3155_36
Bison_bison_BS200_AB_145_37
Bison_bison_BS342_CHL_10340_40
Bison_bison_BS348_CHL_10505_45
Bison_bison_BS368_0_2000
Bison_bison_BS417_AB_909_29
Bison_bison_BS419_AB_7475_45
Bison_bison_BS421_AB_8145_45
Bison_bison_BS422_AB_908_31
Bison_bison_BS423_AB_4660_38
Bison_bison_BS424_AB_202_32
Bison_bison_BS426_AB_7060_45
Bison_bison_BS428_AB_7105_45
Bison_bison_BS429_AB_6775_40
Bison_bison_BS430_9270_50
Bison_bison_BS432_AB_7310_45
Bison_bison_BS433_AB_10450_55
Bison_bison_BS434_AB_809_32
Bison_bison_BS439_AB_5845_45
Bison_bison_BS441_AB_1273_32
Bison_bison_BS444_AB_636_29
Bison_bison_BS445_AB_378_30
Bison_bison_BS449_6195_45
Bison_bison_BS454_AB_287_29
Bison_bison_BS456_AB_125_30
Bison_bison_BS460_AB_10425_50
Bison_bison_BS464_AB_5205_45
Bison_bison_BS465_AB_7115_50
Bison_bison_BS466_AB_3298_37
Bison_bison_BS503_BlR_2776_36
Bison_bison_BS560_AB_2807_28
Bison_bison_BS569_AB_3600_70
Bison_bison_BS570_AB_11300_290
Bison_bison_BS99_26_5
Bison_bison_U12935_0_0
Bison_bison_U12936_0_0
Bison_bison_U12941_0_0
Bison_bison_U12943_0_0
Bison_bison_U12944_0_0
Bison_bison_U12945_0_0
Bison_bison_U12946_0_0
Bison_bison_U12947_0_0
Bison_bison_U12948_0_0
Bison_bison_U12955_0_0
Bison_bison_U12956_0_0
Bison_bison_U12957_0_0
Bison_bison_U12958_0_0
Bison_bison_U12959_0_0

Steppe bison

Bison_priscus_A3133_Yukon_26360_220
Bison_priscus_BS105_F_23380_460
Bison_priscus_BS107_F_19570_290
Bison_priscus_BS108_F_21020_360
Bison_priscus_BS109_F_20730_350
Bison_priscus_BS111_F_21580_370
Bison_priscus_BS121_F_19360_280
Bison_priscus_BS123_BlR_1730_60
Bison_priscus_BS124_BlR_11900_70
Bison_priscus_BS125_F_27440_790
Bison_priscus_BS126_F_19150_280
Bison_priscus_BS130_BlR_9000_250
Bison_priscus_BS133_F_33800_1900
Bison_priscus_BS145_NS_12270_50
Bison_priscus_BS146_NS_11810_50
Bison_priscus_BS147_NS_28120_290
Bison_priscus_BS148_NS_6400_50
Bison_priscus_BS149_NS_46100_2200
Bison_priscus_BS150_NS_10510_50
Bison_priscus_BS151_NS_21530_130
Bison_priscus_BS161_NS_21040_120
Bison_priscus_BS163_LC_13240_75
Bison_priscus_BS164_LC_19540_120
Bison_priscus_BS165_LC_26460_160
Bison_priscus_BS170_YT_13040_70
Bison_priscus_BS172_LC_12525_70
Bison_priscus_BS176_LC_12380_60
Bison_priscus_BS178_LC_17960_90
Bison_priscus_BS192_F_26300_300
Bison_priscus_BS193_NS_49600_4000
Bison_priscus_BS195_NS_29040_340
Bison_priscus_BS196_NS_19420_100
Bison_priscus_BS198_Y_2460_40
Bison_priscus_BS201_Y_12960_60
Bison_priscus_BS202_AB_10460_65
Bison_priscus_BS206_Sibh_23780_140
Bison_priscus_BS211_Sibh_43800_1100
Bison_priscus_BS216_NS_47000_2900
Bison_priscus_BS218_Si_14605_75
Bison_priscus_BS222_NWT_6110_45
Bison_priscus_BS223_Si_53300_1900
Bison_priscus_BS224_AK_13125_75
Bison_priscus_BS233_SW_16685_80
Bison_priscus_BS235_BlR_43400_900
Bison_priscus_BS236_SW_19420_100
Bison_priscus_BS237_AB_11240_70
Bison_priscus_BS243_SW_37550_400
Bison_priscus_BS244_LC_26210_170
Bison_priscus_BS248_OCr_12350_70
Bison_priscus_BS249_F_39200_550
Bison_priscus_BS253_LC_12665_65
Bison_priscus_BS254_CHL_10230_55
Bison_priscus_BS258_F_22120_130
Bison_priscus_BS260_D_30750_290
Bison_priscus_BS261_LC_12915_70
Bison_priscus_BS262_D_29150_500
Bison_priscus_BS281_BlR_40800_600
Bison_priscus_BS282_Si_56700_3200
Bison_priscus_BS284_Y_13135_65
Bison_priscus_BS286_Sim_49500_1300
Bison_priscus_BS287_BlR_49100_1700
Bison_priscus_BS289_BlR_2172_37
Bison_priscus_BS291_NS_49700_1400
Bison_priscus_BS292_NS_35710_730
Bison_priscus_BS294_BlR_58200_3900
Bison_priscus_BS297_NS_10990_50
Bison_priscus_BS311_BlR_12425_45
Bison_priscus_BS316_SI_57700_3000
Bison_priscus_BS318_NS_12410_50
Bison_priscus_BS320_SI_49600_1500
Bison_priscus_BS321_AK_9506_38
Bison_priscus_BS323_SI_37810_380
Bison_priscus_BS327_D_31530_230
Bison_priscus_BS328_SIdy_31690_180
Bison_priscus_BS329_D_27060_190
Bison_priscus_BS337_CHL_10378_36
Bison_priscus_BS340_NS_24500_180
Bison_priscus_BS345_NS_39800_1200
Bison_priscus_BS350_NS_38700_1000
Bison_priscus_BS351_BlR_57700_3200
Bison_priscus_BS359_NTC_20020_150
Bison_priscus_BS364_NS_38800_1100
Bison_priscus_BS365_NS_47000_2900
Bison_priscus_BS387_NS_33320_540
Bison_priscus_BS388_NS_27590_280
Bison_priscus_BS389_NS_17160_80
Bison_priscus_BS390_NS_31630_440
Bison_priscus_BS392_NS_36320_780
Bison_priscus_BS393_NS_39850_1200
Bison_priscus_BS394_NS_37460_890
Bison_priscus_BS395_NS_40700_1300
Bison_priscus_BS396_NS_23680_170
Bison_priscus_BS397_NS_32370_470
Bison_priscus_BS398_NS_27400_260
Bison_priscus_BS400_NS_46100_2600
Bison_priscus_BS405_SI_23040_120
Bison_priscus_BS407_NWT_55500_3100
Bison_priscus_BS412_Y_30500_250
Bison_priscus_BS414_BlR_4495_60
Bison_priscus_BS415_D_30810_975
Bison_priscus_BS418_China_26560_670
Bison_priscus_BS438_AB_53800_2200
Bison_priscus_BS440_AB_60400_2900
Bison_priscus_BS443_AB_34050_450
Bison_priscus_BS459_China_47700_1000
Bison_priscus_BS469_AB_305_24
Bison_priscus_BS472_F_13235_65
Bison_priscus_BS473_AB_56300_3100
Bison_priscus_BS477_D_33710_240
Bison_priscus_BS478_D_34470_200
Bison_priscus_BS490_BlR_2415_25
Bison_priscus_BS493_NS_50000_4200
Bison_priscus_BS494_NS_44800_2200
Bison_priscus_BS495_NS_29570_340
Bison_priscus_BS497_NS_30000_540
Bison_priscus_BS498_NS_25980_230
Bison_priscus_BS499_NS_31410_420
Bison_priscus_BS500_NS_35580_550
Bison_priscus_BS517_BlR_2526_26
Bison_priscus_BS564_Si_24570_90
Bison_priscus_BS571_SIdy_32910_170
Bison_priscus_BS592_Urals_42500_450
Bison_priscus_BS605_NTC_20380_90
Bison_priscus_BS660_Urals_29500_140
Bison_priscus_BS662_SI_20000_0
Bison_priscus_BS674_Urals_29060_140
Bison_priscus_BS708_Urals_47050_750
Bison_priscus_BS713_Urals_30970_180
Bison_priscus_IB179_LC_12465_75

European bison

Bison_bonasus_AF083356_0_0
Bison_bonasus_AY428860_0_0
Bison_bonasus_EF693811_0_0
Bison_bonasus_EU272053_0_0
Bison_bonasus_EU272054_0_0
Bison_bonasus_EU272055_0_0
Bison_bonasus_U12953_0_0
Bison_bonasus_U12954_0_0
Bison_bonasus_U34294_0_0

Yak

Bos_grunniens_AY521140_0_0
Bos_grunniens_AY521149_0_0
Bos_grunniens_AY521150_0_0
Bos_grunniens_AY521151_0_0
Bos_grunniens_AY521152_0_0
Bos_grunniens_AY521154_0_0
Bos_grunniens_AY521155_0_0
Bos_grunniens_AY521156_0_0
Bos_grunniens_AY521160_0_0
Bos_grunniens_AY521161_0_0
Bos_grunniens_DQ007210_0_0
Bos_grunniens_DQ007221_0_0
Bos_grunniens_DQ007222_0_0
Bos_grunniens_DQ856594_0_0
Bos_grunniens_DQ856599_0_0
Bos_grunniens_DQ856600_0_0
Bos_grunniens_DQ856603_0_0
Bos_grunniens_DQ856604_0_0
Bos_grunniens_EF494177_0_0
Bos_grunniens_EF494178_0_0

Zebu

Bos_indicus_AB085923_0_0
Bos_indicus_AB268563_0_0
Bos_indicus_AB268564_0_0
Bos_indicus_AB268566_0_0
Bos_indicus_AB268571_0_0
Bos_indicus_AB268574_0_0
Bos_indicus_AB268578_0_0
Bos_indicus_AB268580_0_0
Bos_indicus_AY378134_0_0
Bos_indicus_AY378135_0_0
Bos_indicus_DQ887765_0_0
Bos_indicus_EF417971_0_0
Bos_indicus_EF417974_0_0
Bos_indicus_EF417976_0_0
Bos_indicus_EF417977_0_0
Bos_indicus_EF417979_0_0
Bos_indicus_EF417981_0_0
Bos_indicus_EF417983_0_0
Bos_indicus_EF417985_0_0
Bos_indicus_EF524120_0_0
Bos_indicus_EF524125_0_0
Bos_indicus_EF524126_0_0
Bos_indicus_EF524128_0_0
Bos_indicus_EF524130_0_0
Bos_indicus_EF524132_0_0
Bos_indicus_EF524135_0_0
Bos_indicus_EF524141_0_0
Bos_indicus_EF524152_0_0
Bos_indicus_EF524156_0_0
Bos_indicus_EF524160_0_0
Bos_indicus_EF524166_0_0
Bos_indicus_EF524167_0_0
Bos_indicus_EF524170_0_0
Bos_indicus_EF524177_0_0
Bos_indicus_EF524180_0_0
Bos_indicus_EF524183_0_0
Bos_indicus_EF524185_0_0
Bos_indicus_L27732_0_0
Bos_indicus_L27736_0_0

Aurochs

Bos_primigenius_DQ915522_ALL1_12030_52
Bos_primigenius_DQ915523_CAT1_5650_0
Bos_primigenius_DQ915524_CHWF_3905_185
Bos_primigenius_DQ915537_CPC98_5936_34
Bos_primigenius_DQ915542_EIL06_5830_29
Bos_primigenius_DQ915543_EIL14_5830_29
Bos_primigenius_DQ915554_LJU3_8020_50
Bos_primigenius_DQ915558_NORF_3370_30
Bos_primigenius_EF187280_PVL04_3204_56

Cattle

Bos_taurus_DQ124372_T4_0_0
Bos_taurus_DQ124375_T4_0_0
Bos_taurus_DQ124381_T3_0_0
Bos_taurus_DQ124383_T2_0_0
Bos_taurus_DQ124388_T3_0_0
Bos_taurus_DQ124394_T3_0_0
Bos_taurus_DQ124398_T3_0_0
Bos_taurus_DQ124400_T4_0_0
Bos_taurus_DQ124401_T4_0_0
Bos_taurus_DQ124412_T4_0_0
Bos_taurus_EU177822_T3_0_0
Bos_taurus_EU177841_T1_0_0
Bos_taurus_EU177842_T1_0_0
Bos_taurus_EU177845_T1_0_0
Bos_taurus_EU177847_T1_0_0
Bos_taurus_EU177848_T1_0_0
Bos_taurus_EU177853_T2_0_0
Bos_taurus_EU177854_T2_0_0
Bos_taurus_EU177860_T2_0_0
Bos_taurus_EU177861_T2_0_0
Bos_taurus_EU177862_T5_0_0
Bos_taurus_EU177863_T5_0_0
Bos_taurus_EU177864_T5_0_0
Bos_taurus_EU177865_T5_0_0

Buffalo

Bubalus_bubalis_AF197208_0_0
Bubalus_bubalis_AF475212_0_0
Bubalus_bubalis_AF475256_0_0
Bubalus_bubalis_AF475259_0_0
Bubalus_bubalis_AF475278_0_0
Bubalus_bubalis_AY488491_0_0
Bubalus_bubalis_EF536327_0_0
Bubalus_bubalis_EF536328_0_0
Bubalus_bubalis_EU268899_0_0
Bubalus_bubalis_EU268909_0_0

Supplementary Table 5. List of published whole mitochondrial genome sequences used for phylogenetic analysis.

American bison	**Cattle**	**Yak**
GU947000_Bison_bison_Plains_Nebraska_0	FJ971080_Bos_Q_Italy_Romagnola_0	KJ704989_Bos_grunniens_ChinaGansu_Gannan_0
GU946976_Bison_bison_Plains_Montana_0	FJ971085_Bos_R_Italy_Cinisara_0	KR011113_Bos_grunniens_ChinaTibet_QinghaiPlateau_0
GU947004_Bison_bison_Plains_Wyoming_0	EU177841_Bos_T1_Italy_chianina_0	KR052524_Bos_grunniens_ChinaTibet_Pali_0
GU947006_Bison_bison_Wood_ElkIsland_0	DQ124383_Bos_T2_Korea_0	KJ463418_Bos_grunniens_ChinaQinghai_Dantong_0
GU946987_Bison_bison_Plains_Montana_0	EU177815_Bos_T3_Italy_piemontese_0	KM233417_Bos_mutus_ChinaTibet_Yakow_0
GU947005_Bison_bison_Wood_ElkIsland_0	DQ124372_Bos_T4_Korea_0	**Buffalo**
GU947002_Bison_bison_Plains_Texas_0	EU177862_Bos_T5_Italy_valdostana_0	GU947003_Bison_bison_Plains_Texas_0
GU947003_Bison_bison_Plains_Texas_0	**Aurochs**	AY488491_Bubalus_bubalis
Wisent	GU985279_Bos_P_England_6760	AY702618_Bubalus_bubalis
JN632602_Bison_bonasus_0	JQ437479_Bos_P_Poland_1500	AF547270_Bubalus_bubalis
HQ223450_Bison_bonasus_0	**Zebu**	
HM045017_Bison_bonasus_Poland_0	FJ971088_Bos_I1_Mongolia_0	
Steppe bison	EU177870_Bos_I2_Iran_0	
KM593920_Bison_priscus_SGE2_France_TroisFreres_19151		

Supplementary Table 6. f4 ratio estimates, f4(A,O,X,C) is the numerator, f4(A,O,B,C) is the denominator.

S6-A. Including heterozygotes

A	**O**	**X**	**C**	**:**	**A**	**O**	**B**	**C**	**alpha**	**std.err**	**Z**
AmericanBison	Ovis_aries	AllWisent+CladeX	Aurochs	:	AmericanBison	Ovis_aries	Steppe	Aurochs	0.890988	0.025788	34.551
AmericanBison	Ovis_aries	AllWisent+CladeX	Steppe	:	AmericanBison	Ovis_aries	Aurochs	Steppe	0.109012	0.025788	4.227
AmericanBison	Ovis_aries	AllWisent	Aurochs	:	AmericanBison	Ovis_aries	Steppe	Aurochs	0.884257	0.02918	30.304
AmericanBison	Ovis_aries	AllWisent	Steppe	:	AmericanBison	Ovis_aries	Aurochs	Steppe	0.115743	0.02918	3.967
AmericanBison	Ovis_aries	CladeX	Aurochs	:	AmericanBison	Ovis_aries	Steppe	Aurochs	0.893978	0.022763	39.273
AmericanBison	Ovis_aries	CladeX	Steppe	:	AmericanBison	Ovis_aries	Aurochs	Steppe	0.106022	0.022763	4.658
AmericanBison	Ovis_aries	AncientWisent	Aurochs	:	AmericanBison	Ovis_aries	Steppe	Aurochs	0.812638	0.054701	14.856
AmericanBison	Ovis_aries	AncientWisent	Steppe	:	AmericanBison	Ovis_aries	Aurochs	Steppe	0.187362	0.054701	3.425
AmericanBison	Ovis_aries	HistoricalWisent	Aurochs	:	AmericanBison	Ovis_aries	Steppe	Aurochs	0.773802	0.032319	23.943
AmericanBison	Ovis_aries	HistoricalWisent	Steppe	:	AmericanBison	Ovis_aries	Aurochs	Steppe	0.226198	0.032319	6.999
AmericanBison	Ovis_aries	ModernWisent	Aurochs	:	AmericanBison	Ovis_aries	Steppe	Aurochs	0.899149	0.031184	28.834
AmericanBison	Ovis_aries	ModernWisent	Steppe	:	AmericanBison	Ovis_aries	Aurochs	Steppe	0.100851	0.031184	3.234

S6-B. Haploidisation by randomly sampling an allele at heterozygous sites

A	**O**	**X**	**C**	**:**	**A**	**O**	**B**	**C**	**alpha**	**std.err**	**Z**
AmericanBison	Ovis_aries	AllWisent+CladeX	Aurochs	:	AmericanBison	Ovis_aries	Steppe	Aurochs	0.894329	0.027147	32.944
AmericanBison	Ovis_aries	AllWisent+CladeX	Steppe	:	AmericanBison	Ovis_aries	Aurochs	Steppe	0.105671	0.027147	3.893
AmericanBison	Ovis_aries	AllWisent	Aurochs	:	AmericanBison	Ovis_aries	Steppe	Aurochs	0.88342	0.030518	28.947
AmericanBison	Ovis_aries	AllWisent	Steppe	:	AmericanBison	Ovis_aries	Aurochs	Steppe	0.11658	0.030518	3.82
AmericanBison	Ovis_aries	CladeX	Aurochs	:	AmericanBison	Ovis_aries	Steppe	Aurochs	0.912424	0.025204	36.202
AmericanBison	Ovis_aries	CladeX	Steppe	:	AmericanBison	Ovis_aries	Aurochs	Steppe	0.087576	0.025204	3.475
AmericanBison	Ovis_aries	AncientWisent	Aurochs	:	AmericanBison	Ovis_aries	Steppe	Aurochs	0.813521	0.059078	13.77
AmericanBison	Ovis_aries	AncientWisent	Steppe	:	AmericanBison	Ovis_aries	Aurochs	Steppe	0.186479	0.059078	3.156
AmericanBison	Ovis_aries	HistoricalWisent	Aurochs	:	AmericanBison	Ovis_aries	Steppe	Aurochs	0.786183	0.035363	22.232
AmericanBison	Ovis_aries	HistoricalWisent	Steppe	:	AmericanBison	Ovis_aries	Aurochs	Steppe	0.213817	0.035363	6.046
AmericanBison	Ovis_aries	ModernWisent	Aurochs	:	AmericanBison	Ovis_aries	Steppe	Aurochs	0.899281	0.032252	27.883
AmericanBison	Ovis_aries	ModernWisent	Steppe	:	AmericanBison	Ovis_aries	Aurochs	Steppe	0.100719	0.032252	3.123

Supplementary Table 7: Bootstrap resampling of genotypes for testing topologies using D statistics. The table shows the fraction of bootstrap replicates for which the original result was not recapitulated, from 10000 bootstraps, for 10%, 20%, etc. subsets of the genotypes. A topology is considered to be simple if it either has a non-significant D statistic (see Supplementary Figure 11), or has a D statistic closest to zero with confidence intervals that do not overlap the D statistic for the other two topologies.

Most parsimonious topology	Simple topology	10%	20%	30%	40%	50%	60%	70%	80%	90%
((CladeX, Steppe), ModernWisent)	True	0.0067	0.0001	0.0	0.0	0.0	0.0	0.0	0.0	0.0
((Steppe, HistoricalWisent), ModernWisent)	False	0.0575	0.0573	0.0284	0.0036	0.0005	0.0	0.0	0.0	0.0
((ModernWisent, CladeX), HistoricalWisent)	False	0.1753	0.371	0.485	0.4427	0.3039	0.1564	0.0549	0.0072	0.0

((CladeX, Steppe), HistoricalWisent)	True	0.0182	0.0174	0.0154	0.016	0.0113	0.0072	0.0022	0.0004	0.0
((AncientWisent, HistoricalWisent), ModernWisent)	True	0.0565	0.0152	0.0042	0.0012	0.0	0.0	0.0	0.0	0.0
((Steppe, HistoricalWisent), AncientWisent)	False	0.0151	0.0039	0.0001	0.0002	0.0	0.0	0.0	0.0	0.0
((AncientWisent, Steppe), ModernWisent)	True	0.0484	0.0086	0.0014	0.0002	0.0	0.0	0.0	0.0	0.0
((CladeX, Steppe), AncientWisent)	False	0.0304	0.0142	0.0086	0.0063	0.0033	0.0025	0.0015	0.0001	0.0
((AncientWisent, CladeX), ModernWisent)	True	0.0703	0.0213	0.0062	0.0015	0.0007	0.0	0.0	0.0	0.0
((HistoricalWisent, CladeX), AncientWisent)	False	0.0184	0.0053	0.001	0.0005	0.0	0.0	0.0	0.0	0.0
((ModernWisent, HistoricalWisent), Aurochs)	False	0.0591	0.0031	0.0005	0.0	0.0	0.0	0.0	0.0	0.0
((Aurochs, ModernWisent), CladeX)	False	0.2229	0.2476	0.0824	0.0115	0.0009	0.0	0.0	0.0	0.0
((HistoricalWisent, CladeX), Aurochs)	True	0.0061	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
((Steppe, CladeX), Aurochs)	True	0.0001	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
((Steppe, HistoricalWisent), Aurochs)	True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
((Steppe, ModernWisent), Aurochs)	False	0.1362	0.0535	0.0048	0.0007	0.0002	0.0	0.0001	0.0	0.0
((Steppe, AncientWisent), Aurochs)	True	0.0441	0.0082	0.0001	0.0001	0.0	0.0	0.0	0.0	0.0
((AncientWisent, CladeX), Aurochs)	True	0.0276	0.0058	0.0004	0.0001	0.0	0.0	0.0	0.0	0.0

Supplementary Table 8: Hypergeometric test for shared derived steppe alleles. Steppe derived sites were filtered for coverage depth in the wisent lineages 1 and 2, for which the test was performed. In the last row, wisent represents all wisent other than CladeX.

1	2	Steppe	Derived 1	Derived 2	Common	P
Ancient Wisent	CladeX	161	111	133	108	1.72E-12
Ancient Wisent	Historical Wisent	174	115	119	108	1.37E-24
Ancient Wisent	Modern Wisent	178	124	108	95	5.12E-11
CladeX	Historical Wisent	529	448	385	370	3.09E-29
CladeX	Modern Wisent	556	469	350	326	2.79E-13
Historical Wisent	Modern Wisent	618	436	372	342	5.50E-48
Wisent	CladeX	557	357	468	332	4.18E-14

Supplementary Table 9: Hypergeometric test for shared derived aurochs alleles. Aurochs derived sites were filtered for coverage depth in the wisent lineages 1 and 2, for which the test was performed. In the last row, wisent represents all wisent other than CladeX.

1	**2**	**Aurochs**	**Derived 1**	**Derived 2**	**Common**	**P**
Ancient Wisent	CladeX	758	20	9	4	4.11E-05
Ancient Wisent	Historical Wisent	822	22	11	8	1.01E-11
Ancient Wisent	Modern Wisent	826	25	22	12	1.49E-14
CladeX	Historical Wisent	2517	36	47	16	7.34E-20
CladeX	Modern Wisent	2580	39	73	15	1.99E-14
Historical Wisent	Modern Wisent	2845	58	83	39	2.66E-50
Wisent	CladeX	2634	93	41	15	1.58E-12

Supplementary Table 10: The weighted sample median $\hat{M}$, the weighted sample mode $\hat{Mo}$, and the prediction error E_{pred}, for each ABC analysis.

Trio	$\hat{\mathbf{M}}$	$\hat{\mathbf{Mo}}$	$\mathbf{E}_{\mathbf{pred}}$
A875, 6A, Aurochs	0.8660	0.9204	0.4534
A3133, 6A, Aurochs	0.8480	0.9172	0.4881
A875, Historical Wisent, Aurochs	0.8636	0.9323	0.4187
A3133, Historical Wisent, Aurochs	0.8646	0.9384	0.4921
All	0.8250	0.9034	0.5111

Supplementary Table 11: Empirical posterior probabilities for levels of hybridisation 1%-5%, for each trio.

Trio	**1%**	**2%**	**3%**	**4%**	**5%**
A875, 6A, Aurochs	0.9620	0.9340	0.8720	0.8400	0.8120
A3133, 6A, Aurochs	0.9600	0.9600	0.8840	0.8440	0.7980
A875, Historical Wisent, Aurochs	0.9660	0.9340	0.8860	0.8520	0.7940
A3133, Historical Wisent, Aurochs	0.9580	0.9100	0.8580	0.8080	0.7640
All	0.9720	0.9440	0.9140	0.8760	0.8760

Supplementary Note 1:

Samples, DNA extraction and sequencing

Samples and radiocarbon dating

For clarity purposes we kept the most commonly used taxonomic nomenclature of bovine throughout the study. Although not yet widely accepted, it has been proposed to sink the genus *Bison* into *Bos* based on the shallow time depth of their evolutionary history [7]. The validity of such genetic separation is further tested in this study.

Samples from a total of 87 putative bison bones were collected from 3 regions across Europe: Urals, Caucasus, and Western Europe (Supplementary Data 1). As shown in the Supplementary Data 1, most of the samples were from bones identified as bison or bovid post-cranial samples, because cranial material is rare for this time period.

The main set of samples, from northeastern Europe, represents isolated bones excavated from a wide variety of cave deposits throughout the Ural Mountains and surrounding areas. These samples are housed at the Zoological Museum of the Institute of Plant and Animal Ecology (ZMIPAE) in Ekaterinburg, Russia.

In southeastern Europe, bovid bone fragments were excavated in Mezmaiskaya Cave in the Caucasus Mountains. Samples were obtained from the Laboratory of Prehistory in St Petersburg. Additional six samples from the Caucasus are identified as Caucasian bison (B. bonasus caucasicus, hereafter referred to as historical wisent): two of them are from the National History Museum (NHM) in London, and four come from hunts in the Kuban Oblast in the early 20th century (one collected by scientist Viktor Iwanovich Worobjew in 1906 and three hunted during the Kuban Hunt under the Grand Duke Sergei Mikhailovich of Russia), currently held at the Zoological Institute of the Russian Academy of Sciences (ZIRAS - Saint Petersburg, Russia). Four additional bones from the Caucasus region comes from the eastern border with Ukraine and are held at the Institute of Archeology (IAKiev), Ukrainian Academy of Sciences, Kiev.

Most western European bones come from late Pleistocene deposits on the North Sea bed. These specimens, now curated by the North Sea Network (NSN) in the Netherlands, were recovered by trawling operations and as such have little stratigraphic information. Specimens were selected on the basis of their morphological similarities with the ‘small form’ described by Drees and Post [8].

Three bones held in the collections of the Vienna Natural History Museum (VNHM), and three bones held in the Museum National d’Histoire Naturelle (Paris) come from central European Holocene sites.

Finally, one bone comes from the Monti Lessini rock-shelter site Riparo Tagliente in the North of Italy, one bone comes from the Swiss site of Le Gouffre de la combe de la racine in the Jura mountains (Swiss Institute for Speleology and Karst Studies, ISSKA), and one bone comes from l’Aven de l’Arquet in the Gard region of France (Musée de Préhistoire d’Orgnac).

In addition, two samples from the Beringian region were used: one sample, a steppe bison astragalus from the Yukon territory (Canada), has previously been used in a study of cytosine methylation in ancient DNA [9]; and another steppe bison from Alyoshkina Zaimka in Siberia.

All non-contemporaneous samples from which bison mitochondrial control region sequences were successfully amplified were sent for accelerator mass spectrometry (AMS) radiocarbon dating (except for seven samples from level 3 of the Mezmaiskaya cave, which were expected to be older than AMS dating capabilities [10,11]). The dating was performed by the AMS facility at the Oxford Radiocarbon Accelerator Unit at the University of Oxford (OxA numbers), the Eidgenössische Technische Hochschule in Zürich for a Ukrainian sample (ETH number), and the Ångström Laboratory of the University of Uppsala, Sweden, for the Swiss sample (Ua number). The results are shown in Supplementary Data 1, with all dates reported in kcal yr BP unless otherwise stated. The calibration of radiocarbon dates was performed using OxCal v4.1 with the IntCal13 curve [12].

In addition, two bones identified as bison were previously dated at the Centre for Isotope Research, Radiocarbon Laboratory, University of Groningen, Netherlands, with infinite radiocarbon age, consistently with the dating performed at Oxford (A2808-JGAC26=GrA-34533; A2809-JGAC27= GrA-34524).

Ancient DNA extraction

All ancient DNA work was conducted in clean-room facilities at the University of Adelaide's Australian Centre for Ancient DNA, Australia (ACAD), and at the University of Tuebingen, Germany (UT) following published guidelines [13].

University of Adelaide:

Samples were UV irradiated (260 nm) on all surfaces for 30 min. Sample surface was wiped with 3% bleach, then ~1 mm was removed using a Dremel tool and carborundum cutting disks. Each sample was ground to a fine powder using a Mikro-Dismembrator (Sartorius). Two DNA extraction methods were used during the course of the project (see Supplementary Data 1for the method used for specific samples):

- *Phenol-chloroform method*: Ancient DNA was extracted from 0.2-0.5g powdered bone using phenol-chloroform and centrifugal filtration methods according to a previously published method [2].

- *In solution silica based method*: Ancient DNA was extracted from 0.2-0.3g powdered bone according to a previously published method [14].

University of Tuebingen:

Samples were UV-irradiated overnight to remove surface contamination. DNA extraction was performed following a guanidinium-silica based extraction method [15] using 50mg of bone powder. A DNA library was prepared using 20µl of extract for each sample according to [16]. Sample-specific indexes were added to both library adapters to differentiate between individual samples after pooling and multiplex sequencing [17]. Indexed libraries were amplified in 100µl reactions, followed by purification over Qiagen MinElute spin columns (Quiagen, Hilden, Germany).

Sequencing of the mitochondrial control region

A ~600 bp fragment of the mitochondrial control region was amplified in one or up to four overlapping fragments, depending on DNA preservation. PCR amplifications were performed using primers designed for the bovid mitochondrial control region, following the method described in [2].

One-step simplex PCR amplifications using Platinum *Taq* Hi-Fidelity polymerase were performed on a heated lid thermal cycler in a final volume of 25 µl containing 1 µl of aDNA extract, 1mg/ml rabbit serum albumin fraction V (RSA; Sigma-Aldrich, Sydeny, NSW), 2 mM $MgSO_4$ (Thermo Fisher, Scoresby VIC), 0.6 µM of each primer (Supplementary Table 1), 250 µM of each dNTP (Thermo Fisher), 1.25 U Platinum *Taq* Hi-Fidelity and 1 × Hi-Fidelity PCR buffer (Thermo Fisher). The conditions for PCR amplification were initial denaturation at 95°C for 2 min, followed by 50 cycles of 94°C for 20 sec, 55°C for 20 sec and 68°C for 30 sec, and a final extension at 68°C for 10 min at the end of the 50 cycles.

Multiplex primer sets A and B were set up separately (Supplementary Table 1). Multiplex PCR was performed in a final volume of 25 µl containing 2 µl of aDNA extract, 1 mg/ml RSA, 6 mM $MgSO_4$, 0.2 µM of each primer (Supplementary Table 1), 500 µM of each dNTP, 2 U Platinum *Taq* Hi-Fidelity and 1 × Hi-Fidelity PCR buffer. Multiplex PCR conditions were initial denaturation at 95°C for 2 min, followed by 35 cycles of 94°C for 15 sec, 55°C for 20 sec and 68°C for 30 sec, and a final extension at 68°C for 10 min at the end of the 35 cycles. Multiplex PCR products were then diluted to 1:10 as template for the second step of simplex PCR. The simplex PCR, using Amplitaq Gold (Thermo Fisher) or Hotmaster™ *Taq* DNA polymerase (5Prime, Milton, Qld), was conducted in a final volume of 25 µl containing 1 µl of diluted multiplex PCR product, 2.5 mM $MgCl_2$, 0.4 µM of each primer (Supplementary Table 1), 200 µM of each dNTP, 1 U Amplitaq Gold/Hotmaster *Taq* polymerase and 1 × PCR buffer. The PCR conditions were initial denaturation at 95°C for 2 min, followed by 35 cycles of 94°C for 20 sec, 55°C for 15 sec and 72°C for 30 sec, and a final extension at 72°C for 10 min at the end of the 35 cycles. Multiple PCR fragments were cloned to evaluate the extent of DNA damage and within-PCR template diversity.

PCR products were then checked by electrophoresis on 3.5-4.0% agarose TBE gels, and visualized after ethidium bromide staining on a UV transilluminator. PCR amplicons were purified using Agencourt® AMPure magnetic beads (Beckman Coulter, Lane Cove, NSW) according to the manufacturer's instructions. Negative extraction controls and non-template PCR controls were used in all experiments.

All purified PCR products were bi-directionally sequenced with the ABI Prism® BigDye™ Terminator Cycle Sequencing Kit version 3.1 (Thermo Fisher). The sequencing reactions were performed in a final volume of 10 µl containing 3.2 pmol of primer (Supplementary Table 1), 0.25 µl Bigdye terminator premixture, and 1.875 µl of 5 × sequencing buffer. The reaction conditions included initial denaturation at 95°C for 2 min, 25 cycles with 95°C for 10 sec, 55°C for 15 sec, and 60°C for 2 min 30 sec. Sequencing products were purified using Agencourt® Cleanseq magnetic beads (Beckman Coulter) according to the manufacturer's protocol. All sequencing reactions were analysed on an ABI 3130 DNA capillary sequencer (Thermo Fisher).

Mitochondrial control region sequences (>400bp) were successfully amplified from 65 out of 87 analysed samples. Three samples produced a mixture of cattle and bison amplification products; these were identified as contaminated and removed from all analyses. Sequences from two individuals did not match bovid haplotypes and were identified as brown bear and elk in BLAST searches (see Supplementary Data 1). This is presumably due to the source postcranial elements being morphologically ambiguous and misidentified.

Sequencing of the whole mitochondrial genome

To provide deeper phylogenetic resolution and further examine the apparent close relationship between *Bos* and wisent mitochondria, full mitogenome sequences of 13 CladeX specimens, as well as one ancient wisent, one historical wisent, and one steppe bison were generated using hybridisation capture with RNA probes.

Samples A001, A004, A018, A4089 (CladeX)

DNA library preparation

DNA repair and polishing were performed in a reaction that contained 20 µl DNA extract, 1x NEB Buffer 2 (New England Biolabs, Ipswich, MA), 3U USER enzyme cocktail (New England Biolabs), 20U T4 polynucleotide kinase (New England Biolabs), 1mM ATP, 0.1 mM dNTPs (New England Biolabs), 8 µg RSA, and H_2O to 38.5 µl. The reaction was incubated at 37°C for 3 hours then 4.5U of T4 DNA polymerase (New England Biolabs) was added and the reaction incubated at 25°C for a further 30 min. Double-stranded libraries were then built with truncated Illumina adapters containing dual 5-mer internal barcodes as in [16].

Amplification of Bos taurus mitochondrial in vitro transcription (IVT) templates

RNA probes were generated from long-range PCR products of *Bos taurus* mitochondrial DNA. The NCBI Primer-Blast program (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) was used to design primers to amplify the *Bos taurus* mitochondrial genome (NC_006853.1) in three overlapping sections: mito-1 (6568 bp), mito-2 (6467 bp), and mito-3 (5390 bp). Primer pairs were designed with a high melting temperature to permit amplification with 2-stage PCR and the T7 RNA promoter was attached to the 5' end of one primer from each pair [18](Supplementary Table 1). Amplification of each mitochondrial section was performed using a heated lid thermal cycler in multiple PCRs containing 1x Phire Buffer (Thermo Fisher), 25 ng calf thymus DNA (Affymetrix, Santa Clara, CA), 200 µM dNTPs , 500 nM forward and reverse primers, 0.5 µl Phire Hot Start II DNA polymerase (Thermo Fisher), and H_2O to 25 µl. The mito-1 and mito-2 sections were amplified with a thermal cycler program of 1 cycle: 98°C for 30 sec; 26 cycles: 98°C for 10 sec and 72°C for 70 sec; and 1 cycle: 72°C for 180 sec whilst the program for mito-3 was 1 cycle: 98°C for 30 sec, 28 cycles: 98°C for 10 sec and 72°C for 60 sec, and 1 cycle: 72°C for 180 sec. After amplification, 2 □l of each PCR was agarose gel electrophoresed and the product visualized with Gel-Red (Biotium, Hayward, CA) staining and UV illumination. Amplification of mito-1 and mito-2 produced a single band and the PCRs for these mitochondrial sections were separately pooled and then purified with QiaQuick columns (Qiagen, Chadstone Centre, VIC) following the provided PCR cleanup protocol. Amplification of mito-3 produced unwanted products and the correct size amplicon was size selected using gel excision followed by purification with QiaQuick columns using the gel extraction protocol. Purified amplicons from each mitochondrial section were quantified using a NanoDrop 2000 Spectrophotometer (Thermo Fisher).

Transcription of Bos taurus mitochondrial IVT templates

Each of the three mitochondrial IVT templates were transcribed using a T7 High Yield RNA Synthesis Kit (New England Biolabs) in multiple reactions containing 150-200 ng purified amplicon, 1x Reaction Buffer, 10 mM rNTPs, 2 µl T7 enzyme mix, and H_2O to 20 µl. The IVT reactions were incubated for 16 hours at 37°C and then the DNA template was destroyed by incubating for an additional 15 min at 37°C with 2U Turbo Dnase (Thermo Fisher). IVT reactions for each mitochondrial section were separately pooled and purified with Megaclear spin columns (Thermo Fisher) except that H_2O was used to elute the RNA instead of the provided elution buffer. The elution buffer provided with the Megaclear kit was found to inhibit fragmentation in the next step. Integrity of the RNA was verified on an acrylamide gel and the mass quantified with a Nanodrop 2000 Spectrophotometer.

Fragmentation of mitochondrial IVT RNA

RNAs from the IVT transcription were fragmented with a NEBNext Magnesium RNA Fragmentation Module (New England Biolabs) in reactions that contained 1x Fragmentation buffer, 45 µg RNA, and H_2O to 20 µl. Reactions were incubated at 94°C for 10 min and fragmentation stopped with the addition of 2 µl Stop Buffer. After fragmentation, each reaction was purified with a RNeasy MinElute spin column (Qiagen) by following the provided cleanup protocol except for the final elution. To elute, 20 µL H_2O was pipetted into the column and the column was heated at 65°C for 5 min and then centrifuged at 15,000 g for 1 min. The flow-through was transferred to a 1.5 ml tube and stored at -80°C. The fragmented RNA was quantified on a NanoDrop 2000 Spectrophotometer and 100 ng was visualized on an acrylamide gel producing a smear in the range of 80-300 bases.

Biotinylation of fragmented RNA

Biotinylation was performed in several reactions containing 6.7 µg each of mito-1, mito-2, and mito-3 fragmented RNA, 40 µl Photoprobe Long Arm (Vector Laboratories, Burlingame, CA), and H_2O to 80 µl in 200 µl PCR tubes. The tubes were placed in a 4°C gel cooling rack and then incubated under the bulb of a UV sterilization cabinet for 30 min. Organic extractions were performed on the labelling reactions by adding 64 µl H_2O, 16 µl 1 M Tris buffer, and 160 µl sec-butanol to each tube and shaking vigorously for 30 sec followed by centrifugation for 1 minute at 1000 g. The upper organic layers were discarded and the extraction repeated with an additional 160 µl sec-butanol. After the second organic layers were discarded, the remaining aqueous phases were purified with RNeasy MinElute spin columns following the provided reaction cleanup protocol but with a modified elution procedure described in the previous step. Elutions with similar RNA were pooled and then quantified with a NanoDrop Spectrophotometer 2000 and the RNA, which will now be called probe, was stored at -80°C in 5 µl aliquots at 100 ng/µl.

Repetitive sequence blocking RNA

RNA to block repetitive sequences in bison aDNA was transcribed from Bovine HyBlock^{TM} DNA (i.e. Cot-1 DNA, Applied Genetics Laboratories Inc., Melbourne, FL) using a published linear amplification protocol [19]. Briefly, the HyBlock DNA was polished in a reaction containing T4 polynucleotide kinase and T4 DNA polymerase and purified with MinElute spin columns following the PCR cleanup protocol provided. Tailing was performed on the polished DNA with terminal transferase and a tailing solution containing 92 µM dTTP (Thermo Fisher) and 8 µM ddCTP (Affymetrix). After tailing, the Hybloc DNA was purified with MinElute spin columns as before. The HyBlock DNA was then heat denatured and the T7-A18B primer (Supplementary Table 1), containing the T7 RNA polymerase promoter, was allowed to anneal to the poly-T tail with slow cooling. A second-strand synthesis reaction was then performed on the HyBlock DNA using DNA polymerase I Klenow fragment (New England Biolabs) and the product was purified with MinElute spin columns. The double stranded HyBlock DNA was transcribed using a T7 High Yield RNA Synthesis Kit in multiple reactions containing 75 ng DNA, 1x Reaction Buffer, 10 mM rNTPs, 2 µl T7 enzyme mix, and H_2O to 20 µl. IVT reactions were incubated for 16 hours at 37°C and then the DNA template was destroyed by adding 2U Turbo Dnase and incubating for an additional 15 min at 37°C. The RNA was purified with RNeasy MinElute spin columns as above. Purified RNA was quantified on a NanoDrop 2000 and 100 ng visualized on an acrylamide gel, which produced a smear 80 to 500 bp in length.

Primary mitochondrial hybridisation capture

Truncated versions of the Illumina adapters were used for hybridisation capture because full-length adapters reduce enrichment efficiency [20]. For the primary hybridisation capture, three Reagent Tubes were prepared for each bison library with the following materials: Reagent Tube #1- 3.5 µl of 35-55 ng/µl DNA library; Reagent Tube #2- 5 µl probes, 1 µl HyBlock RNA, and 0.5 µl of 50 µM P5/P7 RNA blocking oligonucleotides (Supplementary Table 1); Reagent Tube #3- 30 µl Hybridisation Buffer [21]: 75% formamide (Thermo Fisher), 75 mM HEPES , pH 7.3, 3 mM EDTA (Thermo Fisher), 0.3% SDS (Thermo Fisher), and 1.2 M NaCl (Thermo Fisher). Hybridisation capture was performed in a heated lid thermal cycler programmed as follows: Step 1- 94°C for 2 min, Step 2- 65°C for 3 min, Step 3- 42°C for 2 min, Hold 4- 42°C hold. To start hybridisation capture, Reagent Tubes were placed in the thermal cycler at the start of each program Step in the following order: Step 1- Reagent Tube #1; Step 2- Reagent Tube #2; Step 3- Reagent Tube #3. For each library, once the Hold cycle started 20 µl of hybridisation buffer from Reagent Tube #3 was mixed with the RNA in Reagent Tube #2. The entire content of Reagent Tube #2 was then pipetted into Reagent Tube #1 and mixed with the bison library to begin the hybridisation capture. Hybridisation capture was carried out at 42°C for 48 hours.

Magnetic streptavidin beads (New England Biolabs) were washed just prior to the end of the hybridisation capture incubation. For each library, 50 µl of beads were washed twice using 0.5 ml Wash Buffer 1(2X SSC+0.05% Tween-20, all reagents Thermo Fisher) and a magnetic rack. We also saturated all magnetic bead sites that could potentially bind nucleic acid in a non-specific fashion using yeast tRNA, to optimise the expected and specific streptavidin-biotin binding. Briefly, the beads were blocked by incubation in 0.5 ml Wash Buffer 1+ 100 µg yeast tRNA (Thermo Fisher) for 30 min on a rotor. Blocked beads were washed once as before and then suspended in 0.5 ml Wash Buffer. At the end of the hybridisation capture, each reaction was added to a tube of blocked beads and incubated at room temperature for 30 min on a rotor. The beads were then taken through a series of stringency washes as follows: Wash 1 - 0.5 ml Wash Buffer 1 at room temperature for 10 min; Wash 2 - 0.5 ml Wash Buffer 2 (0.75X SSC + 0.05% Tween-20) at 50°C for 10 min; Wash 3 - 0.5 ml Wash Buffer 2 at 50°C for 10 min; Wash 4 - 0.5 ml Wash Buffer 3 (0.2X SSC + 0.05% Tween-20) at 50°C for 10 min. After the last wash, the captured libraries were released from the probe by suspending the beads in 50 µl of Release buffer (0.1 M NaOH, Sigma Aldrich) and incubating at room temperature for 10 min. The Release buffer was then neutralized with the addition of 70 µl Neutralization buffer (1 M Tris-HCl pH 7.5, Thermo Fisher). Captured libraries were then purified with MinElute columns by first adding 650 µl PB buffer and 10 µl 3 M sodium acetate to adjust the pH for efficient DNA binding. Libraries were purified using the provided PCR cleanup protocol and eluting with 35 µl EB+0.05% Tween-20.

Primary hybridisation capture amplification

Amplification of each primary hybridisation capture was performed in five PCRs containing 5 µl of primary captured library, 1X Phusion HF buffer (Thermo Fisher), 200 µM dNTPs, 200 µM each of primers IS7_short_amp.P5 and IS8_short_amp.P7 (Supplementary Table 1), 0.25 U Phusion Hot Start II DNA polymerase (Thermo Fisher), and H_2O to 25 µl. The five PCR products were pooled and DNA was purified using AMPure magnetic beads.

Secondary mitochondrial hybridisation capture

Amplified primary libraries were taken through a second round of hybridisation capture using the same procedure as describe in *Primary mitochondrial hybridisation capture* step.

Secondary hybridisation capture amplification

Indexed primers were used to convert the DNA from the secondary hybridisation capture to full length Illumina sequencing libraries. Each library was amplified in three PCRs containing 5 µl secondary hybridisation capture library, 1X Phusion HF buffer, 200 µM dNTPs, 200 µM each of primers GAII_Indexing_*x* (library specific index) and IS4 (Supplementary Table 1), 0.25 U Phusion Hot Start II DNA polymerase, and H_2O to 25 µl. Amplification was performed in a heated lid thermal cycler programmed as follows 1 cycle: 98°C for 30 sec; 10 cycles: 98°C for 10 sec, 60°C for 20 sec, 72°C for 20 sec; and 1 cycle: 72°C for 180 sec. The five PCR products were pooled and DNA was purified using AMPure magnetic beads.

Samples A003, A005, A006, A007, A017, A15526, A15637, A15668 (CladeX), A4093 (*ancient wisent*) and A15654 (*historical wisent*)

DNA library preparation

Double-stranded Illumina libraries were built from 20 µl of each DNA extract using partial UDG treatment [22] and truncated Illumina adapters with dual 7-mer internal barcodes, following the protocol from [23].

Hybridisation capture

Commercially synthesised biotinylated 80-mer RNA baits (MYcroarray, MI, USA) were used to enrich the target library for mitochondrial DNA. Baits were designed as part of the commercial service using published mitochondrial sequences from 24 placental mammals, including *Bison bison* and *Bos taurus*.

One round of hybridisation capture was performed according to the manufacturer's protocol (MYbaits v2 manual) with modifications. We used P5/P7 RNA blocking oligonucleotides (Supplementary Table 1) instead of the blocking oligonucleotides provided with the kit. We also incubated the magnetic beads with yeast tRNA to saturate all potential non-specific sites on the magnetic beads that could bind nucleic acids and increase the recovery of non-specific DNA and therefore decrease the final DNA yield.

Indexed primers were used to convert the capture DNA to full length Illumina sequencing libraries. Each library was amplified in eight PCRs containing 5 µl hybridisation capture library, 1x Gold Buffer II, 2.5mM $MgCl_2$, 200 µM dNTPs, 200 µM each of primers GAII_Indexing_*x* (library specific index) and IS4 (Supplementary Table 1), 1.25 U Amplitaq Gold DNA polymerase, and H_2O to 25 µl. Amplification was performed in a heated lid thermal cycler programed as follows 1 cycle: 94°C for 6 min; 15 cycles: 98°C for 30 sec, 60°C for 30 sec, 72°C for 40 sec; and 1 cycle: 72°C for 180 sec. The PCR products were pooled and DNA was purified using AMPure magnetic beads (Agencourt®, Beckman Coulter).

<u>Samples LE237, LE242 and LE257 (CladeX)</u>

Target DNA enrichment was performed by capture of the pooled libraries using DNA baits generated from bison (*Bison bison*) mitochondrial DNA [24]. The baits were generated using three primer sets (Supplementary Table 1, f) designed with the Primer3Plus software package [25]. All extractions and pre-amplification steps of the library preparation were performed in clean room facilities and negative controls were included for each reaction.

<u>Sample A3133 (steppe bison)</u>

DNA repair and polishing were performed in a reaction that contained 20 µl bison A3133 extract, 1x NEB Buffer 2, 3U USER enzyme cocktail, 20U T4 polynucleotide kinase, 1mM ATP, 0.1 mM dNTPs, 8 µg RSA, and H_2O to 38.5 µl. The reaction was incubated at 37°C for 3 hours then 4.5U of T4 DNA polymerase was added and the reaction incubated at 25°C for a further 30 min. Double-stranded libraries were then built with truncated Illumina adapters containing dual 5-mer internal barcodes as in [16] with the final amplification with indexed primers using Phusion Hot Start II DNA polymerase to obtain full length Illumina sequencing libraries.

Nuclear locus capture

Genome-wide nuclear locus capture was attempted on DNA repaired libraries of 13 bison samples (as described above - see Supplementary Supplementary Table 2). Two different sets of probe were used (as described below), but ultimately, only the 9908 loci common to both sets were used for comparative analysis (see nuclear locus analysis section).

Probe sets

40k SNP probe set

This probe set was originally designed to enrich 39,294 of the 54,609 BovineSNP50 v2 BeadChip (Illumina) bovine single nucleotide polymorphism (SNP) loci used in a previous phylogenetic study [26], allowing for a direct comparison of the newly generated data to published genotypes. The discrepancy in the number of surveyed targets was due to manufacturing constraints, as the flanking sequences surrounding certain bovine SNP were too degenerate for synthesis with the MyBaits technology. Probes (MYcroarray, Ann Arbor, MI) were 121-mer long, centred on the targeted bovine SNP and with no tiling, as per the original design of the BovineSNP50 v2 BeadChip [27].

The BovineSNP50 v2 BeadChip assay targets SNPs that are variable in *Bos taurus* in order to genotype members of cattle breeds. Consequently, SNPs are heavily ascertained to be common in cattle, and their use in phylogenetic studies of other bovid species results in levels of heterozygosity that decrease rapidly with increased genetic distance between cattle and the species of interest. Decker et al. (2009) found the average minor allele frequency in plains bison and wood bison for the 40,843 bovine SNPs used in the phylogenetic analysis was 0.014 and 0.009, respectively. Average minor allele frequencies ranged from 0.139 to 0.229 in breeds of taurine cattle.

10k SNP probe set

A second set of probes was ordered from MyBaits that targeted a 9,908 locus subset of the previous 39,294 bovine SNPs selected for enrichment. This smaller subset was chosen to minimise ascertainment bias during phylogenetic and population analyses based on their polymorphism within the diversity of available modern genotypes of bison (American and European), Yak, Gaur and Banteng (total of 72 individuals). All of these taxa belong to a monophyletic clade, outside of the cattle diversity, and are consequently all equidistant from the cattle breeds that were used to ascertain the SNP [27], therefore reducing the impact of ascertainment bias when conducting comparisons within the clade. The exclusion of monomorphic sites across specie allows focusing the capture on loci that are more likely to be phylogenetically informative within the bison diversity. Furthermore, singleton sites (only variable for one modern individual, and therefore not informative for the modern phylogeny) were retained on the principle that they might capture some of the unknown ancient diversity of bison when genotyping ancient individuals.

We designed 70-mer probes, and this short length, as well as the limited number of targets, allowed for a tiling of 4 different probes for each targeted locus, within the same MYcroarray custom kit of 40,000 unique probes. Among all potential 70-mer sequences within the original 121-mer probe sequence set, only those containing the targeted bovine SNP no fewer than 10 nucleotides from either end were retained as potential probes. Four probes were then designed using the following criteria: i) Estimated melting temperature closest to the average from the 40k SNP probe set; ii) Optimum proportion of guanine based on the efficiency of the 40k SNP probe set; iii) No two probes can be closer than 7 nucleotides from one another; iv) All 'GGGG' and 'CTGGAG' motifs were modified to 'GTGT' and 'CTGTAG', respectively. The former change was incorporated on the recommendation from MyBaits to avoid poly G stretches because their synthesis technology has difficulty with this type of motif and the latter variation was included to remove a restriction site that will be used in a future protocol to produce these probes from an immortalized DNA oligo library [28].

DNA library preparation

All DNA libraries were used for capture of both the mitochondrial genome and genome-wide nuclear loci. See Supplementary Information "Whole mitochondrial genome sequencing" for protocols.

Hybridisation capture

One round of hybridisation capture was performed according to the manufacturer's protocol (MYbaits v2 manual) with modifications. We used P5/P7 RNA blocking oligonucleotides (Supplementary Table 1) instead of the blocking oligonucleotides provided with the kit. We also incubated the magnetic beads with yeast tRNA (see above) to saturate all potential non-specific sites on the magnetic beads that could bind nucleic acids and increase the recovery of non-specific DNA.

Indexed primers were used to convert the capture DNA to full length Illumina sequencing libraries. Each library was amplified in eight PCRs containing 5 µl hybridisation capture library, 1C Gold Buffer II, 2.5mM $MgCl_2$, 200 µM dNTPs, 200 µM each of primers GAII_Indexing_*x* (library specific index) and IS4 (Supplementary Table 1), 1.25 U Amplitaq Gold DNA polymerase, and H_2O to 25 µl. Amplification was performed in a heated lid thermal cycler programed as follows 1 cycle: 94°C for 6 min; 15 cycles: 98°C for 30 sec, 60°C for 30 sec, 72°C for 40 sec; and 1 cycle: 72°C for 180 sec. The PCR products were pooled and DNA was purified using AMPure magnetic beads.

NGS and data processing

Whole mitochondrial genomes

All libraries enriched for the mitochondrial genome were sequenced in paired-end reactions on Illumina machines (HiSeq 2500 for LE237A, LE242B and LE247B – MiSeq for the rest), except for A017 and A15526 from which the final concentration of DNA obtained after capture was insufficient for sequencing. The mitochondrial genome of the steppe bison A3133 was recovered from shotgun sequencing on an Illumina HiSeq, performed in the context of another study (see Supplementary Table 3).

All NGS reads were processed using the pipeline Paleomix v1.0.1[29]. AdapterRemoval v2[30] was used to trim adapter sequences, merge the paired reads, and eliminate all reads shorter than 25 bp. BWA v0.6.2[31] was then used to map the processed reads to the reference mitochondrial genome of the wisent (NC_014044) or the American bison (NC_012346, only for the steppe bison A3133). Minimum mapping quality was set at 25, seeding was disabled and the maximum number or fraction of gap opens was set to 2.

MapDamage v2[32] was used to check that the expected contextual mapping and damage patterns were observed for each library, depending on the enzymatic treatment used during library preparation (see Supplementary Table 3 and Figures S1-3 for examples), and re-scale base qualities for the non-repaired libraries.

Finally nucleotides at the position of the bovine SNP were called using samtools and bcftools, setting the minimum base quality at 30 and the minimum depth of coverage at 2. Consensus sequences were then generated using the Paleomix script vcf_to_fasta.

Nuclear

Nuclear DNA from historical (historical wisent: A15654) and ancient (ancient wisent: A4093; CladeX: A15526, A001, A003, A004, A005, A006, A007, A017, A018; steppe: A3133, A875) samples, containing HiSeq data (A3133 and A875) and MiSeq data (all samples), was processed using Paleomix v1.0.1[29] to map reads against the *Bos taurus* reference UMD 3.1[33]. Paleomix was configured to use BWA v0.6.2[31] for mapping, with seeding disabled and -n 0.01 -o 2 (see Supplementary Table 2). MapDamage v2[32] was used to check that the expected contextual mapping and damage patterns were observed for each library, and empirically re-scale base qualities at the end of the fragments.

Variants were called using the consensus caller of samtools/bcftools v1.2[34] limiting calls to the 9908 capture sites. Variant calls with a QUAL value lower than 25 were removed. The genotypes for historical and ancient samples were merged with previously published extant bovid 40k capture data[26], and *Bos primigenius* (aurochs) sample CPC98[35]. Only genotypes for the 9908 loci common among all data were retained.

Supplementary Note 2:

DNA analyses

Phylogenetic analysis

Mitochondrial control region phylogeny

The 60 newly sequenced bovid mitochondrial regions (Supplementary Data 1) were manually aligned, using SeaView v4.3.5[36]. These sequences were aligned with 302 published sequences (Supplementary Table 4) representing the following bovid mitochondrial lineages: European bison or wisent (*Bison bonasus*)*,* American bison (*Bison bison*)*,* steppe bison (*Bison priscus*)*,* zebu (*Bos indicus*)*,* and cattle (*Bos taurus*). Among these published sequences, 5 were from steppe bison collected in the Urals (Shapiro et al. 2004, Supplementary Data 1).

The TN93+G6 model of nucleotide substitution was selected by comparison of Bayesian information criterion (BIC) scores in ModelGenerator v0.85[37]. A phylogenetic tree was then inferred using both maximum-likelihood and Bayesian methods (Figure 2A). Bayesian analyses were performed using the program MrBayes v3.2.3[38]. Posterior estimates of parameters were obtained by Markov chain Monte Carlo sampling with samples drawn every 1000 steps. We used 2 runs, each of four Markov chains, comprising one cold and three heated chains, each of 10 million steps. The first 50% of samples were discarded as burn-in before the majority-rule consensus tree was calculated. A maximum-likelihood analysis was performed with the program PhyML v3[39], using both NNI and SPR rearrangements to search for the tree topology and using approximate likelihood-ratio tests to establish the statistical support of internal branches. Complete phylogenies inferred using both methods are shown in Supplementary Figure 4.

Whole mitochondrial genome phylogeny

The 16 newly sequenced bison whole mitochondrial genomes (Supplementary Data 1) were aligned with 31 published sequences (Supplementary Table 5) representing the following bovid mitochondrial lineages: 3 wisent (*Bison bonasus*), 8 American bison (*Bison bison*), 1 steppe bison (*Bison priscus*), 5 yaks (*Bos grunniens – Bos mutus*), 2 zebus (*Bos indicus*), 7 cattle (*Bos taurus*), 2 aurochsen (*Bos primigenius*), and 4 buffalo (*Bubalus bubalis*).

We used the same methods as described above for the control region to align and estimate the phylogeny. The HKY+G6 model of nucleotide substitution was selected through comparison of BIC scores (Figures 2B and S5).

Estimation of evolutionary timescale

To estimate the evolutionary timescale, we used the program BEAST v1.8.1[40] to conduct a Bayesian phylogenetic analysis of all radiocarbon-dated samples from CladeX and wisent (Figure 1C). The GMRF skyride model[41] was used to account for the complex population history, and a strict clock was assumed. We found support for a strict molecular clock based on replicate analyses using a relaxed uncorrelated lognormal clock[42], which could not reject the strict clock assumption.

Mean calibrated radiocarbon dates associated with the sequences were used as calibration points. Some samples appear to be older than 55 ky: one from the Urals, four from the North Sea and five from the Caucasus (Supplementary Data 1). Because these dates have effectively infinite radiocarbon error margins, we allowed them to vary in the analysis by treating them as distinct parameters to be estimated in the model[43]. The dated samples from Mezmaiskaya Cave are from stratigraphic layers 2B4 and 2B3, which lie atop of layer 3. All these lower Middle Palaeolithic layers at Mezmaiskaya have 14C results beyond the radiocarbon limit, reflected in the predominance of greater-than or near-background limit ages[11], and therefore are consistent with the electron spin resonance (ESR) chronology for these levels[10], which suggests mean ages in the range from 53 to 73 ky BP (including error margins). Consequently, for each Caucasian sample, we specified a lognormal prior age distribution (mean=8,000) with an offset of 50 ky and with 95% of the prior probability less than 80 ky. A similar prior distribution (mean=26,000) was used for the five remaining samples that had infinite radiocarbon dates, with a 95% prior probability less than 150 ky. Based on the results of all four phylogenetic analyses described above, which showed strong support for the reciprocal monophyly of CladeX and wisent when outgroups were included, this monophyly was constrained for the BEAST runs.

All parameters showed sufficient sampling (indicated by effective sample sizes above 200) after 5,000,000 steps, with the first 10% of samples discarded as burn-in. In addition, a date-randomization test was conducted to check whether the temporal signal from the radiocarbon dates associated with the ancient sequences was sufficient to calibrate the analysis[44]. This test randomizes all dates and determines whether the 95% high posterior density (HPD) intervals of the rates estimated from the date-randomized data sets include the mean rate estimated from the original data set (Supplementary Figure 6).

The time to the most recent common ancestor (tMRCA) between wisent and CladeX mitochondrial lineages was estimated at 121.6 kyr (92.1 – 152.3) (Figure 2C). The tMRCAs for the two lineages was inferred to be 69.3 kyr (53.4 – 89.4) for wisent and 114.9 kyr (89.2 – 143.1) for CladeX. Furthermore, there is some phylogeographical structure within CladeX, with all individuals from the North Sea forming a basal group, which existed before the population replacement with steppe bison, but complete mixture of genetic diversity between all locations after re-colonization. In addition, the tMRCA of the MIS 3 diversity of CladeX was estimated to be about 53.1 kyr (41.5 – 67.5). This date closely matches the ages of the last observed MIS 4 CladeX individuals across all sampled locations, supporting the idea of a population movement and contraction of wisent individuals towards a refugium during the warmer period of MIS 3 in Europe.

Nuclear phylogeny from bovine SNP locus data

Phylogenetic trees were inferred from nuclear locus data (see next section for information about the data sets). First, a phylogenetic tree of modern representatives of bovid species, and with sheep as an outgroup, was inferred from published 40,843 data[26] (Supplementary Figure 7). Using RAxML v8.1.21[45], the three characters (genotype states AA, AB and BB) from the BovineSNP50 chip were considered as different states in an explicit analogue of the General Time Reversible (GTR) substitution model, with separate substitution parameters for the three possible transformations. For all analyses, 20 maximum likelihood searches were conducted to find the best tree, and branch support was estimated with 500 bootstrap replicates using the rapid bootstrapping algorithm[46].

This species tree, estimated from genome-wide nuclear locus data, shows that the extant bison species (wisent and American bison) are sister taxa, contrary to the phylogenetic signal from the maternally inherited mitochondrial genome. This topology also clearly shows the paraphyletic status of the genus *Bos* (banteng, gaur, yak, zebu and cattle), as it also includes the genus *Bison* (wisent and American bison).

Using the same method, we reconstructed the phylogeny of bison with the inclusion of five pre-modern samples (for which the highest number of nuclear loci were called amongst the ~10k nuclear bovine SNPs). When only the two steppe bison specimens are included they form a sister-lineage to modern American bison (Supplementary Figure 8A). Similarly, when the steppe bison and pre-modern wisent (including ancient, historical and CladeX) are included, all five pre-modern specimens form a clade most closely related to American bison (Supplementary Figure 8C). However, when only the pre-modern wisent is included, the three specimens (ancient, historical and CladeX) form a clade that is most closely related to modern wisent (Supplementary Figure 8B). These conflicting results reflect the complex non-tree like relationships among the modern and pre-modern taxa, and are consistent with the hybridisation origin of wisent/CladeX and the severe bottleneck in the recent history of the wisent. Hence, we used population genomics statistics to study this nuclear locus dataset (see next section). Finally, these topologies are robust to the removal of transitions (see Supplementary Figure 8D), a minimum depth of 2 for variant calling, and haploidisation (data not shown).

Genome wide nuclear locus analysis

Captured nuclear loci corresponding to bovine SNPs for ancient samples were analysed with published genotypes from modern populations: 20 American bison were selected on the criterion that they do not display any detectable signal of recent introgression from cattle (unpublished data); 2 Yak (*Bos gruniens*); 10 water buffalo (*Bubalus bubalis*); and 10 Sheep (*Ovis aries*). Additionally, 7 modern wisent were selected (among 50 sequenced – [47]) as non-related individuals on a known five-generation pedigree (as shown in Supplementary Figure 9).

Principal Component Analysis

PCA (Figures 3A and S10) was performed using EIGENSOFT version 6.0.1 [48]. In Figure 3A, CladeX sample A006 was used as the representative of CladeX, as this sample contained the most complete set of nuclear loci called at the bovine SNP loci (see Supplementary Table 2). Other CladeX individuals, as well as ancient wisent, cluster towards coordinates 0.0, 0.0 (see Supplementary Figure 10), most likely due to missing data.

Topology testing with the D statistic

For three bison populations, assuming two bifurcations and no hybridisations, there are three possible phylogenetic topologies. For this simple case, the D statistic is expected to be significantly different from zero for exactly two of the three topologies, and not significantly different from zero for the most parsimonious topology. We therefore calculate a D statistic [49] for each of these three topologies, using the sheep (*Ovis aries*) as an outgroup.

When D statistics for the set of three topologies do not indicate zero for one topology and non-zero for the other two, the true phylogeny is not treelike. However, the most parsimonious topology may still be apparent when considering only small amounts of introgression from populations of similar size. The interpretation of a most parsimonious tree topology is not valid where confidence intervals around the D statistic closest to zero, contain one or more of the other D statistics.

In this manner, the D statistic was used to indicate the most parsimonious topology for phylogenies including CladeX, ancient wisent, historical wisent, modern wisent, steppe bison and aurochs (Supplementary Figure 11). D statistics were calculated using ADMIXTOOLS version 3.0, git~3065acc5 [50].

Following concern over the limited amount of data for CladeX, particularly in samples other than 6A, we calculated the D statistics with sample 6A omitted from the analysis (Supplementary Figure 12). The most parsimonious topologies match in both cases.

Sensitivity to other factors were also investigated, such as setting a bovine SNP site coverage depth threshold of two (Supplementary Figure 13), changing the outgroup to *Bubalus bubalis* (Asian water buffalo, Supplementary Figure 14), and haploidisation by randomly sampling an allele at heterozygous sites (Supplementary Figure 15). None of these factors had notable influences on the outcome.

We also considered that the obtained topologies may have been caused by the small number of observed loci. To determine how sensitive the topology testing was missing data, we performed bootstrap resampling of the locus calls on decreasingly sized subsets of the data (Supplementary Table 7). For 10,000 bootstraps, we counted how often we obtained a result other than shown in Supplementary Figure 11.

For this bootstrap, a topology is considered to be simple if: (1) It has a D statistic which, uniquely amongst the set of three, is not significantly different from zero, or (2) All three are significantly different from zero but one has a D statistic closest to zero, with confidence intervals that do not overlap the D statistic for the other two topologies.

For simple topologies, we counted how often the bootstrap replicate suggested a simple topology that did not match the most parsimonious topology in Supplementary Figure 11. For non-simple topologies, we counted how often the result suggested any simple topology. In both cases, a lack of support for any simple topology (such as multiple topologies having a D statistic not significantly different from zero) was not counted.

This bootstrapping shows that the D statistics are robust to the small number of observed genotypes.

Admixture proportion determination using an f4 ratio

The proportion of the wisent's ancestry differentially attributable to the steppe bison and the aurochs, was estimated with AdmixTools using an f4 ratio, as described in [50] with sheep (*Ovis aries*) as the outgroup. For the admixture graph shown in Supplementary Figure 16, the admixture proportion, α, is the ratio of two f4 statistics.

$$\alpha y = F4(A, O; X, C)$$

$$y = F4(A, O; B, C)$$

$$\alpha = \frac{\alpha y}{y} = \frac{F4(A, O; X, C)}{F4(A, O; B, C)}$$

For the estimation of admixture proportions using an f4 ratio, it is intended that the ingroup A, while closely related to B, has diverged from B prior to the admixture event. However, in the context of steppe ancestry for wisent, no such population matching ingroup A was available. The admixture graph for wisent is shown in Supplementary Figure 17.

$$\alpha y = F_4(AmericanBison, O; Wisent, Aurochs)$$

$$x + y = F_4(AmericanBison, O; Steppe, Aurochs)$$

$$\alpha \approx \frac{\alpha y}{x + y} = \frac{F_4(AmericanBison, O; Wisent, Aurochs)}{F_4(AmericanBison, O; Steppe, Aurochs)}$$

Where α in Supplementary Figure 17 is approximately determined by the f4 ratio for small branch lengths *x*. The f4 ratio we calculate therefore represents a lower bound on the proportion of steppe bison present in the wisent populations. The steppe ancestry was found to be at least 0.891, with a standard error of 0.026 (Supplementary Table 6-A).

Sensitivity to haploidisation was checked by randomly sampling an allele at heterozygous sites (Supplementary Table 6-B), which had no notable influence on the outcome.

Hypergeometric test for shared derived alleles

To test whether the wisent lineages (including CladeX) have a common hybrid ancestry (Supplementary Figure 18A), or whether multiple independent hybridisation events gave rise to distinct wisent lineages (Supplementary Figure 18B), we identify nuclear loci which have an ancestral state in the aurochs lineage, but a derived state in the steppe lineage (see next section ‘identification of derived alleles’). Under the assumption of a single hybrid origin, we expect a common subset of derived steppe alleles to be present in the various wisent lineages. In contrast, multiple hybridisation events would result in different subsets of derived steppe alleles being present in different wisent lineages. Likewise, we expect the subset of derived aurochs alleles to indicate either one, or multiple hybridisation events.

If the total number of derived steppe alleles is *s*, the number of derived steppe alleles observed in one wisent lineage is *a*, and the number in a second wisent lineage is *b*, then under model B, the number of sites which are found to be in common is a random variable X~HGeom(*a*, *s*-*a*, *b*). Where HGeom is the hypergeometric distribution, having probability mass function:

$$P(X = k) = \frac{\binom{a}{k}\binom{s-a}{b-k}}{\binom{s}{b}}$$

For the number of derived steppe alleles in common between two wisent lineages, c, we calculate P(X ≥ c). This indicates the likelihood of having observed c or more derived steppe alleles in common, if independent hybridisation events gave rise to both wisent and CladeX lineages.

Likelihoods were calculated for steppe derived alleles on all pairwise combinations of wisent lineages (Supplementary Table 8), and then repeated for derived aurochs alleles (Supplementary Table 9). This provides strong support for an ancestral hybridisation event occurring prior to the divergence of the wisent lineages.

We note that parallel genetic drift may also result in a pattern of alleles observed to be derived in the steppe lineage and the wisent lineages, however this is only a confounding factor where the parallel drift occurred in the post hybridisation lineage common to wisent and CladeX in Supplementary Figure 18A. Therefore, this only confounds the determination of genomic positions from a specific parent population, not that the wisent and CladeX lineages have shared ancestry post hybridisation. Alleles under strong selection following distinct hybridisation events would also be shared between lineages more often than if they were randomly distributed. We consider this situation unlikely, as it would require that the same alleles were randomly introgressed repeatedly, and then a strong selective advantage of the alleles at all times and in all environments.

Although we cannot reject the hypothesis that the modern European bison morph may be recent, and only appeared after the LGM as an adaptation to the Holocene environment in Europe, it would mean that the *Bos* mitochondrial lineage has been maintained in the steppe bison diversity throughout the late Pleistocene, and that only individuals carrying this mitochondrial lineage survived in Europe. Therefore, a hybrid origin of the European morph prior to 120 kyr, and maintained during the late Pleistocene, is more parsimonious with the current data.

Identification of derived alleles

The identification of a derived allele in the B lineage of Supplementary Figure 16, for the above analysis, can be performed in a simple way. If the ancestral allele is fixed in both C and the outgroup O, and the derived allele is fixed within B, then the site may be readily identified as derived. However, such fixed alleles are likely to be rare, especially in large populations, and therefore in limited number in our 10K SNP subset. Furthermore, a steppe bison derived allele observed in a wisent population may not be fixed in the wisent, as the population may also contain the ancestral allele from the aurochs lineage.

Relaxing the criterion of allele fixation in any lineage, we identify differential ancestry using the difference in allele frequencies between populations. An ancestral site is one in which the allele frequency closely matches that of the outgroup and a derived site has an allele frequency differing from the outgroup.

For the admixture graph in Supplementary Figure 16, where population X has ancestry from both B and C lineages, with outgroup O, we define an allele frequency shift in B, analogous to a derived state, if

$$\hat{F}_2(C,O) < \hat{F}_2(X,C) \text{ and } \hat{F}_2(C,O) < \hat{F}_2(X,O),$$

where $\hat{F}_2(M,N)$is an unbiased estimate of $(m-n)^2$, for populations M and N with population allele frequencies *m* and *n* at a single locus, as in Appendix A of [50]. Similarly, we define the allele frequency shift in B to have the same shift in X if, in addition to the shift in B:

$$\hat{F}_2(B,X) < \hat{F}_2(B,C) \text{ and } \hat{F}_2(B,X) < \hat{F}_2(B,O) \text{ and}$$

$$\hat{F}_2(B,X) < \hat{F}_2(X,C) \text{ and } \hat{F}_2(B,X) < \hat{F}_2(X,O) \text{ and}$$

$$\hat{F}_2(C,O) < \hat{F}_2(B,C) \text{ and } \hat{F}_2(C,O) < \hat{F}_2(B,O).$$

By observing a shared allele frequency shift instead of shared fixed alleles, we obtain greater sensitivity to the phylogenetic signal that is specific to one ancestral lineage. As for fixed derived alleles, the specific sites showing an allele frequency shift are identified, and can then be compared between multiple daughter populations.

Admixture proportion determination using ABC and simulated data

As the f4 ratio test is giving an upper limit to the amount of aurochs introgression (due to the branch length uncertainty shown in Supplementary Figure 17), we independently test the admixture proportions using simulated data and an ABC approach.

Approximate Bayesian Computation (ABC) is a likelihood-free methodology employed when calculating likelihood functions is either impossible or computationally expensive[51]. The methodology relies on being able to efficiently simulate data, and then compare simulated data to observed data. When simulated data is sufficiently close to the observed data, the parameters used to simulate the data are retained in a posterior distribution.

Consider a single locus, which for three individuals A, B, and C, two different genotypes are observed. The three possible patterns that can be observed are AB, BC, and AC, denoted by the tree tips with shared state. The observed pattern results from a single mutation somewhere on the gene tree, where the position of the mutation relative to the internal node defines which pattern is observed. For example, from the un-rooted gene tree in Supplementary Figure 19c, if a mutation occurs on the branch between C and the internal node, the pattern AB is observed. We assume the relevant time scales are short enough that multiple mutations at a single locus are rare (infinite sites model[52]).

Under the assumption of neutral and independent mutations, the number of fixed mutations accumulating on a branch is Poisson distributed with mean $\mu \times t$, where μ is mutations per locus per generation, and time t is in units of $2N_e$ generations[53,54]. The counts $\boldsymbol{n} = \ (n_{ab}, n_{bc}, n_{ac})$, of observed site patterns AB, BC, and AC, are random variables, which for topology X_1 (Supplementary Figure 19c),

$$n_{ab} \sim Pois(T_m + T_c),$$

$$n_{bc} \sim Pois(T_a),$$

$$n_{ac} \sim Pois(T_b),$$

and topology X_2 (Supplementary Figure 19d),

$$\begin{aligned} n_{ab} &\sim Pois(T_c), \\ n_{bc} &\sim Pois(T_m + T_a), \\ n_{ac} &\sim Pois(T_b), \end{aligned}$$

where $\boldsymbol{T} = (T_a, T_b, T_c, T_m)$ are branch lengths in units of evolutionary time of $2N_e\mu$ generations, and the total number of observed patterns is $N = n_{ab} + n_{bc} + n_{ac}$. Thus for a locus where two genotypes are observed, the probability of patterns AB, BC, AC, is given by $\boldsymbol{p}^{\boldsymbol{T}} = (p_{ab}^{\boldsymbol{T}}, p_{bc}^{\boldsymbol{T}}, p_{ac}^{\boldsymbol{T}}.)$, where for topology X_1 (Supplementary Figure 19c),

$$\begin{aligned} P(\text{AB}|\boldsymbol{T}, X_1) &= p_{ab}^{\boldsymbol{T},X_1} = (T_m + T_c)/(T_m + T_c + T_a + T_b) \\ P(\text{BC}|\boldsymbol{T}, X_1) &= p_{bc}^{\boldsymbol{T},X_1} = T_a/(T_m + T_c + T_a + T_b) \\ P(\text{AC}|\boldsymbol{T}, X_1) &= p_{ac}^{\boldsymbol{T},X_1} = T_b/(T_m + T_c + T_a + T_b) \end{aligned}$$

and for topology X_2 (Supplementary Figure 19d),

$$\begin{aligned} P(\text{AB}|\boldsymbol{T}, X_2) &= p_{ab}^{\boldsymbol{T},X_2} = T_c/(T_m + T_c + T_a + T_b) \\ P(\text{BC}|\boldsymbol{T}, X_2) &= p_{bc}^{\boldsymbol{T},X_2} = (T_a + T_m)/(T_m + T_c + T_a + T_b) \\ P(\text{AC}|\boldsymbol{T}, X_2) &= p_{ac}^{\boldsymbol{T},X_2} = T_b/(T_m + T_c + T_a + T_b). \end{aligned}$$

We simulate site pattern counts for each of the two species trees in Supplementary Figure 19 by drawing from a Multinomial distribution, where for tree topology X_1, $\boldsymbol{n}^{X_1} \sim \text{Mult}(N, \boldsymbol{p}^{\boldsymbol{T},X_1})$, and for tree topology X_2, $\boldsymbol{n}^{X_2} \sim \text{Mult}(N, \boldsymbol{p}^{\boldsymbol{T},X_2})$.

Given a collection of site pattern counts from a hybrid tree with hybridisation parameter $\gamma \in [0,1]$ (Figure S19e), we expect that the combined site pattern counts will be a linear combination of the counts for the different topologies X_1 and X_2. This assumption is reasonable for a large number of total observations N. The simulated counts, $\boldsymbol{n}^{\gamma}$, of site patterns for the hybridised tree is then given by

$$\begin{aligned} \boldsymbol{n}^{\gamma} &= \gamma \boldsymbol{n}^{X_1} + (1 - \gamma)\boldsymbol{n}^{X_2} \\ &= (n_{ab}^{\gamma}, n_{bc}^{\gamma}, n_{ac}^{\gamma}). \end{aligned}$$

As branch lengths are not known (μ, N_e and number of generations are all unknown), we use uninformative priors for the branch lengths. Furthermore, we only require relative branch lengths, so branch lengths $\boldsymbol{T}$ used for simulation were scaled such that $T_b = 1$. Hence we can meaningfully simulate counts of site patterns $\boldsymbol{n}^{\gamma}$ under hybridisation, for comparison to observed site pattern counts.

We perform ABC using the R package ‘abc’, with a ridge regression correction for comparison of the simulated and observed data using the “abc” function[55]. The distance between the observed and simulated data sets is calculated as the Euclidean distance in three-dimensional space. A tolerance $\epsilon = 0.005$ was chosen so that the closest $\ell \times \epsilon$ simulated data sets are retained. For each analysis we had $\ell = 100000$, resulting in 500 posterior samples.

We performed leave-one-out cross-validation using the function “cv4abc” on $\ell' = 250$ randomly selected simulations, and report the prediction error, calculated as

$$E_{\text{pred}} = \frac{\sum_{i=1}^{\ell'} (\hat{\gamma}_i - \gamma_i)^2}{\mathrm{Var}(\gamma_i)}$$

for each analysis. At most the prediction error was 0.5111 standard deviations away from zero, and so we observe that the ridge regression has performed well (see Supplementary Table 11).

Similarly, on inspection of the cross-validation plots, we observe that the ridge regression performs well for γ, as the true simulated values of γ are well estimated by the ridge regression correction. Hence the correction has strengthened the parameter inference methodology when compared to a simple rejection algorithm.

We avoid reporting sample means due to the heavy negative skew in the posterior distributions of γ, and hence report the median (the most central ordered observed value) and mode of each distribution. The mode is estimated using a kernel density estimate of the posterior distribution. Not all simulated data is equally 'close' to the observed data, and the median and mode are weighted according to these distances[56].

The weighted posterior median was between 0.8250 and 0.8660, and the weighted posterior mode was between 0.9034 and 0.9384. These measures of centre indicate evidence for some non-zero level of hybridisation from the Aurochs genome. Evidence against hybridsation must be indicated by overwhelming support for either $\gamma = 0$ or $\gamma = 1$ (no mixing of the tree topologies). However, these values lie on either end of the support for the prior distribution of γ, and hence any resulting posterior distribution for γ. There- fore, classical highest probability density (HPD) intervals cannot be used to indicate uncertainty in the estimates of these measures of centre, as any interval of density less than 100% will result in zero and one being artificially omitted by construction. This is not evidence for or against hybridisation, but rather a consequence of the way in which we calculate HPD intervals.

Supplementary Table 11 gives empirical posterior probabilities for different levels of hybridisation. For example, the first column gives the empirical posterior probability of observing at least 1% hybridisation. This is found for each trio by calculating the total proportion of posterior samples where $0.01 \leq \gamma \leq 0.99$. In general, for some percentage of hybridisation α, Supplementary Table 11 reports

$$[P(\frac{\alpha}{100} \leq \gamma \leq 1 - \frac{\alpha}{100})]$$

for α = 1%, 2%, 3%, 4% and 5%, from the posterior distribution of γ.

As there is no accepted value of γ for which we can claim that significant hybridisation has occurred, we leave it to the reader to consider what they consider to be a significant level of hybridisation, and to find the appropriate probability. However, if one considers 1% hybridisation to be significant, then the observed data indicates that the data has between a 95.80% and 97.20% chance of being from a hybridised topology. Similarly, if one considers 5% hybridisation to be significant, then the observed data has between a 76.40% and 85.00% chance of being from a hybridised topology.

Asymmetrical hybridisation

In this study, we show that wisent and CladeX are of hybrid origin, certainly between ancient aurochs and steppe bison forms. This is consistent with the population structure of most bovids, where a single bull usually breeds with different females of multiple generations. As explained in[57], this usually results in asymmetrical hybridization when males of one species (steppe bison here) dominate males of the other species (aurochs here), therefore preferentially mating with female aurochs, as well as their offspring, potentially over several generations. In addition, male F_1 hybrids are usually sterile or sub-fertile, increasing the amount of steppe bison genomic contribution to the offspring. As illustrated in Supplementary Figure 20, after just a few generations, this mating process results in individuals that are essentially steppe bison for their nuclear genome, but with an aurochs mitochondrial genome (strictly maternally inherited), which is the result that we obtained from the genotyping of historical and ancient wisent individuals (including CladeX).

Supplementary Note 3:

Paleoenvironment reconstruction and stable isotope analyses in the Ural region

The Urals are a well sampled region, with the highest number of genotyped bones through time (Figure 5 and S22). We generated a convex hull based on geo-referenced site locations for all genotyped ancient samples collected from the Urals (Supplementary Figure 21). We used the HadCM3 global circulation model and BIOME4 model to reconstruct paleoclimate and environmental conditions for the Ural region throughout the period from 70,000 years ago to the present day.

We used the HadCM3 global circulation model to reconstructed paleoclimate proxies for the Ural region. The HadCM3 consists of linked atmospheric, ocean and sea ice models at a spatial resolution of 2.5° latitude and 3.75° longitude, resampled at a 1° x 1° latitude/longitude grid cell resolution [58]. The temporal resolution of the raw data is 1,000 year slices back to 22,000BP and 2,000 year slices from 22,000 to 80,000BP [58] We used these palaeo-climate simulations to derive estimates of annual mean daily temperature and Köppen-Geiger climate classifications [59] throughout the period from 70,000 years ago to the present day. We intersected each grid cell in the Ural study region (n = 51) with the derived climate estimates, at each point in time, using ArcGIS 10. We calculated the mean temperature for the region and change in the proportion of the study region represented by four Köppen climate classes, each differing temperature: Dfa (hot summers), Dfb (warm summers), Dfc (cool summers), Dfd (continental temperatures). These are shown in Supplementary Figure 22. Interestingly, our reconstructions for the Urals show a decrease in area with hot and warm summer conditions (Dfa and Dfb) after 35kya.

BIOME4 was used to infer paleovegetation types. BIOME4 is a coupled biogeographical and biogeochemical model that simulates the distribution of 28 plant functional types (PFT) at a global scale [60]. Model inputs for each grid cell are monthly climate (mean annual temperature, mean annual precipitation and mean annual sunshine hours), atmospheric [CO_2], and soil texture class. Ecophysiological constraints determine which PFT is likely to occur in each grid cell. A coupled carbon and water flux model calculates the leaf area index that maximizes net primary production (in gC m^{-2} $year^{-1}$) for each PFT. Competition between PFTs was simulated by using the optimal net primary production of each PFT as an index of competitiveness. Global maps of BIOME4 PFTs were accessed at the same spatial and temporal resolution as the paleoclimate data (http://www.bridge.bris.ac.uk/resources/simulations/). We grouped PFTs into three categories: Grassland (PFT identify numbers = 18-20); Tundra (ID = 22-26); and Forest (ID = 7-11). For each grid cell in the Ural study region, at each point in time, we determined whether the dominant PFT was grassland, tundra or forest. Interestingly the vegetation shift between an all forest-like landscape to a landscape represented by a large proportion of tundra and grassland-like vegetation occurred after 35kya, which coincides with a decrease in hot and warm summer conditions (see above).

These results from the paleovegetation and climate inferences agree with previous landscape reconstructions of the region: In the Middle Urals, where almost all the samplings sites were located, the areas covered with arboreal vegetation underwent changes during MIS3. Spruce and birch open forests were widespread during coolings, and spruce and birch forest-steppe with occurrence of pine formed during warmings. Mesophilic meadows dominated by forbs and grasses were also prevalent during warm climatic events (Lapteva, 2008; 2009; Pisareva and Faustova, 2008). In the south, where one of the sites (Gofmana) is situated, steppe landscapes dominated by Asteraceae, Artemisia, and Poaceae were widespread. Spruce, birch and pine forests covered the areas along the rivers (Smirnov, Bolshakov, Kosintsev et al., 1990). The following was reconstructed for the territory of the Irtysh River: forest-steppe landscapes with pine (Pinus s/g Haploxylon) and spruce forests, as well as meadows with a predominance of Cyperaceae and Poaceae and small quantities of Artemisia and Chenopodiaceae (Araslanov *et al*. 2009).

During MIS2, periglacial forest-steppes dominated by herbaceous communities were typical of the Last Glacial Maximum. Larch, pine and birch covered the river-valleys. Herbaceous vegetation was dominated by goosefoot, sagebrush and grass (Grichuk 2002). Periglacial forest-steppes with arboreal vegetation, including pine-birch forests and small quantities of spruce have been reconstructed for the Last Glacial Termination. Areas covered with sagebrush-goosefoot steppes with small quantities of grass were widespread (Lapteva, 2007).

At later stages of MIS2, periglacial forb-grass forest-steppes with pine, birch and small quantities of spruce have been reconstructed for the Sur'ya 5 and Rasik 1 sites [61]. Periglacial steppes dominated by Artemisia, Rosaceae, Chenopodiaceae, Cichorioideae and Poaceae have been reconstructed for the Voronovka site. Pinus sylvestris and Betula pubescens with occurrence of spruce (Picea), oak (Quercus) and teil (Tilia) covered the river-valleys [62].

The palynological analyses and landscape reconstruction suggest that both bison forms inhabited semi-open landscapes of forest-steppe type, where arboreal vegetation was represented by birch, spruce, pine and sometimes larch, while steppe and meadow herbaceous communities were observed. However, only CladeX (specifically from the Gofmana site, during MIS 3, Rasik 1 and Sur'ya 5, and Voronovka sites, during MIS2) also inhabited steppe-like landscapes, showing a more diverse ecological niche than steppe in this region.

In addition to the paleo-climate and -vegetation reconstructions, stable isotope values (δ13C and δ15N) obtained for all the genotyped bison individuals from the Ural region were compared between steppe bison and wisent (Supplementary Figure 23). Wisent individuals displayed more diverse stable isotope ratios than the steppe bison individuals. This observation is consistent with feeding in more diverse vegetations communities, which correlates well with the reconstructed paleo-environments for the region in the time periods they are found.

Modelled paleo-climate and -vegetation reconstruction at the sampling locations in the southern Urals suggest drastic shifts, which coincide in time with the observed population replacements between steppe bison and wisent. More specifically, between 14 and 31 kya wisent were likely to exist in environmental condition characterised by relatively cold average temperatures, open landscapes with tundra-like flora, and the absence of warm summers. Although modern wisent are found today in wood-like habitats, it has been suggested that they are living in sub-optimal habitat, and paleodiet reconstructions have placed ancient wisent in tundra-like environments, in agreement with our observations [63].

Interestingly, the steppe bison was only recorded when forest vegetation was inferred to dominate the landscape, adding to the evidence that this form of bison might not have been exclusively steppe-adapted [63,64].

Supplementary Note 4:

Cave painting

The present survey, placing wisent across Europe (from the Urals/Caucasus to Ukraine/Italy) during MIS2 and late MIS3, suggests that depictions of bison in European Palaeolithic art, such as cave painting, carving and sculptures, are likely to include representations of wisent. Paleolithic art representations have often been used to infer the morphological appearance of steppe bison, sometimes in great detail [64,4,65–67]. And until now, the steppe bison (i.e., direct ancestor of modern American bison) has always been assumed to be the unique model present at the time of cave painting, and therefore, the diversity within the representations of bison was mainly explained by putative cultural and individual variations of style through time [68–70]. However, in the vast diversity of bison representations (820 pictures representing 20.6% of all known cave ornamentation, according to [71]), two consistent morphological types can be distinguished (see Fig 1 and Fig S24-27). The first type, abundant prior to the last glacial maximum, is characterized by long horns (with one curve), a very oblique dorsal line and a very robust front part of the body (solid shoulders versus hindquarters), all these traits being similar to the modern American bison. The second type, dominating the more recent paintings between 18 and 15 kya, displays thinner sinuous horns (often with double curve), a smaller hump and more balanced dimensions between the front and the rear of the body, similar to the modern wisent lineage, and to some extant the *Bos* lineage. The imposing figure of the steppe bison, with its high hump and long horns stepping out the head profile, certainly was a very strong influence on the artists painting in the cave in Europe before the last glacial maximum. However, later generations thoroughly depicted the slender shape of the more recent form of bison. Considering the geographical and temporal distribution of genotyped steppe bison and wisent presented here, particularly the ~16,000 years old wisent B individual from Northern Italy, it is likely that the variety of bison representations in Paleolithic art does not just come from stylistic evolution, but actually represents different forms of bison (i.e., pre and post-hybridisation) through time.

Supplementary References

1. Wolff, E. W., Chappellaz, J., Blunier, T., Rasmussen, S. O. & Svensson, A. Millennial-scale variability during the last glacial: The ice core record. *Quaternary Science Reviews* **29,** 2828–2838 (2010).
2. Shapiro, B. *et al.* Rise and Fall of the Beringian Steppe Bison. *Science* **306,** 1561–1565 (2004).
3. Leroi-Gourhan, A. & Allain, J. *Lascaux inconnu*. (CNRS, 1979).
4. Capitan, L., Breuil, H. & Peyrony, D. *La caverne de Font-de-Gaume, aux Eyzies (Dordogne)*. (Imprimerie du Chêne, 1910).
5. Lorblanchet, M. *La grotte ornée de Pergouset (Saint-Géry, Lot). Un sanctuaire secret paléolithique*. (Maison des Sciences de l'Homme, 2001).
6. Barrière, C. L'art pariétal de Rouffignac, la grotte aux cent mammouths. *Bulletins et Mémoires de la Société d'anthropologie de Paris* **10,** 144–145 (1983).
7. Groves, C. & Grubb, P. *Ungulate Taxonomy*. (Johns Hopkins University Press, 2011).
8. Drees, M. & Post, K. Bison bonasus from the North Sea, the Netherlands. *Cranium* **24,** 48–52 (2007).
9. Llamas, B. *et al.* High-Resolution Analysis of Cytosine Methylation in Ancient DNA. *PLoS ONE* **7,** e30226 (2012).
10. Skinner, A. R. *et al.* ESR dating at Mezmaiskaya Cave, Russia. *Applied Radiation and Isotopes* **62,** 219–224 (2005).
11. Pinhasi, R., Higham, T. F. G., Golovanova, L. V. & Doronichev, V. B. Revised age of late Neanderthal occupation and the end of the Middle Paleolithic in the northern Caucasus. *PNAS* **108,** 8611–8616 (2011).

12. Reimer, P. J. *et al.* IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. *Radiocarbon* **55,** 1869–1887 (2013).

13. Willerslev, E. & Cooper, A. Ancient DNA. *Proc Biol Sci* **272,** 3–16 (2005).

14. Brotherton, P. *et al.* Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat Commun* **4,** 1764 (2013).

15. Rohland, N. & Hofreiter, M. Ancient DNA extraction from bones and teeth. *Nat. Protocols* **2,** 1756–1762 (2007).

16. Meyer, M. & Kircher, M. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb Protoc* **2010,** pdb.prot5448 (2010).

17. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucl. Acids Res.* **40,** e3–e3 (2012).

18. Cone, R. W. & Schlaepfer, E. Improved In Situ Hybridization to HIV with RNA Probes Derived from PCR Products. *J Histochem Cytochem* **45,** 721–727 (1997).

19. Liu, C., Bernstein, B. & Schreiber, S. *DNA linear amplification*. (Scion Publishin Ltd, 2005).

20. Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* gr.128124.111 (2012). doi:10.1101/gr.128124.111

21. Konietzko, U. & Kuhl, D. A subtractive hybridisation method for the enrichment of moderately induced sequences. *Nucleic Acids Res.* **26,** 1359–1361 (1998).

22. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **22,** 939–949 (2015).

23. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522,** 207–211 (2015).

24. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. *PLoS ONE* **5,** e14004 (2010).

25. Untergasser, A. *et al.* Primer3Plus, an enhanced web interface to Primer3. *Nucl. Acids Res.* **35,** W71–W74 (2007).

26. Decker, J. E. *et al.* Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *PNAS* **106,** 18644–18649 (2009).

27. Matukumalli, L. K. *et al.* Development and Characterization of a High Density SNP Genotyping Assay for Cattle. *PLoS ONE* **4,** e5350 (2009).

28. Shankaranarayanan, P. *et al.* Single-tube linear DNA amplification (LinDA) for robust ChIP-seq. *Nat Meth* **8,** 565–567 (2011).

29. Schubert, M. *et al.* Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Protocols* **9,** 1056–1082 (2014).

30. Lindgreen, S. AdapterRemoval: Easy Cleaning of Next Generation Sequencing Reads. *BMC Research Notes* **5,** 337 (2012).

31. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

32. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29,** 1682–1684 (2013).

33. Zimin, A. V. *et al.* A whole-genome assembly of the domestic cow, Bos taurus. *Genome Biology* **10,** R42 (2009).

34. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27,** 2987–2993 (2011).

35. Park, S. D. E. *et al.* Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and evolution of cattle. *Genome Biology* **16,** 234 (2015).

36. Gouy, M., Guindon, S. & Gascuel, O. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol Biol Evol* **27,** 221–224 (2010).

37. Keane, T. M., Creevey, C. J., Pentony, M. M., Naughton, T. J. & Mclnerney, J. O. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evolutionary Biology* **6,** 29 (2006).

38. Ronquist, F. *et al.* MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst Biol* **61,** 539–542 (2012).

39. Guindon, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol* **59,** 307–321 (2010).

40. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7,** 214 (2007).

41. Minin, V. N., Bloomquist, E. W. & Suchard, M. A. Smooth Skyride through a Rough Skyline: Bayesian Coalescent-Based Inference of Population Dynamics. *Mol Biol Evol* **25,** 1459–1471 (2008).

42. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed Phylogenetics and Dating with Confidence. *PLoS Biol* **4,** e88 (2006).

43. Shapiro, B. *et al.* A Bayesian Phylogenetic Method to Estimate Unknown Sequence Ages. *Mol Biol Evol* **28,** 879–887 (2011).

44. Ho, S. Y. W. *et al.* Bayesian Estimation of Substitution Rates from Ancient DNA Sequences with Low Information Content. *Syst Biol* **60,** 366–375 (2011).

45. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22,** 2688–2690 (2006).

46. Stamatakis, A., Hoover, P. & Rougemont, J. A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Syst Biol* **57,** 758–771 (2008).

47. Pertoldi, C. *et al.* Phylogenetic relationships among the European and American bison and seven cattle breeds reconstructed using the BovineSNP50 Illumina Genotyping BeadChip. *Acta Theriol* **55,** 97–108 (2010).

48. Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLoS Genet* **2,** e190 (2006).

49. Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for Ancient Admixture between Closely Related Populations. *Mol Biol Evol* **28,** 2239–2252 (2011).

50. Patterson, N. *et al.* Ancient Admixture in Human History. *Genetics* **192,** 1065–1093 (2012).

51. Beaumont, M. A., Zhang, W. & Balding, D. J. Approximate Bayesian Computation in Population Genetics. *Genetics* **162,** 2025–2035 (2002).

52. Kimura, M. The Number of Heterozygous Nucleotide Sites Maintained in a Finite Population Due to Steady Flux of Mutations. *Genetics* **61,** 893–903 (1969).

53. Watterson, G. A. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* **7,** 256–276 (1975).

54. Hudson, R. in *Oxford Surveys in Evolutionary Biology* **7,** 1–44 (Oxford University Press, 1990).

55. Csilléry, K., François, O. & Blum, M. G. B. abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution* **3,** 475–479 (2012).

56. Blum, M. G. B. & François, O. Non-linear regression models for Approximate Bayesian Computation. *Stat Comput* **20,** 63–73 (2009).

57. Groves, C. Current taxonomy and diversity of crown ruminants above the species level. *Zitteliana* **B 32,** 5–14 (2014).

58. Singarayer, J. S. & Valdes, P. J. High-latitude climate sensitivity to ice-sheet forcing over the last 120 kyr. *Quaternary Science Reviews* **29,** 43–55 (2010).

59. Peel, M. C., Finlayson, B. L. & McMahon, T. A. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* **11,** 1633–1644 (2007).

60. Kaplan, J. O. Geophysical Applications of Vegetation Modeling. (Lund University, 2001).

61. Lapteva, E. G. Landscape-climatic changes on the eastern macroslope of the Northern Urals over the past 50000 years. *Russ J Ecol* **40,** 267–273 (2009).

62. Lapteva, E. G. & Korona, O. M. Holocene vegetation changes and anthropogenic influence in the forest-steppe zone of the Southern Trans-Urals based on pollen and plant macrofossil records from the Sukharysh cave. *Veget Hist Archaeobot* **21,** 321–336 (2011).

63. Bocherens, H., Hofman-Kamińska, E., Drucker, D. G., Schmölcke, U. & Kowalczyk, R. European Bison as a Refugee Species? Evidence from Isotopic

Data on Early Holocene Bison and Other Large Herbivores in Northern Europe. *PLoS ONE* **10,** e0115090 (2015).

64. Guthrie, R. D. *Frozen fauna of the Mammoth Steppe : the story of Blue Babe*. (University of Chicago Press, 1990).
65. Bandi, H.-G. ; H., W. ;. Sauter, M. R. ;. Sitter, B. *La Contribution de la Zoologie et de L'Ethologie a L'Interpretation de L'Art des Peuples Chasseurs Prehistoriques*. (Editions Universitaires, 1984).
66. Guthrie, R. D. *The nature of Paleolithic art*. (University of Chicago Press, 2005).
67. Paillet, P. *Le bison dans les arts magdaléniens du Périgord*. (CNRS éd, 1999).
68. Breuil, H. *Quatre cents siècles d'art pariétal; les cavernes ornées de l'âge du renne.* (Centre d'études et de documentation préhistoriques, 1952).
69. Leroi-Gourhan, A. *Préhistoire de l'art occidental*. (1965).
70. Petrognani, S. *De Chauvet à Lascaux: l'art des cavernes, reflet de sociétés préhistoriques en mutation*. (Editions Errance, 2013).
71. Sauvet, G. & Wlodarczyk. L'art pariétal, miroir des sociétés paléolithiques. *Zephyrus: Revista de prehistoria y arqueología* **53,** 217–240 (2000).

References in Russian:

Arslanov KH, Laukhin SA, Maksimov FE, *et al.* (2009) Radiocarbon Chronology and Landscapes of Western Siberian Lipovsk-Novoselovsky Interstadial (on evidence of study section near V. Lipovka) // Fundamental Problems of Quaternary: Resultats and Trends of Further Researches. (Ed. A.E. Kantorovich). Novosibirsk. P. 44 – 47. (in Russian).

Grichuk VP (2002) Vegetation of the Late Pleistocene. In: A.A.Velichko (ed.), Dynamics of terrestrial landscape components and inner marine basins of Northern Eurasia during the last 130 000 years. Moscow: GEOS Publishers, pp. 64-88. (in Russian).

Lapteva EG (2007) Реконструкция ландшафтно-климатических изменений на территории Среднего Зауралья в позднеледниковье и голоцене на основе палинологических данных из рыхлых отложений пещеры Першинская-1 // Эколоия древних и традиционных обществ. Вып. 3. (Ред. Н.П. Матвеева). С. 30 – 36. (in Russian).

Lapteva EG (2008) Major palaeogeographical stages and specific landscape-climatic changes on the eastern slope of the Urals during the last 50 kyrs (inferred from palynological data) // Problems of Pleistocene palaeogeography and stratigraphy. (Eds. N.S. Bolikhovskaya and P.A. Kaplin). Vol. 2. P. 196 – 204. (in Russian).

Pisareva VV, Faustova MA (2008) Reconstruction of Landscapes of Northern Russia during the Middle Valday Mega-Interstadial // Way to North: Paleoenvironment and Inhabitants of Arctic and Subarctic (Eds. A.A. Velichko and S.A. Vasil'ev). Moscow.P. 53 – 62. (in Russian).

2.4 Supplementary Data 1: sample details

List of all ancient and historical individuals used in the study. Previously published sequences are shown with their GenBank numbers.

DNA result:

Dloop	≥400 bp of the mitochondrial control region; sanger sequencing
WMG	whole mitochondrial genome; RNA-probe capture and next generation sequencing
nSNP	nuclear single nucleotide polymorphism; probe capture and next generation sequencing

***Calibrated dates:** 95% intervalls are reported after calibration with OxCal v4.2 and the IntCal13 curve (in black) or from the tip date sampling result of the BEAST analysis of Dloop for dates beyond C14 limits (in grey).

	Sample ID	DNA result	Genotype	AMS date		Calibrated dates*	
				Oxdate	Oxerr	Low	High
Urals	A001	Dloop + WMG + nSNP	**CladeX**	12565	55	14510	15152
	A003	Dloop + WMG + nSNP	**CladeX**	12505	55	14321	15075
	A004	Dloop + WMG + nSNP	**CladeX**	19010	80	22586	23155
	A005	Dloop + WMG + nSNP	**CladeX**	15310	70	18400	18750
	A006	Dloop + WMG + nSNP	**CladeX**	18880	90	22490	22992
	A007	Dloop + WMG + nSNP	**CladeX**	58300	2900	53641	70691
	A011	Dloop	**CladeX**	>60900		50081	93622
	A016	Dloop	**CladeX**	49600	1200	47459	52529
	A017	Dloop + nSNP	**CladeX**	18850	90	22472	22963
	A018	Dloop + WMG + nSNP	**CladeX**	13120	60	15485	16000
	BS599	Dloop	**CladeX**	26330	120	30349	30930
	BS604	Dloop	**CladeX**	55400	1800	52281	60568
	BS606	Dloop	**CladeX**	25000	100	28741	29366
	A002	Dloop	**Steppe bison**	51800	1300	49495	55047
	A008	Dloop	**Steppe bison**	31560	210	33001	34022
	A013	Dloop	**Steppe bison**	48400	900	44811	48515
	A014	Dloop	**Steppe bison**	33820	260	35452	36951
	BS592	AY748756 (Dloop)	**Steppe bison**	42500	450	43012	44745
	BS660	AY748766 (Dloop)	**Steppe bison**	29500	140	33433	33980
	BS674	AY748775 (Dloop)	**Steppe bison**	29060	140	32880	33660
	BS708	AY748793 (Dloop)	**Steppe bison**	47050	750	45665	48725
	BS713	AY748795 (Dloop)	**Steppe bison**	30970	180	34514	35288
	BS588	Dloop	**Wisent**	16810	65	20059	20491
	A012	✗	Contamination				
	A015	✗	Contamination				
	A4081	Dloop	**CladeX**	N/A			
	A4082	Dloop	**CladeX**	N/A			
	A4083	Dloop	**CladeX**	N/A			

	Sample ID	DNA result	Genotype	AMS date		Calibrated dates*	
				Oxdate	Oxerr	Low	High
Caucasus	A4084	Dloop	CladeX	N/A			
	A4085	Dloop	CladeX	N/A			
	A4087	Dloop	CladeX	N/A			
	A4088	Dloop	CladeX	N/A			
	A4089	Dloop + WMG + nSNP	CladeX	>59400		50027	93399
	A4091	Dloop	CladeX	>59700		50019	93566
	A4092	Dloop	CladeX	>56600		50030	63620
	A4094	Dloop	CladeX	>56500		50025	60951
	A4104	Dloop	CladeX	12160	40	13906	14186
	A4090	Dloop	Steppe bison	>59400			
	A4093	Dloop + WMG + nSNP	Wisent	>56300		50020	61245
	A4103	Dloop	Cow				
	A4098	Dloop	Brown bear				
	A4086	✗					
	A4095	✗					
	A4096	✗					
	A4097	✗					
	A4099	✗					
	A4100	✗					
	A4101	✗					
	A4102	✗					
	A15644	Dloop	Wisent	Historical (hunted in 1906)			
	A15646	Dloop	Wisent	Historical (hunted in early 20th century)			
	A15648	Dloop	Wisent	Historical (hunted in 1910)			
	A15654	Dloop + WMG + nSNP	Wisent	Historical (hunter in 1911)			
	A3454	Dloop	Wisent	Historical			
	A3455	Dloop	Wisent	Historical			
	A15668	WMG + nSNP	CladeX	13573	36	16182	16547

	Sample ID	DNA result	Genotype	AMS date		Calibrated dates*	
				Oxdate	Oxerr	Low	High
	A15660	Dloop	**CladeX**	18630	220	21962	23012
	LE237A	WMG	**CladeX**	18630	220	21962	23012
	LE242B	WMG	**CladeX**	18630	220	21962	23012
	LE247B	WMG	**CladeX**	18630	220	21962	23012
Western Europe	A2791	Dloop	**CladeX**	>53800		50066	92727
	A2795	Dloop	**CladeX**	29010	160	32789	33652
	A2798	Dloop	**CladeX**	29230	150	33043	33804
	A2808	Dloop	**CladeX**	>61500		50028	65808
	A2809	Dloop	**CladeX**	>61300		50037	73370
	A2811	Dloop	**CladeX**	>62000		50036	68812
	A2792	Dloop	**Steppe bison**	29100	150	32904	33700
	A2793	Dloop	**Steppe bison**	28340	130	31687	32767
	A2796	Dloop	**Steppe bison**	43850	650	45791	48765
	A2797	Dloop	Cow				
	A2799	Dloop	Cow				
	A2810	Dloop	Cow				
	A2800	Dloop	Elk				
	A2801	✗	Contamination				
	A2794	✗					
	A2802	✗					
	A2803	✗					
	A2804	✗					
	A2805	✗					
	A2806	✗					
	A2807	✗					
	BS593	Dloop	**Wisent**	5090	60	5707	5940
	BS600	Dloop	**Wisent**	3430	50	3577	3831
	BS607	Dloop	**Wisent**	1370	50	1227	1369

	Sample ID	**DNA result**	**Genotype**	**AMS date**		**Calibrated dates***	
				Oxdate	**Oxerr**	**Low**	**High**
	A3226	Dloop	**Wisent**	Historical			
	A3227	Dloop	**Wisent**	Historical			
	A3228	Dloop	**Wisent**	Historical			
	A15665	Dloop	**Wisent**	3621	31	3843	3990
	A15526	Dloop + nSNP	**CladeX**	13600	60	16179	16638
	A15637	WMG + nSNP	**CladeX**	>48000			
	SGE2	KM593920 (WMG)	**Steppe bison**	15880	70	18940	19387
Beringia	A875	nSNP	**Steppe bison**	>50000			
	A3133	WMG + nSNP	**Steppe bison**	26360	220	30092	31044

	Sample ID	Origin	Field ID	Museum ID	Type
Urals	A001	Rasik 1 (ZMIPAE)	ACS110	888/117	Pelvis fragment
	A003	Voronovka (ZMIPAE)	ACS121	1871/01	Humerus
	A004	Rasik 1 (ZMIPAE)	ACS88	888/1705	Metacarpal
	A005	Ladeinyi Kamen (ZMIPAE)	ACS108	929/1	Femur
	A006	Sur'ya 5 (ZMIPAE)	ACS100	994/714	Metatarsal
	A007	Sur'ya 3 (ZMIPAE)	ACS94	884/19	Metatarsal
	A011	Sur'ya 5 (ZMIPAE)	ACS99	994/715	Metatarsal
	A016	Gofmana (ZMIPAE)	ACS104	1111/2	Humerus
	A017	Sur'ya 5 (ZMIPAE)	ACS103	994/475	Upper mandible
	A018	Sur'ya 5 (ZMIPAE)	ACS102	994/315	Radius
	BS599	Kholodnyi (ZMIPAE)		816/163	Tibia
	BS604	Sur'ya 5 (ZMIPAE)		994/37	Astralagus
	BS606	Kholodnyi (ZMIPAE)		816/168	Bone fragment
	A002	Sur'ya 5 (ZMIPAE)	ACS101	994/435	Metacarpal
	A008	Dinamitnaya (ZMIPAE)	ACS107	878/28	Metacarpal
	A013	Rasik 1 (ZMIPAE)	ACS109	888/2271	Tibia
	A014	Bobylek (ZMIPAE)	ACS187	528/42256	Tibia
	BS592	Chernye Kosti (ZMIPAE)		887/3	Femur
	BS660	Sur'ya 5 (ZMIPAE)		994/252	Metapodial
	BS674	Kholodnyi (ZMIPAE)		816/166	Phalanx
	BS708	Rasik 1 (ZMIPAE)		888/47	Femur
	BS713	Irtysh River (ZMIPAE)		915/166	Metatarsal
	BS588	Sur'ya 5 (ZMIPAE)		994/716	Metapodial
	A012	Sur'ya 5 (ZMIPAE)	ACS91	994/1003	Metacarpal
	A015	Yurovsk (ZMIPAE)	ACS89	577/7	Femur
	A4081	Mezmaiskaya, level 3	M3M N1		Long Bone
	A4082	Mezmaiskaya, level 3	M3M N2		Long Bone
	A4083	Mezmaiskaya, level 3	M3M N3		Long Bone

	Sample ID	Origin	Field ID	Museum ID	Type
Caucasus	A4084	Mezmaiskaya, level 3	M3M N4		Long Bone
	A4085	Mezmaiskaya, level 3	M3M N5		Long Bone
	A4087	Mezmaiskaya, level 3	M3M N7		Long Bone
	A4088	Mezmaiskaya, level 3	M3M N8		Long Bone
	A4089	Mezmaiskaya, level 2B4	M3M N9		Long Bone
	A4091	Mezmaiskaya, level 2B4	M3M N11		Long Bone
	A4092	Mezmaiskaya, level 2B4	M3M N12		Long Bone
	A4094	Mezmaiskaya, level 2B3	M3M N14		Long Bone
	A4104	Mezmaiskaya, level 1-3	M3M N24		Long Bone
	A4090	Mezmaiskaya, level 2B4	M3M N10		Long Bone
	A4093	Mezmaiskaya, level 2B3	M3M N13		Long Bone
	A4103	Mezmaiskaya, level 1-1	M3M N23		Long Bone
	A4098	Mezmaiskaya, level 2A	M3M N18		Long Bone
	A4086	Mezmaiskaya, level 3	M3M N6		Long Bone
	A4095	Mezmaiskaya, level 2B2	M3M N15		Long Bone
	A4096	Mezmaiskaya, level 2B2	M3M N16		Long Bone
	A4097	Mezmaiskaya, level 2A	M3M N17		Long Bone
	A4099	Mezmaiskaya, level 2A	M3M N19		Long Bone
	A4100	Mezmaiskaya, level 2A	M3M N20		Long Bone
	A4101	Mezmaiskaya, level 1C	M3M N21		Long Bone
	A4102	Mezmaiskaya, level 1C	M3M N22		Long Bone
	A15644	Kuban Oblast (ZIRAS)		7987	Skull
	A15646	Kuban Oblast (ZIRAS)		8834	Skull
	A15648	Kuban Oblast (ZIRAS)		8836	Skull
	A15654	Kuban Oblast (ZIRAS)		8853	Skull
	A3454	Caucasus (NHM)		92.3.15.2	Tooth
	A3455	Caucasus (NHM)		92.3.15.1	Tooth
	A15668	Vinnicki oblast, Ukraine (Vinnytsia)	367BP	Gp-673 (3)	Skull

	Sample ID	Origin	Field ID	Museum ID	Type
	A15660	Amvrosievka, Ukraine (IAKiev)	A88a	A-88 KB XXIII	Mandible
	LE237A	Amvrosievka, Ukraine (IAKiev)	A89a	A-89 KB VI B	Mandible
	LE242B	Amvrosievka, Ukraine (IAKiev)	A89b	A-89 KB 1	Mandible
	LE247B	Amvrosievka, Ukraine (IAKiev)	A93a	A93 K4 b/33	Mandible
Western Europe	A2791	North Sea bed deposit (NSN)	JGAC09		-
	A2795	North Sea bed deposit (NSN)	JGAC13		-
	A2798	North Sea bed deposit (NSN)	JGAC16		-
	A2808	North Sea bed deposit (NSN)	JGAC26		-
	A2809	North Sea bed deposit (NSN)	JGAC27		-
	A2811	North Sea bed deposit (NSN)	JGAC29		-
	A2792	North Sea bed deposit (NSN)	JGAC10		-
	A2793	North Sea bed deposit (NSN)	JGAC11		-
	A2796	North Sea bed deposit (NSN)	JGAC14		-
	A2797	North Sea bed deposit (NSN)	JGAC15		-
	A2799	North Sea bed deposit (NSN)	JGAC17		-
	A2810	North Sea bed deposit (NSN)	JGAC28		-
	A2800	North Sea bed deposit (NSN)	JGAC18		-
	A2801	North Sea bed deposit (NSN)	JGAC19		-
	A2794	North Sea bed deposit (NSN)	JGAC12		-
	A2802	North Sea bed deposit (NSN)	JGAC20		-
	A2803	North Sea bed deposit (NSN)	JGAC21		-
	A2804	North Sea bed deposit (NSN)	JGAC22		-
	A2805	North Sea bed deposit (NSN)	JGAC23		-
	A2806	North Sea bed deposit (NSN)	JGAC24		-
	A2807	North Sea bed deposit (NSN)	JGAC25		-
	BS593	Steiermark, Austria (VNHM)		H-65-5-2	Femur
	BS600	Steiermark, Austria (VNHM)		H-65-5-4	Femur
	BS607	Oberösterreich, Austria (VNHM)		H-79-48-1	Femur

	Sample ID	Origin	Field ID	Museum ID	Type
	A3226	Schönbrunn Zoo, Austria (MNHN)	BV/58	AC-1894-214	Tooth
	A3227	Western Europe (MNHN)	BV/7	AC-1894-230	Tooth
	A3228	Western Europe (MNHN)	A68	AC-1894-239	Tooth
	A15665	Gouffre de la combe de la racine, Switzerland (ISSKA)		165-10.03	Skull
	A15526	Riparo Tagliente, Italy	ITA2	US 352, RT 91 953/3 4884	
	A15637	Aven de l'Arquet, France (Orgnac)	A24671		Metacarpal
	SGE2	Grotte des Trois-Frères, France	SGE2		
Beringia	A875	Alyoshkina Zaimka, Siberia (PIN)		3658-131	Metacarpal
	A3133	Irish Gulch, Yukon, Canada (YPP)		YT03_204	Astralagus

	Sample ID	C14 ID	Extraction method	Isotope values	
				del13	del15
Urals	A001	OxA-14558	PheChlo	-19.887	3.08
	A003	OxA-14948	PheChlo	-18.993	8.608
	A004	OxA-14545	PheChlo	-19.055	3.976
	A005	OxA-14556	PheChlo	-19.171	4.491
	A006	OxA-14550	PheChlo	-19.244	4.532
	A007	OxA-14548	PheChlo	-19.474	3.816
	A011	OxA-14549	PheChlo	-19.185	4.896
	A016	OxA-14554	PheChlo	-19.176	3.653
	A017	OxA-14553	PheChlo	-19.031	5.484
	A018	OxA-14552	PheChlo	-19.627	3.025
	BS599	OxA-12992	PheChlo	-18.597	5.845
	BS604	OxA-12991	PheChlo	-19.285	3.104
	BS606	OxA-12990	PheChlo	-19.181	6.745
	A002	OxA-14551	PheChlo	-19.436	4.614
	A008	OxA-14555	PheChlo	-19.406	7.238
	A013	OxA-14557	PheChlo	-19.918	5.289
	A014	OxA-14559	PheChlo	-18.998	7.111
	BS592	OxA-12986	PheChlo	-18.793	5.122
	BS660	OxA-12987	PheChlo	-18.791	5.185
	BS674	OxA-14559	PheChlo	-19.004	4.642
	BS708	OxA-12985	PheChlo	-19.621	4.421
	BS713	OxA-12989	PheChlo	-19.221	6.602
	BS588	OxA-12122	PheChlo	-19.176	4.464
	A012		PheChlo		
	A015		PheChlo		
	A4081		PheChlo		
	A4082		PheChlo		
	A4083		PheChlo		

	Sample ID	C14 ID	Extraction method	Isotope values	
				del13	del15
Caucasus	A4084		PheChlo		
	A4085		PheChlo		
	A4087		PheChlo		
	A4088		PheChlo		
	A4089	OxA-19197	PheChlo		
	A4091	OxA-19199	PheChlo		
	A4092	OxA-19200	PheChlo		
	A4094	OxA-19124	PheChlo		
	A4104	OxA-20368	PheChlo		
	A4090	OxA-19198	PheChlo		
	A4093	OxA-19201	PheChlo		
	A4103		PheChlo		
	A4098		PheChlo		
	A4086		PheChlo		
	A4095		PheChlo		
	A4096		PheChlo		
	A4097		PheChlo		
	A4099		PheChlo		
	A4100		PheChlo		
	A4101		PheChlo		
	A4102		PheChlo		
	A15644		Silica		
	A15646		Silica		
	A15648		Silica		
	A15654		Silica		
	A3454		PheChlo		
	A3455		PheChlo		
	A15668	ETH-66330	Silica		

	Sample ID	C14 ID	Extraction method	Isotope values	
				del13	del15
	A15660	Indirect date	Silica		
	LE237A	Indirect date	Silica (UT)		
	LE242B	Indirect date	Silica (UT)		
	LE247B	Indirect date	Silica (UT)		
Western Europe	A2791	OxA-19368	PheChlo		
	A2795	OxA-19372	PheChlo		
	A2798	OxA-19374	PheChlo		
	A2808	OxA-19376	PheChlo		
	A2809	OxA-19377	PheChlo		
	A2811	OxA-19326	PheChlo		
	A2792	OxA-19370	PheChlo		
	A2793	OxA-20370	PheChlo		
	A2796	OxA-19373	PheChlo		
	A2797		PheChlo		
	A2799		PheChlo		
	A2810		PheChlo		
	A2800		PheChlo		
	A2801		PheChlo		
	A2794		PheChlo		
	A2802		PheChlo		
	A2803		PheChlo		
	A2804		PheChlo		
	A2805		PheChlo		
	A2806		PheChlo		
	A2807		PheChlo		
	BS593		PheChlo		
	BS600		PheChlo		
	BS607		PheChlo		

	Sample ID	C14 ID	Extraction method	Isotope values	
				del13	**del15**
	A3226		PheChlo		
	A3227		PheChlo		
	A3228		PheChlo		
	A15665	Ua-42583	Silica		
	A15526	OxA-29834	Silica		
	A15637	OxA-32490	Silica		
	SGE2	UCIAMS-144544	Silica		
Beringia	A875	OxA-29064	PheChlo		
	A3133	OxA-22141	PheChlo		

Chapter 3

Population size history from short scaffolds: how short is too short?

3.1 Authorship statement

Statement of Authorship

Title of Paper	Population size history from short genomic scaffolds: how short is too short?
Publication Status	Submitted for publication; available on preprint server
Publication Details	Graham Gower, Simon Tuke, Adam B Rohrlach, Julien Soubrier, Bastien Llamas, Nigel Bean and Alan Cooper, Population size history from short genomic scaffolds: how short is too short? bioRxiv 382036; doi: https://doi.org/10.1101/382036

Principal Author

Name of Principal Author (Candidate)	Graham Gower		
Contribution to the Paper	Designed the study; performed simulations; processed empirical data; interpreted results; wrote and edited the manuscript.		
Overall percentage (%)	75		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	12/10/2018

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Simon 'Jono' Tuke		
Contribution to the Paper	Designed the study; performed mixed-effects modelling; interpreted results; edited the manuscript.		
Signature		Date	12/10/2018

Name of Co-Author	Adam B. Rohrlach		
Contribution to the Paper	Designed the study; interpreted results; edited the manuscript.		
Signature		Date	15/10/2018

Name of Co-Author	Julien Soubrier		
Contribution to the Paper	Supervised work; interpreted results; edited the manuscript.		
Signature		Date	15 / 10 / 2018

Name of Co-Author	Bastien Llamas		
Contribution to the Paper	Supervised work; interpreted results; edited the manuscript.		
Signature		Date	

Name of Co-Author	Nigel Bean		
Contribution to the Paper	Supervised work; designed the study; interpreted results; edited the manuscript.		
Signature		Date	15/10/2018

Name of Co-Author	Alan Cooper		
Contribution to the Paper	Supervised work; interpreted results; edited the manuscript.		
Signature		Date	12·10·18

3.2 Manuscript

Population size history from short genomic scaffolds: how short is too short?

Graham Gower*,[1], Jono Tuke[2,3], AB Rohrlach[2,3], Julien Soubrier[1,4], Bastien Llamas[1], Nigel Bean[2,3], and Alan Cooper[1]

[1]Australian Centre for Ancient DNA, School of Biological Sciences, The Environment Institute, The University of Adelaide, Adelaide, South Australia 5005, Australia
[2]School of Mathematical Sciences, The University of Adelaide, Adelaide, South Australia 5005, Australia
[3]ARC Centre of Excellence for Mathematical and Statistical Frontiers, The University of Adelaide, Adelaide, South Australia 5005, Australia
[4]Genetics and Molecular Pathology, SA Pathology, Adelaide, South Australia 5000, Australia

*Corresponding author: graham.gower@adelaide.edu.au

Abstract

The Pairwise Sequentially Markov Coalescent (PSMC), and its extension PSMC′, model past population sizes from a single diploid genome. Both models have been widely applied, even to organisms with scaffold-level genome reference assemblies of limited contiguity. However it is unclear how PSMC and PSMC′ perform on short scaffolds. We evaluated `psmc` and `msmc`, implementations of the PSMC and PSMC′ models respectively, on simulated genomes with low contiguity, and compared results to those from fully contiguous data. Simulations with scaffolds from 100 Mb to 10 kb revealed that `psmc` maintains high consistency down to lengths of 100 kb, while `msmc` output is consistent down to 1 Mb. The discrepancy is not due to differing models, but stems from an implementation detail of `msmc`—homozygous tracts at the ends of scaffolds are discarded, making `msmc` unreliable for low contiguity genomes. We recommend excluding data that are aligned to shorter scaffolds when undertaking demographic inference.

Introduction

The process of joining (coalescing) and splitting (recombining) lineages backwards-in-time for a sample of homologous sequences is described by the coalescent with recombination (Hudson, 1990). An important consequence of recombination is that there can be many distinct genealogies, known as marginal genealogies, at different locations along the sequence (Griffiths & Marjoram, 1997). The sequentially Markov coalescent (SMC, McVean & Cardin (2005)) models recombination as a Poisson process left-to-right along the sequence, approximating the coalescent with recombination by treating the marginal genealogy on the right of a recombination as a modification of the marginal genealogy on the left of the recombination. In this sense the approximation is a Markovian process along the sequence, and substantially reduces model complexity for long sequences compared to the full coalescent with recombination (Wiuf & Hein, 1999).

The Pairwise Sequentially Markov Coalescent (PSMC) uses a special case of the SMC approximation, restricted to pairs of sequences, to estimate the distribution of coalescent times within a single diploid genome (Li & Durbin, 2011). PSMC scans along a contiguous segment of the genome and considers marginal genealogies, using their distinct pairwise coalescent times as the unknown states in a hidden Markov model (HMM). To enable parameter estimation, continuous time is approximated by a finite partition of time intervals, and transition probabilities are inferred by Baum-Welch iteration of the forward-backward algorithm. Each genotype at consecutive genomic coordinates provides a new observation for the HMM, a homozygote or a heterozygote, with their emission probabilities determined by the pairwise coalescent time at the current locus, and the genome-wide mutation rate. The population size in a given time interval is inversely proportional to the rate of coalescence, as inferred by maximising the fit of the model to both the HMM transition matrix and the emission probabilities.

The Multiple Sequential Markov Coalescent (MSMC, Schiffels & Durbin (2014)) is an extension to PSMC, and models the distribution of first-coalescent times of two or more haploid sequences. If used with only two haploid sequences, MSMC closely matches the PSMC model, with the exception that it implements SMC$'$ (Marjoram & Wall, 2006), a refinement of SMC incorporating recombinations that immediately coalesce back to the same lineage. For this reason the MSMC model, when applied to a diploid genome, is referred to as PSMC$'$. Compared to PSMC, the genome wide recombination rate is more accurately estimated under the PSMC$'$ model, but population size estimates are qualitatively similar (Schiffels & Durbin, 2014).

Other approaches for inferring population size histories typically require either phased genotypes, multiple individuals, or both (Dutheil *et al.*, 2009; Gutenkunst *et al.*, 2009; Sheehan *et al.*, 2013; Boitard *et al.*, 2016; Terhorst *et al.*, 2017). However, in small scale studies of non-model organisms, it is common for only one individual, or a few individuals, from a single population to be sequenced, and genotypes are unlikely to be phased. Population size history, particularly in the recent past, can also be estimated from the length distribution of tracts of identity-by-descent (Palamara *et al.*, 2012), identity-by-state (Harris & Nielsen, 2013), or runs of homozygosity (MacLeod *et al.*, 2013). While potentially useful for a single diploid individual, such approaches are not readily applicable to short scaffolds, where such tracts may be broken across scaffold boundaries. In contrast, PSMC and PSMC′ are very attractive as they require only diploid genotypes for a single individual, which need not be phased.

By using the sequentially Markovian approximation, PSMC and derived methods implicitly assume that genomic information is contiguous. While initially applied to human datasets, which have very high contiguity, PSMC and PSMC′ have since been applied to many non-model organisms where the contiguity of genomic sequences may be poor (Zhao *et al.*, 2013; Dobrynin *et al.*, 2015; Mays *et al.*, 2018; Kozma *et al.*, 2016; Feigin *et al.*, 2018). In particular, demographic history is regularly inferred from a *de novo* assembly as part of genome sequencing projects. Due to time and funding constraints, genome assemblies are often constructed from only short read sequencing data, and assembled into contigs or short scaffolds. These cannot be ordered or oriented with respect to one another (violating the SMC model), nor anchored to physical chromosomes. Where sequencing data is aligned to such assemblies, the genomic information used for population size inference inherits the low contiguity of the assembly. While small gaps in coverage along a scaffold can be handled gracefully, the HMM must be applied separately to each distinct scaffold, and it is not clear what the length threshold is to obtain robust population size inferences.

Results and Discussion

Simulations

To assess the impact of reference genome contiguity on population size estimates, we simulated genomes for populations with three different demographic histories: a constant population size; a bottleneck; and recovery following a bottleneck (Fig. 1A). For each demographic scenario, we simulated 10 independent populations and sampled 20 × 100 Mb haploid chromosomes, representing 10 diploid genomes from each population. New datasets were then created by fragmenting each genome into equally sized scaffolds at four distinct lengths, 10 Mb, 1 Mb, 100 kb, and 10 kb. Population size histories were

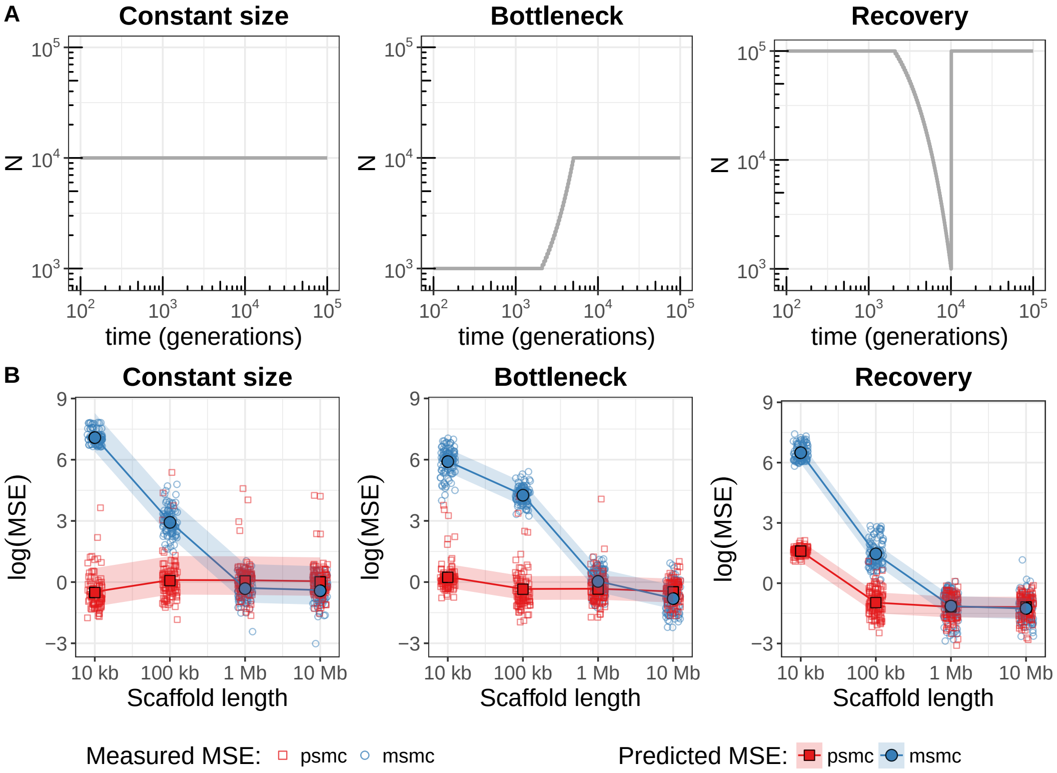


Figure 1: A) Simulated population size histories. **B)** Mean squared error (MSE) of population size inferences from simulations shown immediately above. Larger values indicate a loss of fidelity in the population size estimate. Small hollow markers indicate MSE for distinct simulated individuals (100 Mb per individual; 10 individuals each from 10 populations), with red squares for `psmc` and blue circles for `msmc`. Data from each simulated individual was artificially fragmented to emulate genome sequences aligned to a scaffold-level reference assembly. At each scaffold length, MSE was calculated by comparing to inferences from unfragmented (100 Mb) scaffolds (see methods). Large solid markers and lines show predicted MSE from a linear mixed effect model, with 95% prediction intervals based on simulation.

then inferred for all fragmented and unfragmented datasets using psmc (Li & Durbin, 2011) and msmc (Schiffels & Durbin, 2014), implementations of PSMC and PSMC′ respectively.

Mean squared error

In measuring the error of estimates, Li & Durbin (2011) compared population size inferences to the values that were simulated, but excluded time intervals in the recent and distant past. Population size estimates are expected to be unreliable for times outside a certain range since a typical genome contains relatively few breakpoints corresponding to recombination events in the very recent or very distant past. However, excluding temporal intervals requires advance knowledge of where the method may lose resolution, and this is dependent upon the population size history itself.

To quantify estimation error, we used inferences from the unfragmented datasets as the 'truth', not the values that were simulated. A loess smooth function (Cleveland *et al.*, 1992) was fitted to the unfragmented inferences for each simulated population, separately for psmc and msmc, using population size estimates from all individuals in a given population. Then for each simulated individual, the mean squared error (MSE) was measured between estimates from the fragmented datasets and the loess function for the corresponding population. The MSE was weighted, in discrete time intervals, using the inverse of the sample variance in estimates from the unfragmented datasets (the same individuals as used for the loess fit). This was done to avoid measuring error caused by limited genomic information about the recent and ancient past.

Comparisons of the MSE at each fragmentation level (Fig. 1B) suggest that shorter scaffolds do indeed result in population size estimates that are not consistent with those for longer scaffolds. Qualitatively, msmc appears to decline in fidelity at scaffold lengths between 1 Mb and 100 kb for all demographic scenarios, whereas psmc declines in fidelity only in the Recovery scenario, at scaffold lengths between 100 kb and 10 kb.

Mixed effects model

To determine if the observed differences were significant, we fitted a linear mixed-effects model separately for each demographic scenario. The fixed effects were scaffold length and estimation program (psmc vs. msmc), and a random intercept was necessary to account for the repeated measures of each individual at multiple scaffold lengths. Both scaffold length and estimation program were found to be significant predictors of MSE in all demographic scenarios. Two-way interactions between scaffold length terms and estimation program were also significant in all scenarios.

Empirical data

Arguably, the simulated population history scenarios are unrealistic. Simulated data also provides the best possible case in terms of missing data in that there is none. To gauge the impact of using a scaffold-level assembly with real data, we artificially fragmented chromosome 1 from a high coverage human genome, HG00419, a Southern Han Chinese female (The 1000 Genomes Project Consortium, 2015). Population size histories were again estimated using `psmc` and `msmc`, for each of the fragmented and unfragmented datasets (Fig. 2).

Both programs produced largely the same demographic history when processing long scaffolds, although `msmc` did not estimate population sizes for time intervals as far into the past as `psmc` (3 Mya vs. 10 Mya). For 10 kb scaffold lengths, inferences from `msmc` are substantially different to those using longer scaffolds, and a small departure is also discernible in the recent past for 100 kb scaffolds. Estimates from `psmc` have noticeably poorer resolution at the 10 kb scaffold length, but are remarkably consistent for longer scaffolds.

The data conversion script provided with `psmc` (`fq2psmcfa`) ignores scaffolds having fewer than 10000 genotype calls by default. This excluded most of the 10 kb scaffolds, due to the presence of one or more missing genotypes. Disabling this filter to retain all scaffolds only marginally improved population size estimates, and only in more ancient time intervals (results not shown). We considered the possibility that with 10 kb scaffolds, `psmc` might still closely recapitulate the results from longer scaffolds if provided with more information. To this end chromosome 2 was also partitioned into 10 kb scaffolds and appended to the chromosome 1 data (doubling the information to ∼500 Mb in total). However, the additional information did not alter the result.

`msmc` discards homozygous tracts at the ends of scaffolds

An input file for `msmc` contains lines that specify the coordinate of a heterozygote site and its distance from the previous heterozygote on the same scaffold. Nothing is specified for coordinates after the last heterozygote, and the scaffold is implicitly truncated here. For short scaffolds this causes substantial information loss. Indeed, short scaffolds may contain no heterozygote sites at all, and input files for such scaffolds are empty.

To determine if truncation was a major cause of the different behaviour between `psmc` and `msmc`, we ran `psmc` on 10 kb scaffolds that were artificially

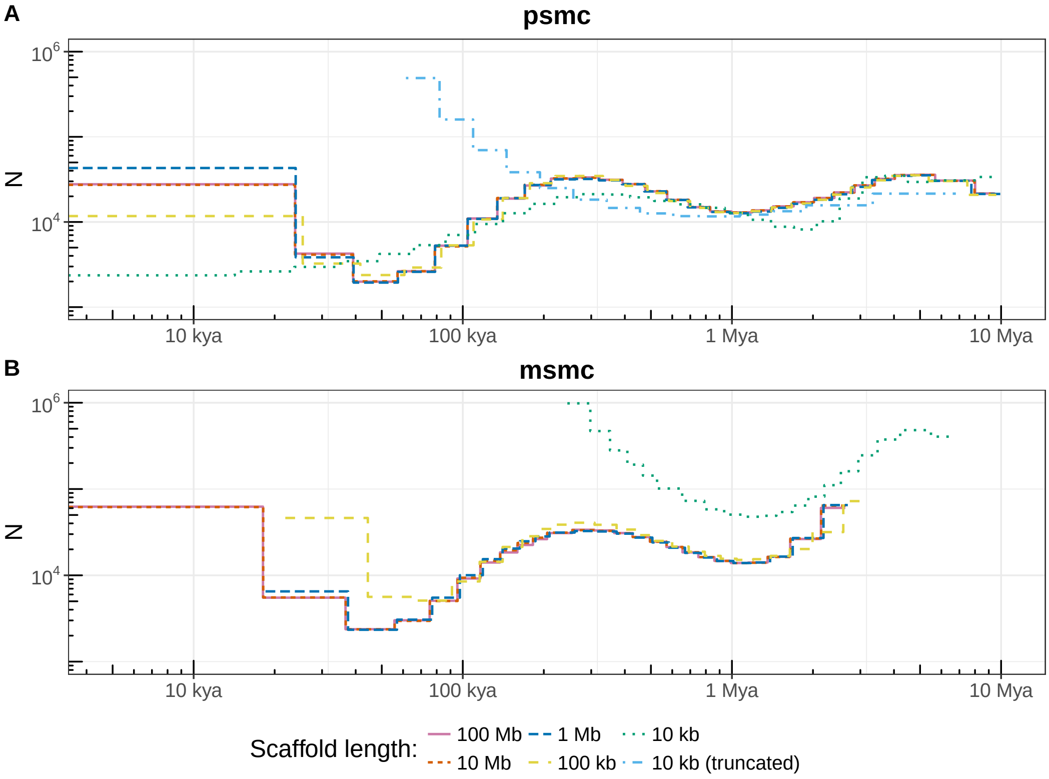


Figure 2: Population size history of HG00419, a Southern Han Chinese individual (The 1000 Genomes Project Consortium, 2015), inferred by **A)** `psmc` and **B)** `msmc`. Empirical data was artificially fragmented to emulate genome sequences aligned to scaffold-level reference assemblies. Population size inferences from `psmc` are consistent down to 100 kb scaffold lengths, with loss of resolution at 10 kb. For `msmc`, stable inferences can be made down to 1 Mb, but fidelity at 100 kb is poor in the recent past, and at 10 kb even broad demographic trends are difficult to discern. Input data to `psmc` for the '10 kb (truncated)' line style had trailing homozygous sites removed from all scaffolds, to match the information content of `msmc` input. Plots were scaled to real time using a 25 year generation time and 1.25*e*-8 mutations per base per generation. kya: thousand years ago; Mya: million years ago;

truncated to match the information available to msmc. Scaffolds containing no heterozygotes were omitted. This output ('10 kb (truncated)' in Fig. 2A), shows a similar trend to that for msmc on 10 kb scaffolds, although differences remain.

Marginal genealogies with recent coalescent times have accumulated few mutations, so corresponding regions of the genome contain mostly homozygote genotypes. Truncation increases the proportion of heterozygotes, hence recent coalescent times appear older. On short scaffolds, all marginal genealogies are near a scaffold end, so inferences from short truncated scaffolds are more strongly biased to not observe recent coalescent events. Since the population size for each time interval is inversely related to the rate at which pairs of haplotypes coalesce, the smaller number of observations of high homozygosity genomic tracts also means that population size inferences are biased upwards. Both artefacts are noticeable, particularly in the more recent time bins, for psmc with artificially truncated 10 kb scaffolds (Fig. 2A) and for msmc with 10 kb and 100 kb scaffolds (Fig. 2B).

Conclusion

Reasonable parameter inference in a hidden Markov model relies on observations leading up to, and following, transitions in state. For PSMC, this corresponds to having sufficient sequence contiguity to observe genomic tracts on both sides of historical recombination breakpoints. The chance that a short scaffold will contain a tract covering a recombination breakpoint depends not only on the completeness of the reference assembly, but also the sparsity of breakpoints.

Several factors contribute to breakpoint density, including population size, the per base recombination rate, and recombination hotspots. A population suffering a recent and very severe bottleneck will give rise to mostly recent pairwise coalescent times, and few recombination breakpoints, both of which are poorly represented within short scaffolds. Our simulations considered a mammalian recombination rate (3.125×10^{-9} per base per generation) and population size histories that are relevant to many taxa. This suggests that PSMC inference can be reasonable from scaffolds as short as 100 kb for a wide range of datasets.

Scaffold level reference assemblies are unlikely to contain equally sized scaffolds, as evaluated here. Generally, a scaffold-level assembly contains tens of long scaffolds and tens of thousands of short scaffolds. In such cases, it is reasonable to exclude scaffolds shorter than 100 kb when running psmc, and scaffolds shorter than 1 Mb for use with msmc. However, we caution that this guideline may be too optimistic for severely bottlenecked populations or genomic data aligned to a very low quality reference assembly.

Materials and Methods

Simulations

Simulations were performed using scrm (Staab *et al.*, 2015), with mutation rate $\mu = 1.25 \times 10^{-8}$ per base per generation and recombination rate $\mu/4$ per base per generation (Schiffels & Durbin, 2014). Simulation output was artificially fragmented during conversion to psmc and msmc input formats, using a custom Perl script. Demographic inferences were obtained from psmc v0.6.5-r67 and msmc v1.0.0 for all inputs. Both psmc and msmc were run with the same time bin parameter (-p 1*2+15*1+1*2), although we note that each program calculates time boundaries for the discrete bins differently, so a completely fair comparison is not possible. Scripts used for simulation, format conversion, and running psmc/msmc are available from https://github.com/grahamgower/psmc-error-analysis/.

Mean squared error

For each simulated population history scenario and each estimation program, estimates from the unfragmented datasets were used to fit a loess function of log population ($\log(N)$) against log time ($\log(t + 10)$). The offset of 10 was based on a sensitivity analysis and the smallest non-zero time. An optimal value for the loess smoothing parameter was selected by maximising the corrected AIC (AICc) (Hurvich *et al.*, 1998). Mean squared error for individual i in population j was calculated as

$$MSE_{ij} = \frac{1}{k} \sum_{m=1}^{k} (n_{ijm} - \tilde{n}_{\cdot jm})^2 / var_j(m),$$

where the sum extends over all k time intervals, n_{ijm} is the log of the population size estimate in interval m, and $\tilde{n}_{\cdot jm}$ is the prediction for the mth time interval from the loess function fitted for the jth population. The variance step function $var_j(m)$ at time interval m, for the jth population, was calculated by splitting time on a log scale into 10 even-width bins and calculating the variance in each bin.

Mixed effects modelling

Scatter-plots of MSE against scaffold length indicated a cubic relationship between MSE and log(scaffold length). This was confirmed by comparing residual plots for linear, quadratic, and cubic models. To help numerical consistency of the fitting process, we performed a location scaling of log(scaffold length).

Bivariate analysis of each of the predictors—log(scaffold length), estimation program, population history scenario, sample ID, and population ID—were used for variable selection. Only log(scaffold length), estimation program, and population history scenario had a significant relationship with MSE.

The linear mixed effects model was fitted using the `lme4` package (Bates *et al.*, 2015) in R (R Core Team, 2017). The fixed effects were log(scaffold length) and estimation program. Up to two-way interaction terms were considered for each of the cubic log(scaffold length) terms with estimation program. To account for repeated measures from each simulated individual due to multiple levels of fragmentation, we included random effects. Both random intercepts and random slopes were considered.

All significance testing was performed using the `lmerTest` package (Kuznetsova *et al.*, 2017). All assumptions of the linear mixed-effects models were assessed and regarded as reasonable. The 95% prediction intervals were based on simulation with the `merTools` package (Knowles & Frederick, 2016).

Empirical dataset

We downloaded the cram alignment file for HG00419, aligned to assembly GRCh38DH, from The 1000 Genomes ftp server, and called genotypes with `samtools -q20 -Q20 -C50 ... | bcftools call -c ...`. The resulting vcf was partitioned into scaffolds of a specific size by modifying the chromosome name and position to which each genotype call corresponded, and was performed separately for each of the scaffold sizes 100 Mb, 10 Mb, 1 Mb, 100 kb, and 10 kb. Input for both `psmc` and `msmc` were filtered to exclude sites with less than half, or greater than double, the mean depth (54.76). The vcf was converted to `psmc` input format with `vcfutils.pl` (distributed with `samtools`) and `fq2psmcfa` (distributed with `psmc`), then `psmc` was run with time bin parameter `-p 4+25*2+4+6`. The same vcf was converted to `msmc` input format with `bamCaller.py` and `generate_multihetsep.py`, both distributed with `msmc-tools`, then `msmc` was run with parameters `-R -p 15*1+15*2`. The time bin parameters for both programs were chosen to be suitable for inferring human demography (Li & Durbin, 2011; Schiffels & Durbin, 2014).

Author Contributions

GG, JT, ABR, JS, BL, and NB designed the study. GG performed simulations and processed the empirical data. JT calculated the MSE and performed mixed effects modelling. All authors interpreted the results. GG wrote the manuscript with feedback from all coauthors.

Acknowledgements

This work was supported by an Australian Government Research Training Program Scholarship [GG], and research fellowships from the University of Adelaide and the Australian Research Council [BL].

References

Bates D, Mächler M, Bolker B, & Walker S (2015). Fitting linear mixed-effects models using lme4. *J Stat Softw.*, **67(1)**:1–48. http://dx.doi.org/10.18637/jss.v067.i01

Boitard S, Rodríguez W, Jay F, Mona S, & Austerlitz F (2016). Inferring Population Size History from Large Samples of Genome-Wide Molecular Data - An Approximate Bayesian Computation Approach. *PLoS Genet.*, **12(3)**:e1005877. http://dx.doi.org/10.1371/journal.pgen.1005877

Cleveland W, Grosse E, & Shyu WM (1992). *Statistical Models in S*, chapter Local regression models, pp. 309–375. Wadsworth & Brooks/Cole, Belmont (CA)

Dobrynin P, Liu S, Tamazian G, Xiong Z, Yurchenko AA, Krasheninnikova K, Kliver S, Schmidt-Küntzel A, Koepfli KP, Johnson W, *et al.* (2015). Genomic legacy of the African cheetah, *Acinonyx jubatus*. *Genome Biol.*, **16**:277. http://dx.doi.org/10.1186/s13059-015-0837-4

Dutheil JY, Ganapathy G, Hobolth A, Mailund T, Uyenoyama MK, & Schierup MH (2009). Ancestral population genomics: The coalescent hidden Markov model approach. *Genetics*, **183(1)**:259–274. http://dx.doi.org/10.1534/genetics.109.103010

Feigin CY, Newton AH, Doronina L, Schmitz J, Hipsley CA, Mitchell KJ, Gower G, Llamas B, Soubrier J, Heider TN, *et al.* (2018). Genome of the

Tasmanian tiger provides insights into the evolution and demography of an extinct marsupial carnivore. *Nat Ecol Evol.*, **2(1)**:182–192. http://dx.doi.org/10.1038/s41559-017-0417-y

Griffiths RC & Marjoram P (1997). An ancestral recombination graph. In *Progress in population genetics and human evolution*, pp. 257–270. Springer, New York (NY)

Gutenkunst RN, Hernandez RD, Williamson SH, & Bustamante CD (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.*, **5(10)**:e1000695. http://dx.doi.org/10.1371/journal.pgen.1000695

Harris K & Nielsen R (2013). Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genet.*, **9(6)**:e1003521. http://dx.doi.org/10.1371/journal.pgen.1003521

Hudson RR (1990). Gene geneologies and the coalescent process. In D Futuyma & J Antonovic, eds., *Oxford Surveys in Evolutionary Biology*, volume 7, pp. 1–44. Oxford University Press, New York (NY)

Hurvich CM, Simonoff JS, & Tsai CL (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J R Stat Soc Series B Stat Methodol.*, **60(2)**:271–293

Knowles JE & Frederick C (2016). *merTools: Tools for Analyzing Mixed Effect Regression Models*. R package version 0.3.0

Kozma R, Melsted P, Magnússon KP, & Höglund J (2016). Looking into the past - the reaction of three grouse species to climate change over the last million years using whole genome sequences. *Mol Ecol.*, **25(2)**:570–580. http://dx.doi.org/10.1111/mec.13496

Kuznetsova A, Brockhoff PB, & Christensen RHB (2017). lmerTest package: Tests in linear mixed effects models. *J Stat Softw.*, **82(13)**:1–26. http://dx.doi.org/10.18637/jss.v082.i13

Li H & Durbin R (2011). Inference of human population history from individual whole-genome sequences. *Nature*, **475(7357)**:493–496. http://dx.doi.org/10.1038/nature10231

MacLeod IM, Larkin DM, Lewin HA, Hayes BJ, & Goddard ME (2013). Inferring demography from runs of homozygosity in whole-genome sequence, with correction for sequence errors. *Mol Biol Evol.*, **30(9)**:2209–2223. http://dx.doi.org/10.1093/molbev/mst125

Marjoram P & Wall JD (2006). Fast "coalescent" simulation. *BMC Genet.*, **7**:16. http://dx.doi.org/10.1186/1471-2156-7-16

Mays HL, Hung CM, Shaner PJ, Denvir J, Justice M, Yang SF, Roth TL, Oehler DA, Fan J, Rekulapally S, *et al.* (2018). Genomic analysis of demographic history and ecological niche modeling in the endangered Sumatran rhinoceros *Dicerorhinus sumatrensis*. *Curr Biol.*, **28(1)**:70–76. http://dx.doi.org/10.1016/j.cub.2017.11.021

McVean GAT & Cardin NJ (2005). Approximating the coalescent with recombination. *Philos Trans R Soc Lond B Biol Sci.*, **360(1459)**:1387–1393. http://dx.doi.org/10.1098/rstb.2005.1673

Palamara PF, Lencz T, Darvasi A, & Pe'er I (2012). Length distributions of identity by descent reveal fine-scale demographic history. *Am J Hum Genet.*, **91(5)**:809–822. http://dx.doi.org/10.1016/j.ajhg.2012.08.030

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria

Schiffels S & Durbin R (2014). Inferring human population size and separation history from multiple genome sequences. *Nat Genet*, **46(8)**:919–925. http://dx.doi.org/10.1038/ng.3015

Sheehan S, Harris K, & Song YS (2013). Estimating variable effective population sizes from multiple genomes: A sequentially Markov conditional sampling distribution approach. *Genetics*, **194(3)**:647–662. http://dx.doi.org/10.1534/genetics.112.149096

Staab PR, Zhu S, Metzler D, & Lunter G (2015). scrm: efficiently simulating long sequences using the approximated coalescent with recombination. *Bioinformatics*, **31(10)**:1680–1682. http://dx.doi.org/10.1093/bioinformatics/btu861

Terhorst J, Kamm JA, & Song YS (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat Genet.*, **49(2)**:303–309. http://dx.doi.org/10.1038/ng.3748

The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature*, **526(7571)**:68–74. http://dx.doi.org/10.1038/nature15393

Wiuf C & Hein J (1999). Recombination as a point process along sequences. *Theor Popul Biol.*, **55(3)**:248–259. http://dx.doi.org/10.1006/tpbi.1998.1403

Zhao S, Zheng P, Dong S, Zhan X, Wu Q, Guo X, Hu Y, He W, Zhang S, Fan W, *et al.* (2013). Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nat Genet.*, **45(1)**:67–71. http://dx.doi.org/10.1038/ng.2494

Chapter 4

Widespread male sex bias in mammal fossil and museum collections

4.1 Authorship statement

Statement of Authorship

Title of Paper	Widespread male sex bias in mammal fossil and museum collections
Publication Status	Unpublished and unsubmitted work written in manuscript style

Principal Author

Name of Principal Author (Candidate)	Graham Gower		
Contribution to the Paper	Performed sex determination; logistic regression analysis; kernel test; interpreted results; wrote and edited the manuscript.		
Overall percentage (%)	80		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	12/10/2018

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Lindsey Fenderson		
Contribution to the Paper	Preparation of bison samples for sequencing; mapped sequencing data.		
Signature		Date	13.10.2018

Name of Co-Author	Alexander Salis		
Contribution to the Paper	Preparation of brown bear samples for sequencing; mapped sequencing data.		
Signature		Date	13/10/18

Name of Co-Author	Kristofer M. Helgen		
Contribution to the Paper	Expertise on mammals and mammal databases; interpreted results; wrote and edited the manuscript.		
Signature		Date	13/10/2018

Name of Co-Author	Ayla L. van Loenen		
Contribution to the Paper	Preparation of bison samples for sequencing; mapped sequencing data.		
Signature		Date	12/8/18

Name of Co-Author	Holly Heininger		
Contribution to the Paper	Preparation of samples for sequencing.		
Signature		Date	15/10/18

Name of Co-Author	Kieren J. Mitchell		
Contribution to the Paper	Supervised work; interpreted results; edited the manuscript.		
Signature		Date	12.10.18

Name of Co-Author	Bastien Llamas		
Contribution to the Paper	Supervised work; interpreted results.		
Signature		Date	15/10/2018

Name of Co-Author	Alan Cooper		
Contribution to the Paper	Supervised work; interpreted results; wrote and edited the manuscript.		
Signature		Date	12.10.18

4.2 Manuscript

Widespread male sex bias in mammal fossil and museum collections

Graham Gower[a,1], Lindsey Fenderson[a], Alexander Salis[a], Kristofer M. Helgen[b], Ayla L. van Loenen[a], Holly Heiniger[a], Kieren J. Mitchell[a], Bastien Llamas[a], Alan Cooper[a,1]

[a]Australian Centre for Ancient DNA, The University of Adelaide.
[b]School of Biological Sciences, The University of Adelaide.
[1]To whom correspondence should be addressed:
graham.gower@adelaide.edu.au, alan.cooper@adelaide.edu.au.

Abstract

The sexing of subfossil material using relatively low coverage high-throughput DNA sequencing methods was recently used to show a male-biased sex ratio in mammoth remains and predict the same pattern for steppe bison (Pečnerová *et al.*, 2017). We genetically sexed subfossil remains of 186 Holarctic bison (*Bison spp.*) and 91 brown bears (*Ursus arctos*), and found that approximately 75% of both are male, very similar to the ratio observed in mammoths (72%). We found no evidence for differences between the sexes with respect to: DNA preservation, sample age, material type, or spatial distribution. However, bison and brown bear remains preserved in caves exhibited a different sex ratio to other sedimentary deposits. We also examined ratios of male and female specimens from four large museum collections of hunted and trapped mammals and again found a strong male bias with the species-averaged percentage greater than 50% in almost all mammalian orders. We suggest: (1) wider male geographic ranges can lead to considerably increased chances of detection in fossil studies, and (2) sexual dimorphic behaviour or appearance can facilitate a bias in fossil and modern mammal collections towards males, or the more visually striking sex. These findings reveal a sex bias on a previously unacknowledged scale, which have major implications for a wide range of studies of museum material that require specimens to be an unbiased representation of their population.

Introduction

Most mammal species have a sex ratio of 1:1 at birth (Karlin & Lessard, 1986), but this may shift demographically according to differential patterns of mortality between the sexes across various life stages. A variety of factors have been identified that may affect sex ratios in mammal populations from birth to adulthood, including competition for mates and local resources, or the physiological condition of mothers (Trivers & Willard, 1973; Charnov, 1975; Karlin & Lessard, 1986). The sex ratios in natural populations are helpful in evaluating the impact of these and other factors, and to illuminate aspects of life history and comparative demographics within and across species. However, it is important that field-based studies of sex ratios capture real, rather than biased, information for both sexes. Pečnerová *et al.* (1) recently demonstrated that males are over-represented in the fossil record of mammoths, and suggested that this also may be the case for the fossil record of other female herd-based mammal species, such as bison. To explore the extent of this problem we examined the relative representation of males and females in the fossil record of Late Pleistocene and Holocene bison (*Bison* spp.) and brown bears (*Ursus arctos*), as well as in museum collections of a range of extant mammals.

Morphological sex determination of fossil and subfossil remains is generally reliable only where sexual dimorphism is apparent, but has been widely used despite this limitation (Frayer & Wolpoff, 1985; Rehg & Leigh, 1999). However, it is also possible to genetically sex subfossil specimens using ancient DNA, either by direct PCR of a sex-linked gene or more powerfully via shotgun sequencing data (Skoglund *et al.*, 2013; Mittnik *et al.*, 2016). In the latter approach, mammalian sex may be inferred by calculating the ratio of the number of reads that map to the Y versus X chromosomes (Skoglund *et al.*, 2013), although because many genome reference assemblies lack a Y chromosome it is often better to calculate the ratio of reads mapping to the X versus non-sex chromosomes (Mittnik *et al.*, 2016). Because females have two X chromosomes, and males have only one, the X chromosomal *read dosage* is approximately double in females compared with males. Read dosage for both X and Y has also been evaluated for ancient DNA in conjunction with nuclear SNP capture data (Fu *et al.*, 2016). The use of read dosage is very convenient for ancient DNA studies, as the method requires relatively little sequencing effort, and is typically generated as part of routine DNA quality screening.

The read dosage approach was recently used to show that male specimens are over-represented in Holarctic mammoth remains (Pečnerová *et al.*, 2017). This was suggested to result from the “lone male model”, originally proposed to explain the excess of young adult males in the Hot Springs mammoth assemblage (Agenbroad & Mead, 1987). This model proposes that after subadult males are expelled from their familial group, they lose the protection of a large herd and experienced group leaders, and consequently engage in riskier behaviour or enter more dangerous territory. As a result, the excess of males is caused by segregation of sexes due to their social behaviour leading to differential mortality, including at taphonomically favourable sites which preserve fossils (such as bogs and tarpits, and river crossings upstream of fluvial deposits). Morphological age profiling has provided support for this model at specific mammoth mass death sites (reviewed by Haynes, 2017), but it has not previously been suggested as a more widespread pattern across the fossil record. Furthermore, the model is not readily falsifiable without profiling age at death, and other possible causes for a male bias also remain untested.

To investigate this issue further, we examined large collections of two other Late Pleistocene Holarctic megafauna, bison (*Bison* spp.) and brown bears (*Ursus arctos*) from across Europe, Beringia, and North America, along with the original mammoth dataset (Pečnerová *et al.*, 2017) and a small dataset of the extinct Balearic bovid *Myotragus balearicus*. Most of the specimens were collected by the authors either directly from the field (most of the North American samples) or from existing museum collections (most of the European and Russian samples), providing some level of control against collection biases. We used these datasets to investigate a number of aspects of sample taphonomy and collection activities that might influence their observed sex ratios.

Late Pleistocene bison thrived on the vast mammoth steppe, leaving a substantial fossil record across Eurasia and North America. Modern bison are polygynous and gregarious, forming large herds comprised mostly of female adults and young of both sexes. Adult males are solitary or form small bachelor groups, joining with the female groups for only 1-2 months of the year. Similar structures have been implied for Pleistocene steppe bison (Guthrie, 1989), and this has led to predictions that, like mammoth, steppe bison remains would also exhibit a pronounced male bias (Pečnerová *et al.*, 2017). We examined this by genetically sexing 188 subfossil bison specimens from across Europe, Beringia, and North America, mostly recovered from alluvial sediments.

Both modern and Late Pleistocene brown bears have a Holarctic distribution, and individuals are typically either solitary or form small family groups, only congregating in large numbers under atypical circumstances of highly abundant food. Dispersal of extant brown bears is density dependent (Støen *et al.*, 2006), with more than one third of females and 80-90% of males, dispersing before adulthood (Støen *et al.*, 2006; Zedrosser *et al.*, 2007). Given that brown bears are facultative carnivores, both their ecology and social structure are clearly different to mammoths and bison and provide an additional test of biased sex ratios. We genetically sexed 92 brown bear subfossils from Europe, Russia, and North America, recovered from caves and alluvial sediments.

Table 1: Male and female sample counts. †Mammoth data is from (Pečnerová *et al.*, 2017).

	Bison			Brown Bears			†Mammoths
	all	postcrania	non cave	all	Alps	non Alps	
Males	139	72	135	58	8	50	67
Females	47	31	39	33	16	17	26
Total	186	103	174	91	24	67	93
% male	74.73	69.90	77.59	63.74	33.33	74.63	72.04
Unassigned	2	0	2	1	0	1	5

Results

Shotgun sequencing data were used to confidently assign sex to 186 subfossil bison and 91 brown bear specimens from across Europe, Beringia, and North America using the ratio of reads mapping to the X chromosome versus non-sex chromosomes (Methods, Table 1). A pronounced male sex bias close in size to that of mammoths (72%) was observed across all bison (75%) and the vast majority of the brown bear specimens (75%) (Table 1). Interestingly, in the small sets of cave-preserved bones a contradictory signal of female bias was observed for both bison (5 males, 8 females), and brown bears from the Alps (8 males, 16 females). The dominance of female brown bears has previously been noted for Austrian caves (Döppes & Pacher, 2014), and is thought to relate to behavioural differences in the Alps region where female bears hibernate in caves, whereas males do not. Outside of the Alps, both male and female brown bears hibernate, and a strong male sex bias was observed in cave sites (50 males, 26 females) while open sites showed a more equal ratio (8 males, 7 females).

To test whether additional information about the samples could explain the excess male ratio we used an intercept-only logistic regression, as a null model, for comparison with logistic regression models containing explanatory variables. Intuitively, this null model can be interpreted as ‘there is a fixed ratio of males to females’, while the alternative models that we construct should be interpreted as ‘the sex ratio changes as the explanatory variable changes’. Alternative models were compared to the null using a likelihood ratio test. Logistic regression models with univariate predictors of sex were constructed for a variety of explanatory variables (Table 2).

Table 2: Logistic regression models with sex as the dependent variable. The row corresponding to an intercept-only model shows p-values for the intercept term, which tests the null hypothesis that there is a 1:1 male to female ratio. All other cells contain p-values from likelihood ratio tests, comparing a logistic regression model of the form 'sex ∼ X', where X is a single explanatory variable, to the intercept-only model above it. Material1 consists of factors such as tooth, leg, astragalus, foot, petrous, other skull, vertebrae, flat bone, horn. Material2 collapses factors from Material1 into crania and non-crania. †Mammoth data is from (Pečnerová *et al.*, 2017).

Explanatory variable	Bison			Brown Bears			†Mammoths
	all	postcrania	non cave	all	Alps	non Alps	
Intercept-only	***1.31E-10***	***8.80E-05***	***8.51E-12***	***0.00973***	0.110	***0.000122***	***4.21E-05***
Cave/non-cave	***0.00176***	***0.00646***		0.367		***0.0399***	
Material1	0.618	0.634	0.716	0.264	0.758	0.0695	
Material2	0.227		0.245	0.594	0.671	0.590	0.132
^{14}C age	0.768	0.534	0.614	***0.0122***	0.133	0.174	0.992
Latitude	0.954	0.657	0.682	0.619	0.494	***0.0244***	
Longitude	0.490	0.527	0.965	***0.0171***	0.708	0.417	
Altitude	0.676	0.802	0.847	***0.0157***	0.158	0.911	
Alps/non-Alps				***0.000363***			
Endogenous	0.707	0.790	0.941	0.137	0.521	0.439	
GC ratio	0.312	0.625	0.468	0.723	0.386	0.168	
DNA fragment length	0.237	0.343	0.705	0.352	0.717	0.514	
5' deamination (C→T)	0.558	0.681	0.644	0.162	0.446	0.148	

Bison

For the bison, only the type of site (cave versus non-cave) was found to be significantly better than the intercept-only model, due to the female bias in the 12 cave specimens noted above. We searched for site specific factors that might contribute to differential mortality of males and females, but rejected univariate models with the following explanatory variables: latitude, longitude, and altitude. Univariate models may not reveal differences that arise only when jointly considering latitude and longitude, so we implemented a Gaussian kernel two-sample test (Gretton *et al.*, 2012), to look for more complex spatial differences between the sexes. This multivariate test has good sensitivity to test such differences (see Supplementary Information), but was unable to reveal any sex specific patterns for bison remains.

To examine whether larger bison might generate a 'trophy' collection bias we searched for an increase in the proportion of male bone samples where sexual dimorphism is more apparent (e.g. skulls). Due to the small sample size of many types of bone used for DNA extraction, we also collapsed the categories into either 'crania' or 'postcrania' with teeth placed into the crania category as they are regularly taken from full or partial skulls. Neither the model containing all bone categories, nor collapsed categories, was significantly better than the null.

Brown bears

While several variables (^{14}C age, longitude, and altitude) explained the brown bear male sex bias better than an intercept-only model (Table 2), these are all related to the strong female bias in the Alps cave samples (p=0.0003). Outside of the Alps region, the only variables significantly better than an intercept-only model (Table 2) were latitude and cave sites. The male bias was more extreme at lower latitudes, which is consistent with the lone male model as female home ranges are larger in higher latitudes due to food scarcity, particularly after emerging from dens (Bunnell & Tait, 1981). Interestingly, brown bear bones found in caves outside the Alps showed a male bias, suggesting the female hibernation behaviour in the Alps may indeed be producing the female sex bias, while elsewhere males dominated caves as preferred denning sites.

The kernel two-sample test applied to bison was also applied to brown bears, which identified the sex specific spatial distribution caused by sites in the Alps. However, when applied to only brown bear remains outside the Alps, no spatial differences between the sexes could be identified.

Mammoth

We also reanalysed the mammoth samples from the previous study (Pečnerová *et al.*, 2017) for comparison, using our methods for consistency. Of 98 samples, 93 were unambiguously assigned to a sex (Table 1). We evaluated the two variables given, material type and ^{14}C date, as possible explanations of the sex ratio. Neither were significantly better than an intercept-only model (Table 2).

Myotragus

We sexed nine bones of the fossil dwarf bovid *Myotragus balearicus* from several different Mallorcan deposits (Balearic Islands, Spain). Larger bones were deliberately chosen from available collections (as part of another study Bover et al. submitted) in an effort to identify specimens with good DNA preservation. All nine bones were found to be male, suggesting the deliberate choice of

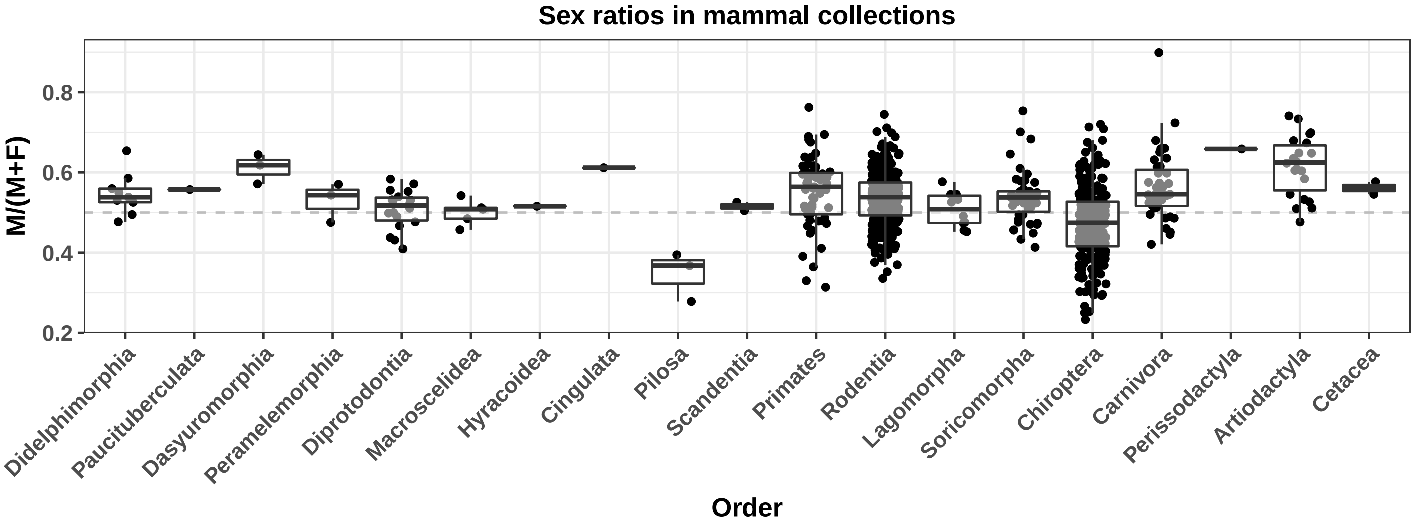


Figure 1: Boxplot showing the proportion of males samples for distinct species in modern mammal collections, grouped by order. Black dots represent the proportion of males for a single species, and are jittered horizontally. Only species with more than 100 sexed samples were included.

large bones in medium-small size species can result in a substantial male bias for taxa that have obvious sexual size dimorphism.

Modern mammal collections

To further explore the potential for biases in museum collections we counted male and female samples in the online databases of large mammalogy collections from: the American Museum of Natural History (AMNH), New York; the Natural History Museum (NHM), London; the Smithsonian Institution National Museum of Natural History (USNM), Washington; and the Royal Ontario Museum (ROM), Ontario. These specimens of modern and historical mammal samples were obtained during the past few hundred years, largely from hunted or trapped individuals. Many were sexed at the time of collection, or subsequently, based on preserved genitalia, or clearly distinguishing secondary sexual characters (such as antlers for most deer species). The ratio of males was calculated for each species represented by more than 100 individuals (Fig. 1). The male ratio, averaged across species, was greater than 1:1 in most mammalian orders, with notable exceptions for Chiroptera (bats) and Pilosa (sloths and anteaters). However, there was extreme variability across taxa, which may result from the method of collection (hunting vs. trapping),

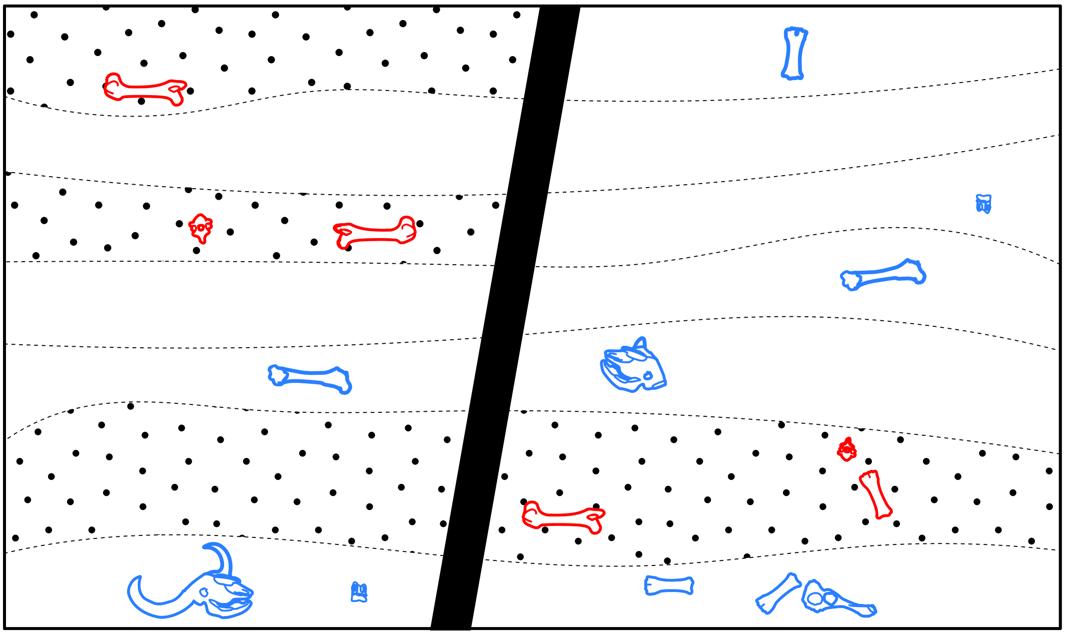


Figure 2: Hypothetical cross sections from two dig sites, with female bones in **red** and males in **blue**. Female home ranges (shown with polkadots) vary over time, and this is reflected vertically in the different sedimentary layers. For the site shown on the left, both male and female dominated zones encompass the site regularly, so a sampling episode here would yield equal proportions of males and females. On the right, female dominated zones have irregular/infrequent interaction with the site, so sampling here would make females appear rarer than males.

or the source of the samples (zoo vs. wild).

Discussion

A bias towards males appears to be a pervasive feature in both subfossil and live-collected mammal collections, and could be due to a range of plausible explanations. Perhaps the simplest explanation of the male sex bias in the subfossil datasets is a taphonomic artefact, where male bones in sexually dimorphic data such as bison are larger or denser and more likely to be better preserved or identified as likely to contain ancient DNA. If this was the case, male bias might be expected to correlate with factors associated with post-mortem DNA preservation, such as sample age, average DNA fragment length, and cytosine deamination rate. Greater bone density might also be expected to inhibit microbial intrusion, and thus increase the proportion of endogenous DNA (host species versus microbial DNA). No such trends were observed here

(Table 2), and thus it is reasonable to conclude that DNA preservation is equal between the sexes.

Given the evidence of equal post-mortem preservation, the observed male bias could relate to differences in either deposition rates or collection activities. Regarding the latter, we found no evidence of a decreased male sex bias in smaller skeletal elements where sexual dimorphism is less readily apparent, suggesting that size-biased collecting or sampling is unlikely to be a major driver of the observed sex ratios in bison or brown bears. Consequently, our data would appear to support a biased male deposition rate in both bison and brown bears, consistent with the landscape ranging hypothesis proposed for mammoth (Pečnerová *et al.*, 2017), where male deaths are more broadly distributed. This bias is expected to be particularly strong for female-herding taxa, where female ranges are potentially clustered geographically. While the latter will change distribution over time, random sampling across the landscape is still more likely to locate male remains (Fig. 2). A corollary of this model is that locations dominated by large female groups should be encountered occasionally, yielding female biased ratios for such sites. The only such site we observed was brown bears in the Alps (for which a behavioural explanation is available), however, there are very few sites for which we have multiple samples in our dataset. The female-biased sex ratios observed for bats may derive from collections dominated by sampling of single roosts, which at certain times of the year may be inhabited only by one sex, particularly maternity colonies (Kunz, 1982).

Cave sites appear to provide different sex biases from open alluvial systems, possibly related to behavioural traits such as the differential denning activities for bears in the Alps and elsewhere. For example, the dominance of male brown bears in cave sites outside the Alps may reflect the ability of males to drive off females from preferred denning locations such as caves. While the lone male model is consistent with the observed data, it technically only applies to herd animals and also probably can only be differentiated from the landscape ranging model, where males simply have bigger ranges, by determining the age at death for specimens. The lone male model predicts that the age at death will be younger for male than for females, due to lack of experience and herd protection. Age at death can be measured morphologically from tooth eruption and wear, and in mammoths by dating the enamel layers of tusks. However, large collections of subfossil teeth, preserving ancient DNA are unlikely to be readily available. Certain methylomic loci can be used to indicate age in humans (Horvath & Raj, 2018), so cytosine methylation in ancient DNA (Llamas *et al.*, 2012) could be used to age subfossil specimens. Currently, without detailed age of death data, it remains challenging to support the lone male model over the landscape ranging model.

Collection bias

Where we deliberately sampled thicker and larger *Myotragus balearicus* bones to maximise DNA preservation in a warm climate, all were found to be male ($n = 9$), indicating that this bias can potentially affect subfossil collections. It is highly likely that a similar collection bias affects modern mammalian collections arising from mostly hunted and trapped individuals. For modern mammals, this bias need not only be driven by deliberate selection of large 'impressive' male specimens, but could also be due to other factors such as hunters or trappers avoiding females tending young because of legislation or other motivation. At the same time, museum collections do not only represent the choices of collectors and hunters. Museum curators may act judiciously to select materials for accession with a goal of representing both sexes (as well as representing different localities, times, or ages) for species in their collections, a factor that may in fact counteract, to some extent, any tendency for extreme male bias in some collections. Whatever the cause, the pervasiveness of male over-representation in mammal collections requires attention. The use of museum specimens as the major platform for comparative anatomy, morphological variability, ontogenetic development, parasitology, stable isotope chemistry, stomach contents, and many other aspects of biology in mammalian species (McLean *et al.*, 2016) raises the question of the extent that previous studies may be impacted by an undetected male bias.

We have not examined the extent of male bias in modern bird collections, but suspect that the remarkable sexual dimorphism in colour in many bird species may lead to similar male bias, as males typically exhibit more visually striking plumage. However, data available for the extinct moas of New Zealand suggest a different pattern for ratite birds, where sex roles are reversed. Moa exhibit pronounced reverse sexual size dimorphism, with females two or more times heavier than males (Bunce *et al.*, 2003). Fossil remains of four different moa species show heavily female-dominated sex ratios across two different deposits, with suggestions that female territoriality led to an increased death rate near watering holes (Allentoft *et al.*, 2010). Importantly, this provides a further indication that differential sexual morphology and behavioural ecology of large vertebrates, rather than sex *per se*, may be important drivers of sex ratios observed in the fossil record.

Conclusion

We observed a substantial excess of male bison and brown bear subfossils across a range of Late Pleistocene Holarctic deposits, consistent with a landscape ranging hypothesis. The female-herd structure of bison, like mammoths, explains the high ratio of male subfossils as females are expected to be clustered geographically, and therefore more heterogeneous on the landscape. In the case of brown bears, the lack of a herd structure leads to a more equal distribution of subfossil remains in open sites but a pronounced male sex bias in cave sites, which may reflect preferred denning sites. Within caves in the European Alps a reversed situation is observed, potentially due to a lack of male hibernation in caves.

Regardless of the actual mechanisms, a substantial male sex bias exists in both the subfossil record and modern mammalian collections. The biases are highly taxon specific, and are likely to differ between collections. This has implications for studies that assume their samples are representative for the whole of the population under consideration, such as comparisons of taxa or studies of factors such as bone dietary isotopes where sexes differ in their behaviour or distribution. Our results suggest that sex biases are ubiquitous in collections, and should not be ignored. The routine application of genetic sexing will allow the possible confounding effects of cryptic sexual dimorphism to be identified when working with subfossils or museum collections.

Materials and Methods

Laboratory procedures

All ancient DNA work was performed in the purpose-built isolated ancient DNA facility at the University of Adelaides Australian Centre for Ancient DNA following previously published guidelines (Cooper & Poinar, 2000; Shapiro & Hofreiter, 2012). DNA was extracted from bison samples using either a phenol-chloroform or in house silica based method as described in (Soubrier *et al.*, 2016). Brown bear samples were extracted using a phenol-chloroform based extraction protocol (Bray *et al.*, 2013) or an in-house silica-based protocol (Dabney *et al.*, 2013). Double-stranded Illumina sequencing libraries were built from 25 μL of DNA extract following the partial uracil-DNA-glycosylase (UDG) treatment protocol (Rohland *et al.*, 2015), modified to include the use of dual 7-mer internal barcode sequences as per (Soubrier *et al.*, 2016). The libraries were pooled and sequenced using paired-end reactions on an Illumina MiSeq, NextSeq, or HiSeq.

Alignment and filtering

Demultiplexed reads were mapped using the Paleomix pipeline (Schubert *et al.*, 2014) configured to use BWA-aln (Li & Durbin, 2009) with typical ancient DNA parameters (`-l 16384 -o 2 -n 0.01`). Alignments were subsequently filtered to exclude those with mapping quality lower than 30, and fragments longer than 100 bp. We considered only samples with at least 5000 reads mapped to the nuclear genome, and subsampled down to approximately 20000 reads for sex determination.

Bison

Bison reads were mapped to a composite cattle reference assembly formed by concatenating the assembly UMD3.1 (Zimin *et al.*, 2009), with the Y chromosomal sequence from Btau4.6.1 (Elsik *et al.*, 2009). As very few reads map to this Y sequence, we were unable to do genetic sexing using counts of reads mapping to the Y chromosome vs. counts of those mapping to the X chromosome as in (Skoglund *et al.*, 2013). We instead counted reads mapping to the X chromosome vs. the autosome, in an approach similar to (Mittnik *et al.*, 2016).

We counted the reads that mapped to the X chromosome, N_{X}, and the reads that mapped to the autosome, N_{A}, using `samtools idxstats` (Li, 2011). Assuming reads are drawn from the genome uniformly along its length, the observed ratio $R_{\text{X}} = N_{\text{X}}/(N_{\text{X}} + N_{\text{A}})$ can be predicted from the length of the X chromosome, L_{X}, and the length of the autosome, L_{A}. Conditional on the sex, the expected ratios are,

$$\begin{aligned} p_{\text{XY}} &= \mathbb{E}\left[R_{\text{X}} \mid \text{sex} = \text{XY}\right] = L_{\text{X}}/(L_{\text{X}} + 2\,L_{\text{A}}) \quad \text{or} \\ p_{\text{XX}} &= \mathbb{E}\left[R_{\text{X}} \mid \text{sex} = \text{XX}\right] = L_{\text{X}}/(L_{\text{X}} + L_{\text{A}}). \end{aligned}$$

The likelihood of the male ratio p_{XY} given the observed counts N_{X} and N_{A} can thus be described using the Binomial probability mass function,

$$\mathcal{L}(p_{\text{XY}} \mid N_{\text{X}}, N_{\text{A}}) = \frac{(N_{\text{X}} + N_{\text{A}})!}{N_{\text{X}}!\,N_{\text{A}}!}\, p_{\text{XY}}^{N_{\text{X}}}\,(1 - p_{\text{XY}})^{N_{\text{A}}},$$

and similarly for the female ratio. We determined if one sex fit the data best using a likelihood ratio test (LRT), requiring that the LRT result in a p-value < 0.001 for one or the other sex, in order that a sex be assigned. Further, we considered

$$M_{\text{X}} = \begin{cases} 0.5\,R_{\text{X}}/p_{\text{XY}} & \text{for males;} \\ 1.0\,R_{\text{X}}/p_{\text{XX}} & \text{for females;} \end{cases}$$

depending on the result of the LRT, to cluster males near 0.5 and females near 1.0. We did not assign a sex to samples that had $0.6 < M_{\text{X}} < 0.8$, under the assumption that they violated both male and female models. Our Python code implementation for the sex assignment is available from `https://github.com/grahamgower/sexassign`.

Mammoths

Mammoth sexing was done using the same method as for bison. Read counts N_{X} and N_{A} were taken from Supplementary Table 1 of (Pečnerová *et al.*, 2017), which also lists material type and ^{14}C age for each sample. L_{X} and L_{A} were derived from the African elephant reference loxAfr4. A total of 398 360 mapped reads were reported for sample L285, which is likely missing a digit. We appended a zero, placing this sample into the male range, which matches the inferred sex from (Pečnerová *et al.*, 2017).

Bears

Brown bear reads were mapped to the polar bear reference UrsMar1.0 (Liu *et al.*, 2014), a scaffold-level reference assembly. For sex determination, we counted reads that mapped to X-linked scaffolds as N_{X}, and applied the same method as for Bison. Only scaffolds longer than 1 Mbp were used in calculations of N_{X}, N_{A}, L_{X}, L_{A}.

A list of UrsMar1.0 X-linked scaffolds (Table S1) was obtained by mapping all UrsMar1.0 scaffolds to the dog reference CanFam3.1 (Lindblad-Toh *et al.*, 2005), with minimap2 (Li, 2018). The default mapping parameters were used (`minimap2 CanFam3.1.fasta UrsMar1.0.fasta > aln.paf`), which provides an approximate alignment lacking base-level precision. We retained only UrsMar1.0 scaffolds having more than 100 kbp cumulative ‘approximate’ matches to the CanFam3.1 chrX, resulting in 28 putatively X-linked scaffolds comprising 102 Mbp of sequence.

Model violations

While care was taken to minimise contamination from exogenous sources, such model violations may yet occur due to sample cross-contamination. Other factors that may contribute to sample specific model violations include chromosome translocations, aneuploidy, and unanticipated post-mortem preservation artifacts that (dis)favour one chromosome over another.

Systematic model violations may also be present, such as due to reference assembly errors, or post-mortem preservation artifacts. Inactivated copies of chromosome X are heavily methylated, which may lead to additional post-mortem DNA fragmentation compared to the active copy and hence fewer reads mapping from the inactivated chromosome. Conversely an inactivated chromosome is condensed into heterochromatin, which may facilitate greater post-mortem preservation than the active copy.

We note that the UrsMar1.0 assembly was derived by sequencing a male, and thus Y-linked scaffolds may be present, while the CanFam3.1 assembly was derived by sequencing a female and thus lacks a chrY. This leaves open the possibility that the pseudoautosomal region (PAR) on Y-linked UrsMar1.0 scaffolds could have mapped to CanFam1.0 chrX. The dog PAR region is ˜6.6 Mbp (Young *et al.*, 2008), small compared with the size of chrX, but this could yet artificially inflate R_X values for males.

Nonetheless we observed a clear separation of R_X values into two cohorts, with few intermediate values, suggesting model violations are rare, or do not notably influence sex determination.

GLM

Logistic regression models were implemented in R (R Core Team, 2017) using the `bayesglm` function with default parameters, from the `arm` package (Gelman *et al.*, 2008). For categorical variables with three or more levels, we constructed multiple models, each with different reference levels, to verify this did not have a notable influence on the outcome.

Testing spatial distribution

We implemented the two-sample kernel test described by (Gretton *et al.*, 2012) with a Gaussian kernel, and obtained a p-value by comparing the test statistic to 1000 permutations. The Gaussian kernel $k(x, y) = \exp\left(-\left(d(x, y)/\sigma\right)^2\right)$, where $d(x, y)$ is the great circle distance between x and y, has a scaling parameter σ, which was chosen to maximise the test statistic in each permutation. More details regarding the test statistic, and validation of its performance for spatial data, can be found in the Supplementary Information. Our R code implementation for the kernel test is available from `https://github.com/grahamgower/kernel-test`.

Mammalian databases

For mammalian species listed in the PanTHERIA WR05 database (Jones *et al.*, 2009), we downloaded sample information from three museum databases: the American Museum of Natural History (AMNH) (American Museum of Natural

History, 2018); the Natural History Museum, London (NHM) (London Natural History Museum, 2014); and the Royal Ontario Museum (ROM) v11.5 (Royal Ontario Museum, 2018). In addition, samples for 38 species were manually downloaded from the Smithsonian Institution National Museum of Natural History (USNM) (Smithsonian National Museum of Natural History, 2018). We excluded juveniles and hybrids, and sex ratios were calculated only for species represented by more that 100 sexed samples.

Author contributions

GG performed sex determination and statistical analyses. LF, AS, ALvL, and HH processed samples. GG, KMH, KJM, BS, and AC interpreted the results. GG, KMH, and AC wrote the manuscript with input from all coauthors. GG wrote the supplementary material.

Acknowledgements

We thank Pere Bover for providing information on *Myotragus balearicus*. This work was supported by Australian Government Research Training Program Scholarships [GG,LF,AS,ALvL], research fellowships from the University of Adelaide and the Australian Research Council (ARC) [BL], and Linkage Program funding from the ARC.

References

Agenbroad L & Mead J (1987). Age structure analyses of *mammuthus columbi*, Hot Springs Mammoth Site, South Dakota. *Current Research in Pleistocene*, **4**:101–102

Allentoft ME, Bunce M, Scofield RP, Hale ML, & Holdaway RN (2010). Highly skewed sex ratios and biased fossil deposition of moa: ancient DNA provides new insight on New Zealand's extinct megafauna. *Quat Sci Rev*, **29(5-6)**:753–762. http://dx.doi.org/10.1016/j.quascirev.2009.11.022

American Museum of Natural History (2018). AMNH Vertebrate Zoology Database. http://sci-web-001.amnh.org/db/emuwebamnh/index.php. [Online; accessed 29-May-2018]

Bray SCE, Austin JJ, Metcalf JL, Østbye K, Østbye E, Lauritzen SE, Aaris-Sørensen K, Valdiosera C, Adler CJ, & Cooper A (2013). Ancient DNA identifies post-glacial recolonisation, not recent bottlenecks, as the primary driver of contemporary mtDNA phylogeography and diversity in Scandinavian brown bears. *Divers Distrib*, **19(3)**:245–256. http://dx.doi.org/10.1111/j.1472-4642.2012.00923.x

Bunce M, Worthy TH, Ford T, Hoppitt W, Willerslev E, Drummond A, & Cooper A (2003). Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature*, **425(6954)**:172–175. http://dx.doi.org/10.1038/nature01871

Bunnell FL & Tait DEN (1981). Population dynamics of bears – implications. In CW Fowler & TD Smith, eds., *Dynamics of large mammal populations*, pp. 75–98. John Wiley and Sons, Inc.

Charnov EL (1975). Sex ratio selection in an age-structured population. *Evolution*, **29(2)**:366–368. http://dx.doi.org/10.1111/j.1558-5646.1975.tb00216.x

Cooper A & Poinar HN (2000). Ancient DNA: Do it right or not at all. *Science*, **289(5482)**:1139–1139. http://dx.doi.org/10.1126/science.289.5482.1139b

Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, Nickel B, Valdiosera C, García N, Pääbo S, Arsuaga JL, *et al.* (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci U S A*, **110(39)**:15758–15763. http://dx.doi.org/10.1073/pnas.1314445110

Döppes D & Pacher M (2014). 10,000 years of *Ursus arctos* in the Alps – a success story? Analyses of the late glacial and early holocene brown bear remains from alpine caves in Austria. *Quat Int*, **339-340**:266–274. http://dx.doi.org/10.1016/j.quaint.2013.11.039

Elsik CG, Tellam RL, & Worley KC (2009). The genome sequence of taurine cattle: A window to ruminant biology and evolution. *Science*, **324(5926)**:522–528. http://dx.doi.org/10.1126/science.1169588

Frayer DW & Wolpoff MH (1985). Sexual dimorphism. *Annual Review of Anthropology*, **14(1)**:429–473. http://dx.doi.org/10.1146/annurev.an.14.100185.002241

Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, Furtwängler A, Haak W, Meyer M, Mittnik A, *et al.* (2016). The genetic history of Ice

Age Europe. *Nature*, **534(7606)**:200–205. http://dx.doi.org/10.1038/nature17993

Gelman A, Jakulin A, Pittau MG, & Su YS (2008). A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat*, **2(4)**:1360–1383. http://dx.doi.org/10.1214/08-AOAS191

Gretton A, Borgwardt KM, Rasch MJ, Schölkopf B, & Smola A (2012). A kernel two-sample test. *J Mach Learn Res*, **13**:723–773

Guthrie RD (1989). *Frozen Fauna of the Mammoth Steppe: The Story of Blue Babe*. University of Chicago Press, Chicago. ISBN 978-0-226-31123-4

Haynes G (2017). Finding meaning in mammoth age profiles. *Quat Int*, **443**:65–78. http://dx.doi.org/10.1016/j.quaint.2016.04.012

Horvath S & Raj K (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet*, **19(6)**:371–384. http://dx.doi.org/10.1038/s41576-018-0004-3

Jones KE, Bielby J, Cardillo M, Fritz SA, O'Dell J, Orme CDL, Safi K, Sechrest W, Boakes EH, Carbone C, *et al.* (2009). PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, **90(9)**:2648–2648. http://dx.doi.org/10.1890/08-1494.1

Karlin S & Lessard S (1986). *Theoretical Studies on Sex Ratio Evolution*. Princeton University Press, Princeton, New Jersey

Kunz TH, ed. (1982). *Ecology of Bats*. Plenum Press, New York

Li H (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27(21)**:2987–2993. http://dx.doi.org/10.1093/bioinformatics/btr509

Li H (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. http://dx.doi.org/10.1093/bioinformatics/bty191

Li H & Durbin R (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, **25(14)**:1754–1760. http://dx.doi.org/10.1093/bioinformatics/btp324

Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Iii EJK, Zody MC, *et al.* (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, **438(7069)**:803. http://dx.doi.org/10.1038/nature04338

Liu S, Lorenzen ED, Fumagalli M, Li B, Harris K, Xiong Z, Zhou L, Korneliussen TS, Somel M, Babbitt C, *et al.* (2014). Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell*, **157(4)**:785–794. http://dx.doi.org/10.1016/j.cell.2014.03.054

Llamas B, Holland ML, Chen K, Cropley JE, Cooper A, & Suter CM (2012). High-resolution analysis of cytosine methylation in ancient DNA. *PLoS One*, **7(1)**. http://dx.doi.org/10.1371/journal.pone.0030226

London Natural History Museum (2014). Dataset: Collection specimens. resource: Specimens. Natural History Museum Data Portal. http://data.nhm.ac.uk. http://dx.doi.org/10.5519/0002965. [Online; accessed 29-May-2018]

McLean BS, Bell KC, Dunnum JL, Abrahamson B, Colella JP, Deardorff ER, Weber JA, Jones AK, Salazar-Miralles F, & Cook JA (2016). Natural history collections-based research: progress, promise, and best practices. *J Mammal*, **97(1)**:287–297. http://dx.doi.org/10.1093/jmammal/gyv178

Mittnik A, Wang CC, Svoboda J, & Krause J (2016). A molecular approach to the sexing of the triple burial at the upper paleolithic site of Dolní Věstonice. *PLoS ONE*, **11(10)**:e0163019. http://dx.doi.org/10.1371/journal.pone.0163019

Pečnerová P, Díez-del Molino D, Dussex N, Feuerborn T, von Seth J, van der Plicht J, Nikolskiy P, Tikhonov A, Vartanyan S, & Dalén L (2017). Genome-based sexing provides clues about behavior and social structure in the woolly mammoth. *Curr Biol*, **27(22)**:3505–3510.e3. http://dx.doi.org/10.1016/j.cub.2017.09.064

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria

Rehg JA & Leigh SR (1999). Estimating sexual dimorphism and size differences in the fossil record: A test of methods. *Am J Phys Anthropol*, **110(1)**:95–104. http://dx.doi.org/10.1002/(SICI)1096-8644(199909)110:1<95::AID-AJPA8>3.0.CO;2-J

Rohland N, Harney E, Mallick S, Nordenfelt S, & Reich D (2015). Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philos Trans R Soc Lond B Biol Sci*, **370(1660)**:20130624. http://dx.doi.org/10.1098/rstb.2013.0624

Royal Ontario Museum (2018). Mammalogy Collection - Royal Ontario Museum. http://gbif.rom.on.ca/ipt/resource.do?r=mamm. [Online; accessed 29-May-2018]

Schubert M, Ermini L, Sarkissian CD, Jónsson H, Ginolhac A, Schaefer R, Martin MD, Fernández R, Kircher M, McCue M, *et al.* (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat Protoc*, **9(5)**:1056–1082. http://dx.doi.org/10.1038/nprot.2014.063

Shapiro B & Hofreiter M, eds. (2012). *Ancient DNA: Methods and Protocols*. Methods in Molecular Biology. Humana Press. ISBN 978-1-61779-515-2

Skoglund P, Storå J, Götherström A, & Jakobsson M (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J Archaeol Sci*, **40(12)**:4477–4482. http://dx.doi.org/10.1016/j.jas.2013.07.004

Smithsonian National Museum of Natural History (2018). Mammals Collections Search. https://collections.nmnh.si.edu/search/mammals/. [Online; accessed 30-May-2018]

Soubrier J, Gower G, Chen K, Richards SM, Llamas B, Mitchell KJ, Ho SYW, Kosintsev P, Lee MSY, Baryshnikov G, *et al.* (2016). Early cave art and ancient DNA record the origin of European bison. *Nat Commun*, **7**:13158. http://dx.doi.org/10.1038/ncomms13158

Støen OG, Zedrosser A, Sæbø S, & Swenson JE (2006). Inversely density-dependent natal dispersal in brown bears *ursus arctos*. *Oecologia*, **148(2)**:356. http://dx.doi.org/10.1007/s00442-006-0384-5

Trivers RL & Willard DE (1973). Natural selection of parental ability to vary the sex ratio of offspring. *Science*, **179(4068)**:90–92. http://dx.doi.org/10.1126/science.179.4068.90

Young AC, Kirkness EF, & Breen M (2008). Tackling the characterization of canine chromosomal breakpoints with an integrated in-situ/in-silico approach: The canine PAR and PAB. *Chromosome Res*, **16(8)**:1193–1202. http://dx.doi.org/10.1007/s10577-008-1268-9

Zedrosser A, Støen OG, Sæbø S, & Swenson JE (2007). Should I stay or should I go? Natal dispersal in the brown bear. *Animal Behaviour*, **74(3)**:369–376. http://dx.doi.org/10.1016/j.anbehav.2006.09.015

Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS, *et al.* (2009). A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol*, **10(4)**:R42. http://dx.doi.org/10.1186/gb-2009-10-4-r42

4.3 Supplementary Information

Table S1: Putatively X-linked scaffolds of the polar bear, identified by mapping the `UrsMar1.0` reference to the dog reference `CanFam3.1`.

KK498507.1
KK498524.1
KK498558.1
KK498591.1
KK498592.1
KK498613.1
KK498620.1
KK498621.1
KK498625.1
KK498626.1
KK498633.1
KK498654.1
KK498655.1
KK498666.1
KK498668.1
KK498669.1
KK498670.1
KK498681.1
KK498702.1
KK498740.1
KK498766.1
KK498779.1
KK498782.1
KK498829.1
KK498842.1
KK499341.1
KK499355.1
KK499613.1

Testing for differences in spatial distribution

In mammoths and bison, large groups are comprised predominantly of mature females and sub-adults (of both sexes), while most mature males are excluded from the group by an oligarchy. Excluded males are solitary, or form minor groups, and may inhabit more marginal locations compared to those inhabited by the larger groups. Pečnerová *et al.* (2017) hypothesised that an excess of male samples is observed for mammoths (and will be observed for bison), because their social structure gives rise to differences in the modes of death for males and females. This implies either taphonomic differences between sexes due to differing habitats, or simply differences in their spatial extent. In either case, inter-sample distances within sexes should be smaller than for the population as a whole, and this should be discernible from the fossil record.

To test for differences in spatial distribution between males and females, it is possible to apply univariate tests, separately to latitude and longitude. However, the Kolmogorov-Smirnov test, Cramér-von Mises test, and similar two-sample univariate tests are known to be conservative tests that perform poorly on many datasets. In addition, a univariate test may not reveal differences that arise only when jointly considering latitude and longitude. We note also that the spatial distributions for both bison and brown bear samples are multimodal, with population centres in Europe and North America. Thus we sought a mulitivariate two-sample test, which is adequate for testing non-symmetric multimodal distributions. The kernel two-sample test described in Gretton *et al.* (2012) can be readily applied to high dimensional data, and has few assumptions on the data itself.

We implemented the kernel test in R (R Core Team, 2017), and our code is available from `https://github.com/grahamgower/kernel-test`. The test statistic that we used is

$$T(\mathbf{X}, \mathbf{Y}) = \frac{1}{m^2} \sum_{i,j=1}^{m} k(x_i, x_j) - \frac{2}{mn} \sum_{i,j=1}^{m,n} k(x_i, y_j) + \frac{1}{n^2} \sum_{i,j=1}^{n} k(y_i, y_j).$$

Which can be interpreted as the distance between the two probability distributions that produced samples $\mathbf{X}$ and $\mathbf{Y}$. Where $\mathbf{X} = x_1, \ldots, x_m$, and $\mathbf{Y} = y_1, \ldots, y_n$, are the coordinates of male and female samples, respectively (e.g. x_i is the latitude and longitude for the ith male sample). The kernel function $k(u, v)$ measures similarity of two individuals u and v, which for our purposes is 1 for two samples with identical locations, and decreases towards 0 as their distance increases. The kernel function must be *positive definite* in order that it embed the sample distances into an Hilbert space. Both Gaussian and Laplacian kernel functions are known to be appropriate choices, and we evaluated both. For the Gaussian kernel, this is

$$k_G(u, v) = \exp\left(-\left(d(u, v)/\sigma\right)^2\right),$$

and for the Laplacian kernel,

$$k_L(u, v) = \exp\left(-d(u, v)/\sigma\right).$$

Where $d(u, v)$ is the distance between individuals u and v. Both kernels have a scaling parameter σ (known as the bandwidth), which we chose to maximise the test statistic, as inspired by: `https://normaldeviate.wordpress.com/2012/07/14/modern-two-sample-tests/`. In each case, the test statistic has a single maxima with respect to σ, and maximisation was done using the `optimise` function in R (R Core Team, 2017).

We evaluated two metrics, the Euclidean distance, $d_E(u, v)$, and the great-circle distance (as traveled on the surface of a sphere), $d_{gc}(u, v)$. If lat_u and lon_u are the latitude and longitude for sample u, then

$$\begin{aligned} d_E(u, v) &= \sqrt{(\text{lat}_u - \text{lat}_v)^2 + (\text{lon}_u - \text{lon}_v)^2}, \\ d_{gc}(u, v) &= \cos^{-1}\left(\sin\left(\text{lat}_u\right)\sin(\text{lat}_v) + \cos(\text{lat}_u)\cos(\text{lat}_v)\cos(\text{lon}_u - \text{lon}_v)\right). \end{aligned}$$

The significance of the test statistic, $T(\mathbf{X}, \mathbf{Y})$, was evaluated with a permutation test. I.e. the male/female labels were randomly reassigned to new individuals, keeping the total number of males and females the same, then the test statistic was recomputed. A null distribution was obtained by repeating this procedure many times. In each permutation, the scaling parameter σ was reestimated.

To determine how well the kernel test performs compared to various other two-sample tests, we simulated spatial distributions with two population centres by drawing random latitudes and longitudes from a mixture of two multivariate normal distributions. The sample counts, means, and covariance matrices, were taken from the data observed for European and American bison, and we did 1000 simulations under each of two distinct scenarios: (1) male and female locations were drawn from the same distribution (same mean and covariance

Table S2: Proportion of simulations in which a two-sample test rejected the null hypothesis, from 1000 simulations, at a specified false positive rate $\alpha = 0.05$. The configuration highlighted in bold was used for the results reported in the main text.

Test	Type 1	Power
Kolmogorov-Smirnov (lat)	0.040	0.063
Kolmogorov-Smirnov (lon)	0.059	0.109
Cramér-von Mises (lat)	0.042	0.056
Cramér-von Mises (lon)	0.060	0.071
Cramér test	0.050	0.061
Energy distance (d_E)	0.049	0.058
Energy distance (d_{gc})	0.045	0.112
kernel test ($k_G, d_E, \sigma = median$)	0.043	0.135
kernel test ($k_G, d_{gc}, \sigma = median$)	0.043	0.170
kernel test ($k_L, d_E, \sigma = median$)	0.044	0.186
kernel test ($k_L, d_{gc}, \sigma = median$)	0.042	0.335
kernel test (k_G, d_E)	0.038	0.661
kernel test (k_G, d_{gc})	**0.034**	**0.956**
kernel test (k_L, d_E)	0.038	0.678
kernel test (k_L, d_{gc})	0.035	0.947

matrices); and (2) male and female locations were drawn from different distributions (same mean, but different covariance matrices). Using a prespecified false positive rate $\alpha = 0.05$, the actual false positive rate was estimated by counting how often a test rejected the null for scenario (1), and the relative power of the tests was established by identifying how often the null was rejected under scenario (2) (see **Table S2**).

The Energy distance (Szekely & Rizzo, 2004) test was performed using the `eqdist.test` function from the `energy` R package, and the Cramér test (distinct from the Cramér-von Mises test) was calculated with the `cramer.test` from the `cramer` R package (Baringhaus & Franz, 2004). The Kolmogorov-Smirnov test used the `ks.test` function from base R (R Core Team, 2017). The Cramér-von Mises test was implemented in R following the description from Anderson (1962). Except the Kolmogorov-Smirnov test, all tests use permutations to obtain a p-value, and are expected to have false positive rates close to the prespecified value. We note that the `kernlab` R package (Karatzoglou *et al.*, 2004) also implements a two-sample kernel test, but we were unable to obtain reliable results for our test cases.

The results of **Table S2** suggest that the power of the kernel test, for this data, is sensitive to the choice of metric, and the kernel bandwidth, but not the kernel function (Laplacian or Gaussian). Using a bandwidth based on the median distance between individuals is much faster than maximising the test statistic. It is plausible that a fixed bandwidth could be chosen which attains similar power to a variable bandwidth, although it is not clear how this might be chosen in a way that performs well for different types of data. We note that both the Energy distance and Cramér tests have the same form as the test statistic we used for the kernel test (Sejdinovic *et al.*, 2012), and it might to be possible to improve their power by transforming the pairwise distance matrix used for these tests.

In the main text, we report kernel test results for a Gaussian kernel, great-circle distance, and variable bandwidth chosen by maximising the test statistic. This configuration is shown to be a good choice compared to several alternatives. Many other choices of kernel function are possible, and our investigation was far from exhaustive. We also chose only one specific scenario with which to evaluate the relative power of the two-sample tests. However, the scenario was chosen by mildly perturbing our empirical dataset, in order to make the evaluation as realistic as possible.

References

Anderson TW (1962). On the distribution of the two-sample Cramer-von Mises criterion. *Ann Math Statist*, **33(3)**:1148–1159. http://dx.doi.org/10.1214/aoms/1177704477

Baringhaus L & Franz C (2004). On a new multivariate two-sample test. *J Multivar Anal*, **88(1)**:190–206. http://dx.doi.org/10.1016/S0047-259X(03)00079-4

Gretton A, Borgwardt KM, Rasch MJ, Schölkopf B, & Smola A (2012). A kernel two-sample test. *J Mach Learn Res*, **13**:723–773

Karatzoglou A, Smola A, Hornik K, & Zeileis A (2004). kernlab – an S4 package for kernel methods in R. *J Stat Softw*, **11(9)**:1–20. http://dx.doi.org/10.18637/jss.v011.i09

Pečnerová P, Díez-del Molino D, Dussex N, Feuerborn T, von Seth J, van der Plicht J, Nikolskiy P, Tikhonov A, Vartanyan S, & Dalén L (2017). Genome-based sexing provides clues about behavior and social structure in the woolly mammoth. *Curr Biol*, **27(22)**:3505–3510.e3. http://dx.doi.org/10.1016/j.cub.2017.09.064

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria

Sejdinovic D, Gretton A, Sriperumbudur B, & Fukumizu K (2012). Hypothesis testing using pairwise distances and associated kernels. *arXiv:1205.0411 [cs, stat]*. ArXiv: 1205.0411

Szekely GJ & Rizzo ML (2004). Testing for equal distributions in high dimensions. *InterStat*, **5(Nov)**

Chapter 5

PP5mC: preprocessing hairpin-ligated bisulfite-treated DNA sequences

5.1 Authorship statement

Statement of Authorship

Title of Paper	PP5mC: preprocessing hairpin-ligated bisulfite-treated DNA sequences
Publication Status	Unpublished and unsubmitted work written in manuscript style

Principal Author

Name of Principal Author (Candidate)	Graham Gower		
Contribution to the Paper	Designed and implemented the software; performed simulations; interpreted results; wrote the manuscript.		
Overall percentage (%)	100		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	19/10/2018

5.2 Manuscript

PP5mC: preprocessing hairpin-ligated bisulfite-treated DNA sequences

Graham Gower

Australian Centre for Ancient DNA, School of Biological Sciences, The Environment Institute, The University of Adelaide, Adelaide, 5005, Australia.

Abstract

Ligation of a hairpin adapter onto double stranded DNA enables simultaneous sequencing of the top and bottom strands. This is a useful precursor step to bisulfite sequencing as the original four nucleotide states can be recovered prior to mapping, in addition to the cytosine methylation status. PP5mC is a collection of tools written in C for processing hairpin-ligated bisulfite-treated sequencing data prior to analysis. We provide shell scripts implementing a basic pipeline for use on Unix workstations, and a more sophisticated pipeline written in Python for compute clusters using the Slurm job manager. Pipeline stages include: reconstruction of original nucleotide sequences from paired-end reads (`foldreads`); alignment to a reference, PCR deduplication, and indel realignment; recording nucleotide pairing statistics for positions upstream, within, and downstream of aligned reads (`scanbp`); and counting methylated/unmethylated cytosines at all CpG, CHG, and CHH contexts covered by alignments (`mark5mC`). PP5mC also includes a read simulator (`simhbs`) for hairpin and regular non-hairpin bisulfite sequencing, which we use to show that `foldreads` reconstructs sequences with greater accuracy, and is an order of magnitude faster, than HBS-tools, the only comparable software.

Availability: https://github.com/grahamgower/PP5mC

Introduction

DNA methylation plays an important role in the regulation of gene expression in eukaryotes, particularly with regard to transposon silencing, cell differentiation, and stress response pathways (Jeon *et al.*, 2015; Edwards *et al.*, 2017; Zhang *et al.*, 2018). To profile the genome-wide occurrence of 5-methylcytosines (5mC), with base-level precision, sequencing libraries may be treated using sodium bisulfite prior to amplification and sequencing (Urich *et al.*, 2015). Bisulfite treatment converts unmethylated cytosines into uracils, which are converted into thymines during subsequent PCR (C→T); but cytosines methylated at their $5'$ position are protected from conversion, and the cytosine is retained by PCR (5mC→C). Once sequenced, data are typically aligned to a reduced complexity reference, with cytosines translated into thymines, whereupon the methylation status is determined by considering the base composition of alignments in conjunction with the untranslated reference sequence (Krueger *et al.*, 2012). Compared with data from ordinary DNA sequencing experiments, bisulfite-treated sequencing data has lower information content, which decreases the proportion of reads that can be mapped to a unique reference position (Krueger *et al.*, 2012).

Laird *et al.* (2004) introduced the use of a hairpin adapter for bisulfite sequencing projects, which enables the sequencing of both top and bottom strands of the DNA molecule simultaneously. By computationally folding the sequence back together at the hairpin, the original molecule may be reconstructed for use with regular mapping software, while also recording the methylation status. This approach was recently modified for use with high- throughput sequencing (Zhao *et al.*, 2014), where paired-end libraries are produced by ligating an Illumina Y-adapter on one end of the molecule, and ligating a hairpin adapter on the other end (**Figure 1**). As the ligation process can result in molecules with Y-adapters on both ends, Zhao *et al.* (2014) used a hairpin containing a biotinylated thymine to enrich the library for hairpin-containing molecules. This depletes the concentration of molecules without hairpins, but does not remove them altogether, likely because sequence similarity between molecules can cause daisy chaining. Molecules with hairpins on both ends cannot be amplified, and are lost during subsequent PCR.

To date, hairpin bisulfite sequencing (HBS-seq) has not been widely used, possibly because regular bisulfite sequencing (BS-seq) provides adequate results for many experiments, but HBS-seq is now being considered for single-cell epigenomics (Kelsey *et al.*, 2017) due to its fidelity in assessing hemimethylation (Xu & Corces, 2018). Another potential application of HBS-seq is ancient DNA, for which BS-seq has already been used to show that 5mC signals can be recovered post-mortem with base-level precision (Llamas *et al.*, 2012). However, for source DNA molecules that are short (e.g. < 50 bp), such as those obtained from subfossil remains, reads are already challenging to map (uniquely) to the genome (Li & Freudenberg, 2014; Prüfer *et al.*, 2010). Reducing the sequence complexity with bisulfite treatment compounds this problem, making the approach unappealing for all but the best-preserved samples. HBS-seq provides a distinct advantage in this respect, as it permits alignment using all four nucleotide states, possibly extending epigenetic analyses to a

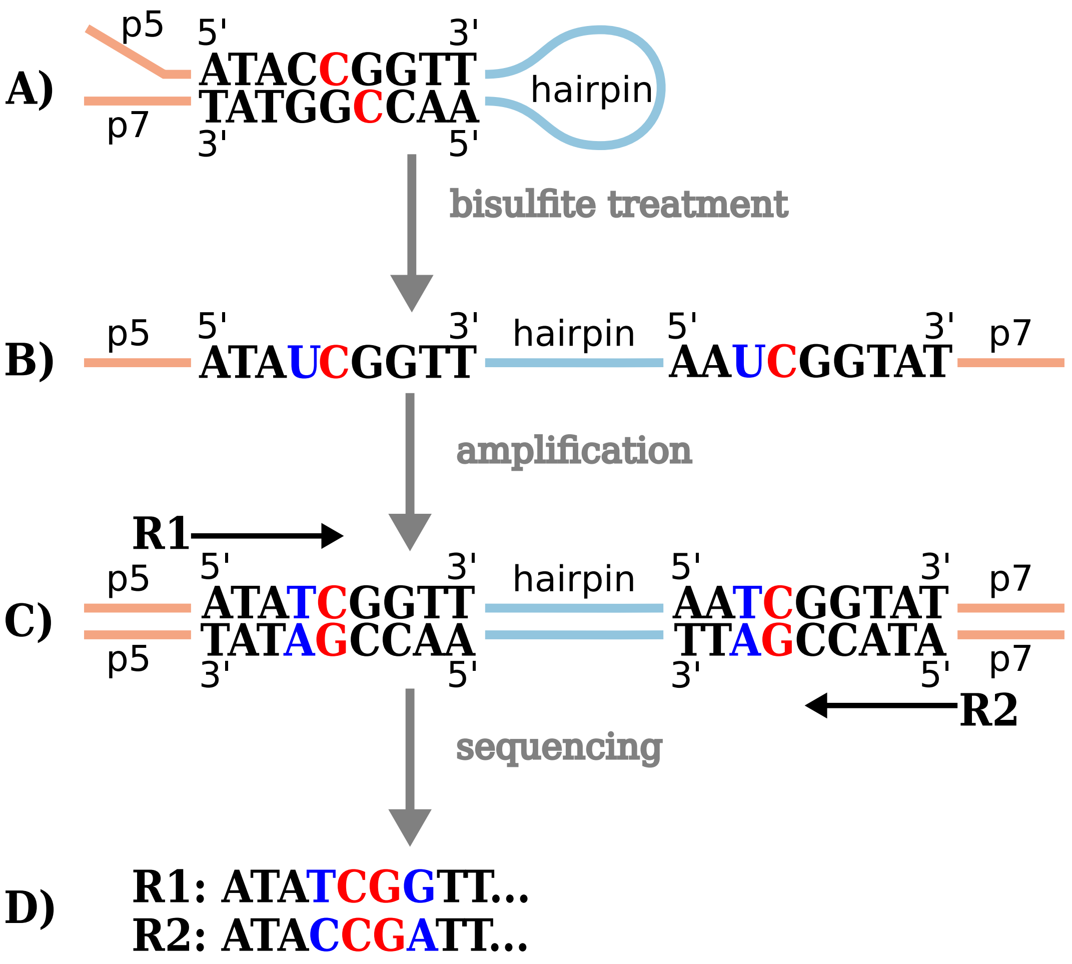


Figure 1: Flow chart of hairpin bisulfite sequencing protocol. **A)** Hairpin and Illumina Y-adapter are ligated to target molecule. Methylated cytosines are shown in **red**. **B)** Bisulfite treatment denatures the double stranded molecule and converts unmethylated cytosines to uracils (shown in **blue**). The connection between top and bottom strand is ensured by the hairpin. **C)** Library molecules are amplified by PCR for sequencing, with the polymerase incorporating thymines instead of uracils. R1 is the top strand sequence (left-to-right) immediately to the right of the p5 sequencing primer, whereas R2 is the bottom strand sequence (right-to-left) immediately to the left of the p7 sequencing primer. **D)** Paired-end sequencing produces R1 and R2 that have sequence identity, up to bisulfite conversion differences. As a visual aid, positions corresponding to cytosines in the original molecule are also coloured in (**C**) and (**D**).

wide range of subfossil and museum samples.

The use of forked Y-adapters, in conjunction with ancient DNA, has been reported to result in interrupted palindrome sequence artefacts (Star *et al.*, 2014). Palindromes are exactly what are expected from HBS-seq, except artefactual sequences would lack the hairpin sequence, and following the model of Star *et al.*, sequences would be complementary only near the ends of the molecules. For HBS-seq to be applicable to ancient DNA, interrupted palindromes must be excluded, either in the laboratory, or in software.

Currently, only one publicly available tool exists for processing HBS-seq data. HBS-tools (Sun *et al.*, 2015), developed for Zhao *et al.* (2014), aligns read one (R1) to read two (R2) using a gapped alignment algorithm (Needleman-Wunsch). As R1 should match R2 up to bisulfite conversion differences, gapped alignment ought to be unnecessary. However, HBS-tools erroneously matches adapter sequences—which must be specified by the user—to the beginning of reads, rather than the end. Consequently, coincidental matches to the adapter can occur in one or a few bases at the start of a read, resulting in trimming and making subsequent gapped alignment between R1 and R2 necessary. This undesirable behaviour can be avoided by specifying an empty adapter sequence.

Despite removal of a few bases at the beginning of reads, HBS-tools performs adequately for reads derived from long inserts. Indeed Zhao *et al.* size selected library molecules between 400–600 bp, presumably avoiding many molecules not containing a hairpin. However, for ancient DNA, where median fragment lengths are on the order of 50 bp, desirable library molecules would have length ∼129 bp (assuming a 29 bp hairpin, as used by Zhao *et al.*). Due to the long tailed distribution of ancient DNA fragment lengths (roughly lognormal, see Renaud *et al.*, 2014), non-hairpin library molecules are more difficult to exclude using size selection, so they must be excluded during read processing. Furthermore, 2x100 or 2x150 paired-end sequencing of libraries derived from short molecules will frequently read through the hairpin and into the other strand, providing additional base calls for some nucleotide positions, which are not considered by HBS-tools during sequence reconstruction. Finally, HBS-tools depends upon a closed source component (`cross_match` from Gordon *et al.*, 1998), preventing modifications and a more detailed assessment.

Here, we present PP5mC for preprocessing HBS-seq data. It uses a straightforward sequence reconstruction approach, matching nucleotides between R1 and R2 according to their position in the reads. An explicit probabilistic model is used for base calls and quality scores, and base calls following the hairpin sequence are considered when they are present. The source code is freely available online and is distributed under a permissive MIT license.

Materials and Methods

`foldreads`

The `foldreads` program attempts to 'fold' paired-end HBS-seq reads at the hairpin sequence, reconstructing the original nucleotides from the homology between R1 and R2. We first search R1 for the hairpin sequence, and R2 for the reverse complement hairpin sequence. If the position of the hairpin differs between R1 and R2, the reads are discarded. This scenario can arise due to polymerase slippage on poly-A and poly-T homopolymers, which are common in bisulfite-treated sequences. If no hairpin sequences are found, R1 and R2 are searched for trailing Y-adapter sequences. If adapters are present, this indicates no hairpin was contained in the library molecule, and reads are discarded on the assumption that they derive from non-canonical HBS-seq molecules.

The position of the hairpin indicates the length of the molecule to be reconstructed. When the molecule length is short relative to the read length, the position of the hairpin is known, and valid bases follow the hairpin sequence.

```
R1 -> ---s1---hairpin---s3---
      ---s4---hairpin---s2--- <- R2
```

In this case, `foldreads` matches the top strand to the bottom strand ($s1$ to $s4$ and $s2$ to $s3$). These sequences are complementary, and errors stem from the sequencing platform. Properly paired nucleotides are one of A/T, T/A, C/G, or G/C (top/bottom strand). Once matched, two sequences remain, one upstream of the hairpin and one downstream.

```
R1 -> ---s1---hairpin
              hairpin---s2--- <- R2
```

This now corresponds to what is observed for long molecules, as the hairpin is absent from reads, or perhaps partially present. In any case, `foldreads` then matches $s1$ with $s2$. They both correspond to the same strand of the original DNA fragment, but may have mismatches resulting from bisulfite conversion of cytosines. Differences may also arise during library amplification or sequencing. C/C and G/G indicate methylated cytosines on the top and bottom strands respectively. T/C and G/A are also valid pairs, indicating unmethylated cytosines on the top and bottom strands respectively. Note that C/T and A/G are not valid because strand orientation following Y-adapter/hairpin ligation is maintained throughout (**Figure 1**).

To identify the hairpin and Y-adapter sequences, and to match sequences $s1/s4$ and $s2/s3$, `foldreads` calculates the most probable base from the FASTQ quality scores, using the model described in Renaud *et al.* (2014). For matching $s1$ to $s2$, the model was extended to permit differences due to bisulfite conversion. The number of mismatches between top and bottom strands is calculated by summing posterior base-error probabilities over all bases in the output sequence. If there are too many mismatches, the read pairs are discarded. We determine the mismatch threshold for different sequence lengths using the Poisson approximation to a binomial distribution, as used by BWA-aln, assuming a 1 % sequencing error rate, with the threshold at 4 % of the Poisson distribution's tail (Li & Durbin, 2009).

Sequences reconstructed by `foldreads` are output in FASTQ format (Cock *et al.*, 2010), with lower case letters in the sequence designating a methylated cytosine (c), or a methylated cytosine on the opposite strand (g). Quality scores for the reconstructed sequences are derived from the posterior base-error probability as in Renaud *et al.* (2014). Additional fields after the read name in the FASTQ file are used to indicate the hairpin sequence, the original read sequences, and their quality scores.

Alignment

The sequence alignment/map (SAM) format (Li *et al.*, 2009), does not permit mixed upper and lower case nucleotides in the SEQ(uence) field (The SAM/BAM Format Specification Working Group, 2018), thus methylation status cannot be directly encoded here. However, when reads are aligned with BWA-mem (Li, 2013) using the `-C` flag, any text in the FASTQ file following the read name is appended verbatim to the optional SAM fields for that read's alignment, allowing information to be stored (and sorted) with the alignments. Hence the additional fields in the FASTQ file output by `foldreads` follow the format required for optional SAM fields.

`scanbp`

During (H)BS-seq library preparation, molecules with single stranded overhangs are 'polished' prior to adapter ligation, typically removing $3'$ overhangs and repairing $5'$ overhangs. For $5'$ overhangs containing guanines, this repair step will result in unmethylated cytosines being incorporated on the other strand, which are subsequently bisulfite converted. Plots showing the empirical frequency of 5mC vs. C along the reads are known as M-bias (methylation bias) plots (Hansen *et al.*, 2012), and can help to identify which parts of the reads should be used to infer methylation status.

In ancient DNA, unmethylated cytosines may spontaneously deaminate into uracils, and when sequenced, this has a similar effect to bisulfite conversion (C→T substitutions). Single-stranded DNA suffers deamination at a higher frequency than does double-stranded DNA, and Briggs *et al.* (2007) showed that the frequency of observed C→T substitutions increases towards the ends of molecules, where single-stranded DNA prevails after post-mortem fragmentation. In addition, Briggs *et al.* reported an excess of purines immediately 5′ of read mapping locations, suggesting a fragmentation bias 3′ of depurinated sites. These characteristic patterns of DNA damage are now routinely assessed for the purpose of authenticating the source of DNA in ancient DNA studies (Llamas *et al.*, 2017).

Using the original R1 and R2 sequences from the optional SAM fields of a sorted alignment, `scanbp` measures the frequency of all sixteen possible nucleotide pairs in reconstructed molecules. Nucleotide frequencies are calculated for each position within the molecule, plus positions upstream and downstream of the alignment. Plots based on this information can be used to simultaneously observe both M-bias and post-mortem damage profiles.

`mark5mC`

Methylation calls are made by `mark5mC`. By moving sequentially along each contig in the sorted alignment, C and 5mC counts are produced for each CpG, CHG, and CHH context, on both strands of the reference sequence. Only reference positions covered by alignments are printed. User-specified parameters indicate how many bases at either end of the reads should not be considered, as identified from M-bias plots.

`simhbs`

The simulator `simhbs` can produce regular BS-seq data, HBS-seq data, and palindromic artefactual reads following the model proposed by Star *et al.* (2014), but with bisulfite conversion. Molecules are drawn from a user-specified reference sequence, bisulfite converted, adapters ligated, and sequencing error (substitutions) applied. The length of the molecules are lognormally distributed, with user specified μ and σ parameters representing the mean and standard deviation of log(read length). Sequencing errors can be specified as a mean error rate, with the probability of sequencing error at each position in each read drawn from a normal distribution, and reflected in the quality scores.

Alternately, an empirical sequencing profile can be used, whereupon the quality scores for a read are modelled as a multivariate Normal (MVN) distribution, with means and covariance matrix estimated from external FASTQ

Table 1: Recovery and mapping of molecules from simulated paired-end reads for each of three molecule types: hairpin ligated (HBS); hairpin missing (BS); and interrupted palindrome (PAL). Two distinct fragment length distributions were simulated: a short length representing ancient DNA ($\mu = 4.0$, median length ∼55 bp); and a longer length where many molecules exceeded the read length ($\mu = 5.0$, median length ∼150 bp). For each combination of molecule type and length distribution, 100 000 2x150 paired-end reads were simulated. The original pre-bisulfite molecules were recovered using HBS-tools' `hbs_process/hbs_mapper` and PP5mC's `foldreads`. Recovered molecules were aligned with BWA-mem and Bowtie1, using default parameters, and we calculated the percentage of simulated molecules that mapped to within 10 bp of the originating location.

mol. type	read length		recovered %		BWA-mapped %		Bowtie-mapped %	
	μ	σ	HBS-tools	PP5mC	HBS-tools	PP5mC	HBS-tools	PP5mC
HBS	4.0	0.25	99.996	100.000	94.138	94.173	93.253	97.117
BS	4.0	0.25	0.000	0.000	0.000	0.000	0.000	0.000
PAL	4.0	0.25	1.562	0.326	1.497	0.319	0.000	0.000
HBS	5.0	0.40	99.827	99.983	97.866	98.351	97.510	99.244
BS	5.0	0.40	0.015	0.000	0.000	0.000	0.000	0.000
PAL	5.0	0.40	28.253	15.994	19.985	10.793	2.383	2.826

files. This method of constructing an empirical error profile produces quality scores that reflect the characteristics of the sequencing technology and chemistry. For Illumina data, sequencing errors are more likely towards the ends of reads, and tend to be highly correlated. Different error profiles may be specified for R1 and R2.

Results & Discussion

Empirical data

We extracted DNA from the petrous bone of a 50 thousand-year-old steppe bison, and constructed an HBS-seq library following the protocol of Zhao *et al.* (2014), with three modifications. Firstly, the DNA extract was treated with a cocktail of uracil-DNA-glycosylase and endonuclease VIII (UDG/endoVIII) prior to library construction, which cleaves single-stranded overhangs at uracil nucleotides when they are present (Briggs *et al.*, 2010). This was done to mitigate the possible confounding effect of post-mortem cytosine deamination on methylation calls. Secondly, we ligated a methylated hairpin, so that incomplete conversion of unmethylated cytosines during bisulfite treatment would have no impact on the hairpin sequence The hairpin was otherwise the same sequence as in Zhao *et al.*, with a biotin-modified thymine for enrichment with streptavidin beads. Finally, we size-selected for 200–400 bp library molecules, to discard molecules without a hairpin. The resulting library was sequenced using a 2x150 kit on an Illumina NextSeq.

Using `foldreads`, we successfully reconstructed molecules from 77.63 % of reads. Examples of canonical HBS-seq molecules, of varying lengths, are shown in **Supplementary Figure S1**. By visual inspection of reads discarded by `foldreads`, we identified several different types of non-canonical HBS-seq library molecules. Displaced hairpins were observed in 4.29 % of molecules—the hairpin in R1 had a different position to the hairpin in R2, resulting from polymerase slippage during amplification (**Supplementary Figure S2.A**). We found that 2.92 % of observed hairpins had a deletion adjacent to the biotinylated thymine (**Supplementary Figure S2.B**). Deletions elsewhere in the hairpin were uncommon, suggesting biotinylated bases contributed to a synthesis issue during hairpin manufacture, or perhaps during library amplification. For the remaining discarded reads, R1 did not match R2 directly, but they were reverse complements, and the lack of a hairpin sequence in these reads suggested that most were regular BS-seq molecules (**Supplementary Figure S2.C,D**). However, some resembled the interrupted palindromes described by Star *et al.* (2014) for ancient DNA libraries constructed with Y-adapters (**Supplementary Figure S2.E,F,G**). The proportion of palindromic artefacts was unclear, as molecules with short palindrome segments were challenging to distinguish from non-hairpin BS-seq molecules.

Nucleotide pairing frequencies from `scanbp` indicated various compositional biases in and around the ends of reconstructed molecules (**Supplementary Figure S3**). The proportion of methylated cytosines increased towards the ends the molecules, as previously reported for some BS-seq libraries (Hansen *et al.*, 2012). This is possibly caused by ligation biases, as the opposite pattern, an increase in unmethylated cytosines towards terminal positions, was anticipated as a side effect from polymerase fill in of 5′ overhangs during end polishing.

While characteristic post-mortem damage patterns were not visible within the reads due to UDG/endoVIII treatment, we did observed a marked increase in the frequency of C/G pairs immediately 5′ of the molecule, and a corresponding increase for G/C pairs immediately 3′ of the molecule. This pattern likely resulted from the removal of uracils by UDG, and subsequent cleavage by endoVIII, in single-stranded overhangs (Briggs *et al.*, 2010). Deamination of unmethylated cytosines into uracils on single-stranded overhangs is a defining characteristic of post-mortem degradation, and hence this supports the

Table 2: Accuracy of pre-mapping base reconstruction and post-mapping C/mC calls. Simulations are the same as those presented in Table 1. Reads were simulated to contain methylated cytosines in a CpG context, and unmethylated cytosines elsewhere, with a 98 % bisulfite conversion efficiency. Methylation calls for a given cytosine context were used to calculate the false positive (FP) mC call rate (where an mC should be called, but was not) and the false negative (FN) mC call rate (where a C should be called, but was not).

mol. type	read length		correct bases %		FP mC call rate %		FN mC call rate %	
	μ	σ	HBS-tools	PP5mC	HBS-tools	PP5mC	HBS-tools	PP5mC
HBS	4.0	0.25	99.942	99.977	0.062	0.012	2.046	2.047
HBS	5.0	0.40	99.500	99.784	0.118	0.106	2.047	2.042

authenticity of the DNA as deriving from a non-modern source.

Simulations

To compare the performance of PP5mC and HBS-tools, we simulated reads derived from molecules with two different length distributions, a 98 % bisulfite conversion rate, and an empirical sequencing error profile taken from an Illumina NextSeq run. BS-seq, HBS-seq, and interrupted palindromes were simulated. Then we evaluated both pipelines on each simulated dataset by counting the number of original molecules recovered, and the number that were subsequently mapped (**Table 1**). The two tools had similar efficacy for the recovery of hairpin-ligated molecules, and the exclusion of molecules without hairpins ligated. PP5mC excluded more interrupted palindromes, although both tools had trouble with this artefact for longer molecules. We note that the initial portion of an interrupted palindrome corresponds to a real endogenous sequence but the putative process generating these artefacts will erase the methylation state on one strand, so exclusion of such molecules is certainly desirable.

PP5mC uses BWA-mem (Li, 2013) as its default mapper, while HBS-tools has Bowtie1 (Langmead *et al.*, 2009) as its default, and we mapped reads recovered from both pipelines using both mappers, in order that fair comparisons be made between the two pipelines. While **Table 1** suggests that Bowtie1 may be a more appropriate mapper for this application when considering the proportion of hairpin reads mapped, and the proportion of palindromic reads excluded, we caution that our simulated dataset is not the most appropriate for evaluating alignment software. Bowtie1 performs end-to-end alignment of reads, without considering indels, whereas BWA-mem is indel aware and can also 'soft clip' reads at either end, to align only part of a read. Soft clipping behaviour likely drives the differences observed for palindromic reads, and we did not simulate indels (neither true differences from the reference, nor sequencing errors), which would certainly skew results in favour of BWA-mem. Reads that align to multiple locations are also treated differently by the two mappers, which may account for the differences with hairpin reads. As Bowtie1 does not calculate mapping quality scores (Li *et al.*, 2008), post-alignment filtering could not be used to improve concordance between the mappers.

From the same simulations, we assessed the proportion of correctly reconstructed bases for sequences that were successfully reconstructed from hairpin-containing molecules. In addition, we evaluated the accuracy of methylation calls, following alignment with HBS-tools' and PP5mC's default mappers (**Table 2**). PP5mC had more accurate base reconstruction, and five-fold lower rate of falsely calling a methylated cytosine for short molecules, but only a small difference between pipelines was found for long molecules. As expected, the ability of PP5mC to use the additional base calls following the hairpin sequence is of much greater utility when molecules are short relative to the read length. In contrast, the frequency of erroneously calling an unmethylated cytosine was almost indistinguishable between the two pipelines, as these errors were dominated by the simulated 2 % of unmethylated sites that weren't bisulfite converted. This highlights the fact that, like BS-seq, accuracy may be limited by biochemical inefficiencies, and not software choice.

To compare the computational efficiency of the pipelines' read reconstruction stages, we modified HBS-tools to exit once the original molecules had been recovered. HBS-tools took 161 seconds to process 100 000 pairs of simulated hairpin reads, with 96 Mb peak memory usage, while PP5mC's `foldreads` took only 2.46 seconds to do the same, with 14 Mb peak memory usage.

Conclusion

Our initial focus for PP5mC was on applicability to short ancient DNA molecules. However, we have shown that PP5mC is equally useful for long molecules, and vastly outperforms HBS-tools in computation time. These factors, and our read simulator, ought to remove existing barriers to the adoption of HBS-seq in new studies.

Acknowledgements

Thanks to Bastien Llamas for ancient bison HBS-seq data and for editing the manuscript; and Dulce Vargas Landin and Gabriel Renaud for discussions and feedback on early versions of this software. This work was supported by an Australian Government Research Training Program Scholarship.

References

Briggs AW, Stenzel U, Johnson PL, Green RE, Kelso J, Prüfer K, Meyer M, Krause J, Ronan MT, Lachmann M, *et al.* (2007). Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A*, **104(37)**:14616–14621. http://dx.doi.org/10.1073/pnas.0704665104

Briggs AW, 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 Res*, **38(6)**:e87. http://dx.doi.org/10.1093/nar/gkp1163

Cock PJA, Fields CJ, Goto N, Heuer ML, & Rice PM (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res*, **38(6)**:1767–1771. http://dx.doi.org/10.1093/nar/gkp1137

Edwards JR, Yarychkivska O, Boulard M, & Bestor TH (2017). DNA methylation and DNA methyltransferases. *Epigenetics Chromatin*, **10(1)**:23. http://dx.doi.org/10.1186/s13072-017-0130-8

Gordon D, Abajian C, & Green P (1998). Consed: A graphical tool for sequence finishing. *Genome Res*, **8(3)**:195–202. http://dx.doi.org/10.1101/gr.8.3.195

Hansen KD, Langmead B, & Irizarry RA (2012). BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol*, **13(10)**:R83. http://dx.doi.org/10.1186/gb-2012-13-10-r83

Jeon J, Choi J, Lee GW, Park SY, Huh A, Dean RA, & Lee YH (2015). Genome-wide profiling of DNA methylation provides insights into epigenetic regulation of fungal development in a plant pathogenic fungus, Magnaporthe oryzae. *Sci Rep*, **5**. http://dx.doi.org/10.1038/srep08567

Kelsey G, Stegle O, & Reik W (2017). Single-cell epigenomics: Recording the past and predicting the future. *Science*, **358(6359)**:69–75. http://dx.doi.org/10.1126/science.aan6826

Krueger F, Kreck B, Franke A, & Andrews SR (2012). DNA methylome analysis using short bisulfite sequencing data. *Nat Meth*, **9(2)**:145–151. http://dx.doi.org/10.1038/nmeth.1828

Laird CD, Pleasant ND, Clark AD, Sneeden JL, Hassan KMA, Manley NC, Vary JC, Morgan T, Hansen RS, & Stöger R (2004). Hairpin-bisulfite PCR: Assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. *Proc Natl Acad Sci U S A*, **101(1)**:204–209. http://dx.doi.org/10.1073/pnas.2536758100

Langmead B, Trapnell C, Pop M, & Salzberg SL (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, **10(3)**:R25. http://dx.doi.org/10.1186/gb-2009-10-3-r25

Li H (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:13033997 [q-bio]*. ArXiv: 1303.3997

Li H & Durbin R (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, **25(14)**:1754–1760. http://dx.doi.org/10.1093/bioinformatics/btp324

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, & Durbin R (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25(16)**:2078–2079. http://dx.doi.org/10.1093/bioinformatics/btp352

Li H, Ruan J, & Durbin R (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res*, **18(11)**:1851–1858. http://dx.doi.org/10.1101/gr.078212.108

Li W & Freudenberg J (2014). Mappability and read length. *Front Genet*, **5**. http://dx.doi.org/10.3389/fgene.2014.00381

Llamas B, Holland ML, Chen K, Cropley JE, Cooper A, & Suter CM (2012). High-resolution analysis of cytosine methylation in ancient DNA. *PLoS One*, **7(1)**. http://dx.doi.org/10.1371/journal.pone.0030226

Llamas B, Valverde G, Fehren-Schmitz L, Weyrich LS, Cooper A, & Haak W (2017). From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *STAR*, **3(1)**:1–14. http://dx.doi.org/10.1080/20548923.2016.1258824

Prüfer K, Stenzel U, Hofreiter M, Pääbo S, Kelso J, & Green RE (2010). Computational challenges in the analysis of ancient DNA. *Genome Biol*, **11(5)**:R47. http://dx.doi.org/10.1186/gb-2010-11-5-r47

Renaud G, Stenzel U, & Kelso J (2014). leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res*, **42(18)**:e141–e141. http://dx.doi.org/10.1093/nar/gku699

Star B, Nederbragt AJ, Hansen MHS, Skage M, Gilfillan GD, Bradbury IR, Pampoulie C, Stenseth NC, Jakobsen KS, & Jentoft S (2014). Palindromic sequence artifacts generated during next generation sequencing library preparation from historic and ancient DNA. *PLoS One*, **9(3)**:e89676. http://dx.doi.org/10.1371/journal.pone.0089676

Sun Ma, Velmurugan KR, Keimig D, & Xie H (2015). HBS-Tools for hairpin bisulfite sequencing data processing and analysis. *Adv Bioinformatics*, **2015**:e760423. http://dx.doi.org/10.1155/2015/760423

The SAM/BAM Format Specification Working Group (2018). Sequence alignment/map format specification. https://samtools.github.io/hts-specs/SAMv1.pdf. [Online; accessed 24-August-2018]

Urich MA, Nery JR, Lister R, Schmitz RJ, & Ecker JR (2015). MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat Protoc*, **10(3)**:475–483. http://dx.doi.org/10.1038/nprot.2014.114

Xu C & Corces VG (2018). Nascent DNA methylome mapping reveals inheritance of hemimethylation at CTCF/cohesin sites. *Science*, **359(6380)**:1166–1170. http://dx.doi.org/10.1126/science.aan5480

Zhang H, Lang Z, & Zhu JK (2018). Dynamics and function of DNA methylation in plants. *Nat Rev Mol Cell Biol*, **19(8)**:489–506. http://dx.doi.org/10.1038/s41580-018-0016-z

Zhao L, Sun Ma, Li Z, Bai X, Yu M, Wang M, Liang L, Shao X, Arnovitz S, Wang Q, *et al.* (2014). The dynamics of DNA methylation fidelity during mouse embryonic stem cell self-renewal and differentiation. *Genome Res*, p. gr.163147.113. http://dx.doi.org/10.1101/gr.163147.113

5.3 Supplementary Figures

A)

```
R1 AGTAAATTAGTTTTTATTAGATTTGGAGTTTTGAGTGGATCGGGTTTAACGCCGGCGGCAAGTGAAGCCGCCGGCGTTAAATTCGATTTATTTAAGGTTTTAAATTTGGTAAGGGTTAATTTACTAGATCGGAAGAGCACACGTCTGAAC
R2 AGTAAATTAACCCTTACCAAATTTAAAACCTTAAATAAATCGAATTTAACGCCGGCGGCTTCACTTGCCGCCGGCGTTAAACCCGATCCACTCAAAACTCCAAATCTAATAAAAACTAATTTACTAGATCGGAAGAGCGTCGTGTAGGGA
FS AgTAAATTAGCCCTTACCAGATTTGGAGCCTTGAGTGGATcgGGTTTA
FQ IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
```

B)

```
R1 GGAGGGTAGAAGGGGTTTAGAGGAAGTGGTATCGGAAAATTTCGGTGTTTTTTTTTAGGGAGATCGGGATGTCGGGGAATTTTGTACGCCGGCGGCAAGTGAAGCCGCCGGCGTATAAAGTTTTTCGGTATTTCGGTTTTTTTGGAGAGG
R2 AAAAAATAAAAAAAACTCAAAAAAAATAATACCGAAAAACCTCGATATTCCTCTCCAAAAAAACCGAAATACCGAAAAACTTTATACGCCGGCGGCTTCACTTGCCGCCGGCGTACAAAATTCCCCGACATCCCGATCTCCCTAAAAAAA
FS GGAGGGTAGAAGGGGCTCAGAGGAAGTGGTACcgGAAAACCTcgGTGTTCCTCTCCAGGGAGACcgGGATGCcgGGGAACTTTGT
FQ CCICCGIIGIIGGGGGIGIGIGGIIGIGGIIGIIGIIIIGGIIIGIGIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
```

C)

```
R1 AGAATTTATATTGGACGTAAATATATTTACGTTGGATATAAATATATATTTAGACGGGTTTTGAGTTTTTTGTTTAAGGATATTGGGTTATTAGGTTTATATTTGTTTGTTTTTTTGATTTTTGTTTTTACGCCGGCGGCAAGTGAAGCC
R2 AAAACTCACACTAAACATAAATACATTCACACTAAACATAAATACACATTCAAACGAACCCTAAACCTCTTACTTAAAAACACTAAATCACCAAATCCACATTTATCTACCTTCTTGATCCCTATTCCTACGCCGGCGGCTTCACTTGCC
FS AGAACTCACACTGGAcGTAAATACATTCAcGCTGGACATAAATACACATTCAGAcgGGCCCTGAGCCTCTTGCTTAAGGACACTGGGTCACCAGGTCCACATTTGTCTGCCTTCTTgATCCCTGTTCCT
FQ 71HICIGIGIGIGGIIG>I>I7IGIIIG7IGGIGGIGI6IIIHICIGI7IGIG6IIG>G1CDGI>CGICD>GGIIIIGGIGIG7GGGIGI>GIGGIG1ICIIIICDGIG>CIIGIDIIICCCIGIIG>D
```

D)

```
R1 GTTTTTGTTTTGAATATGTTGTTTAGGTTGGTTATAATTTTTTTTTTAAGGAGTAAGCGTTTTTAATTTTATGGTTGTAATTATTATTTGTAGTGATTTTGGAGTTTAGAAAAATAAAGTTTGGTATTGTTTTTATTGTTTTTTTATT
R2 ATCTCCACTTTTAAATATACTATCTAAATTAATCATAACTTTCCTTCCAAAAAATAAACGTCTTTTAATTTCATAACTACAATCACCATCTACAATAATTTTAAAACCCAAAAAAATAAAATCTAACACTATTTCCATTATTTCCCCATT
FS GTCTCCGCTTTTGAATATGCTGTCTAGGTTGGTCATAACTTTCCTTCCAAGGAGTAAGcgTCTTTTAATTTCATGGCTGCAATCACCATCTGCAGTGATTTTGGAGCCCAGAAAAATAAAGTCTGGCACTGTTTCCATTGTTTCCCCATT
FQ CICIC1GGIDIIGIIIIIGGIGIGIIGGIIGGIGIDIIGD7IGCIIC17I1CIGIIIGIIICIIIIIIIIICIDGCGIC>IHIGIC1IIGIGGI>IGIIIII1CIGG>>71IHI77IIIHGIGICCG6GIGDIIGC7DD>IIIGGCC7II
```

Supplementary Figure S1: Canonical HBS-seq molecules from a 50 thousand-year-old bison petrosal (sample ACAD16132). The figure shows read pairs (R1 and R2) with a reconstruction of the original molecule (FS) by `foldreads`, and the PHRED+33 quality scores (FQ) that were assigned to the reconstructed bases. Lower case letters in the reconstructed molecule (FS) indicates methylated cytosines on the top (c) or bottom (g) strand. Dark blue text indicates dinucleotides derived from an unmethylated cytosine on the top (T/C) or bottom (G/A) strand. Red text indicates dinucleotides derived from a methylated cytosine on the top (C/C) or bottom (G/G) strand. The hairpin sequence is highlighted in yellow, and the Y-adapter is highlighted in light blue. Underlined nucleotides have a quality score less than or equal to 20 (probability of error is 0.01 or greater). **A)** A short insert, for which both the complete hairpin and the Y-adapter were observed in the reads. **B)** A complete hairpin was sequenced, but no Y-adapter was present. **C)** A partial hairpin was present in the reads. **D)** A long insert where the hairpin was not observed. Quality scores for reconstructed bases in **(A)** were on average higher than those for **(C)** or **(D)**, due to having four observations at every nucleotide position in the original molecule.

A)

```
R1 ATTTTTTTTTTAGGAGATGATGAGATATTATTTATTGTAGGAGTGTAAAGAATATGGTTATTTAGACGCCGGCGGCAAGTGAAGCCGCCGGCGTTTAAATGATTATGTTTTTTATATTTTTGTAATGAGTGGTATTTTATTATTTTTAG
R2 ATTTCTTCCTAAAAAATAATAAAATACCACTCATTACAAAATATAAAAACATAATCATTTAAACGCCGGCGGCTTCACTTGCCGCCGGCGTCTAAATAACCATATTCTTTACACTCCTACAATAAATAATATCTCATCATCTCCTAAA
```

B)

```
R1 AGTTGAGTCGCGCGATATTTGTTTGTTTATTTGTTTGATATTCGTTAGTTAAGTTGACGCCGGCGGCAAGTAAGCCGCCGGCGTCGGTTTAGTTGGCGAGTATTAGGTAGATGAGTAGGTAGGTGTCGCGCGGTTTAGTTAGATCGGAAG
R2 AACTAAACCGCGCGACACCTACCTACTCATCTACCTAATACTCGCCAACTAAACCGACGCCGGCGGCTTACTTGCCGCCGGCGTCAACTTAACTAACGAATATCAAACAAATAAACAAACAAATATCGCGCGACTCAACTAGATCGGAAG
```

C)

```
R1 ACAATNTGGATATGATTGTGTAGATTGGTAATGTTGTGATTTTTGTTGTTTGGTGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACGCCAATATCTCGTATGCCGTCTTCTGCTTGAAAAAAAAAAGGGGGGGGGGGGGGGGGGGGG
R2 CACCAAACAACAAAAATCACAACATTACCAATCTACACAATCATATCCACATTGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
```

D)

```
R1 TGTTGTAGTATTTATATTATGGTGGTGTGTTTTTTGTGTTGGTTGTGTTTGGTAATAAGTTTGGTTATGTTATAGATGGTGAATAGGATAATGATAATTTGAATTTTGATTTGGAAGTTTTTTATTATGGTGTTGTTTTTGGTGGTGGTG
R2 ACAAACCAACCATAAAACCACCACCACCAAAAACAACACCATAATAAAAACTTCCAAATCAAAATTCAAATTATCATTATCCTATTCACCATCTATAACATAACCAAACTTATTACCAAACACCACCAACACAAAACCACACCACCAT
```

E)

```
R1 TTGTAAGAAATTTATTGATTTTGTTGAAGGGTATTAGTTTTTTAGTAAGATTAGTGGATTTTTTATAAAGATCGGAAGAGCACACGTCTGAACTCCAGTCACGCCAATATCTCGTATGCCGTCTTCTGCTTGAAAAAAAAAAGGGGGGGG
R2 TTATAAAAAATCCACTAATCTTACTAAAAAACTAATACCCTTCAACAAAATCAATAAATTTCTTACAAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATTAAAAAAAAAAGGGGGGGGGGGGGG
```

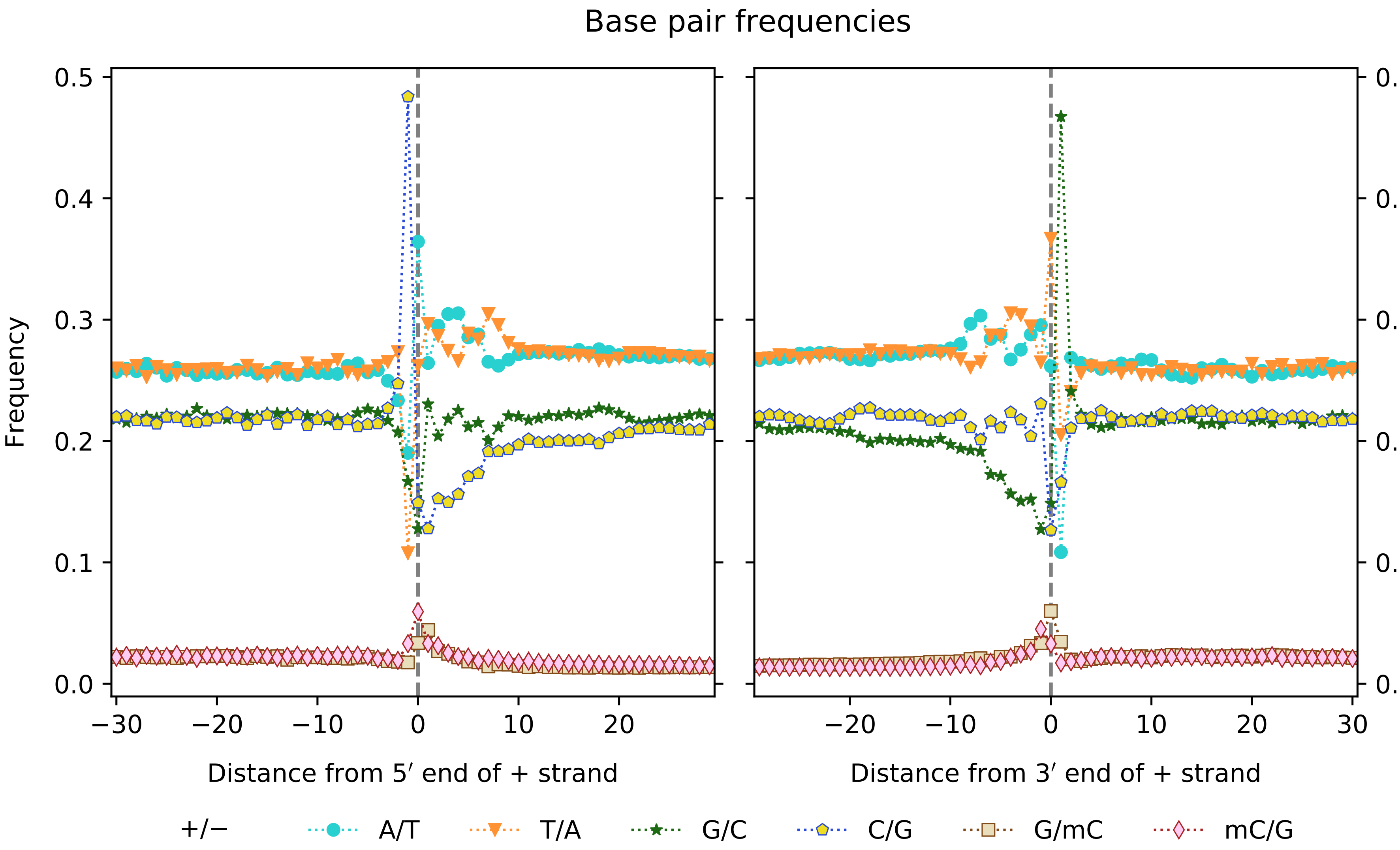
F)

```
R1 AAATATTAAGAAATATTATAAATTATATATTAATGTAATGGATAGGGTAGAAGAAATGGATATAAATTTAAAAATGTATAATTTTTTTTATTTTGTTTATTGTATTGGTATATAATTTGTAATGTTTTTTAGTATTTAGATCGGAAGAGC
R2 AAATACTAAAAAACATTACAAATTATATACCAATACAATAAACAAAATAAAAAAAATTATACATTTTTAAATTTATATCCATTTCTTCTACCCTATCCATTACATTAATATATATAATTTATAATATTTCTTAATATTTAGATCGGAAGAGC
```

G)

```
R1 TAATTTAAAATGAATATGTAATTTGTTAGTTTGAGGTAAATAGAATAGTATTTTTGGGATTAGTTGTTATATGAAGAAATAGTGGGTTAGTTAAAAAAGTTGTTCGGGTTTTTTTGTAATATGGAAAAATCCAAACGAACTTTTGGTTA
R2 TAATCTAAAATAAACATACAATTTACCAATTTAAAATAAACAAAACAATATCTCTAAAATCAATTATTATATAAAATAATAAATTAACCAAAAAATTTATTCAAATTTTTCTATAATATAAAAAATCCAAACGAACTTTTGACCAACC
```

Supplementary Figure S2: Non-canonical HBS-seq molecules (discarded by `foldreads`) in a library prepared from a 50 thousand-year-old bison petrosal (sample ACAD16132). Colours have the same meaning as in **Supplementary Figure S1**, with the addition of pink text to indicate dinucleotide mismatches. **A)** The polymerase inserted or deleted a single nucleotide, as evidenced from differing locations for the hairpin sequence in R1 and R2. **B)** The hairpin sequence in R1 has a one nucleotide deletion immediately 3′ of the biotinylated thymidine, which arose during hairpin manufacture or subsequent library amplification. As `foldreads` permits multiple hairpin sequences to be specified in a single run, the original molecule can be successfully reconstructed provided the modified hairpin sequence is (also) specified. **C)** A short insert without a hairpin sequence, but Y-adapters were present. R1 and R2 are reverse complements. **D)** A long insert with neither a hairpin nor Y-adapters. R1 and R2 are reverse complements for most of their lengths—from position 19 of R1 onwards, nucleotides are complementary with the reverse of R2. This molecule is typical of a regular BS-seq library. **E)** & **F)** Interrupted palindromes. R1 and R2 are reverse complements, however the initial and final portions of R1 and R2 match directly, which is unlikely to have occurred by chance. **G)** Possible long insert version of **(E)** & **(F)**.



Supplementary Figure S3: Frequencies of base pairs observed in an HBS-seq library prepared from a 50 thousand-year-old bison petrosal (sample ACAD16132). Nucleotides in the top/bottom strands of **Figure 1** are represented here as +/− strands. Frequencies were calculated for each position within the reconstructed molecules, as a function of the distance from either end. Vertical gray lines correspond to the first and last base pairs within the reconstructed molecule. From each molecule's alignment to the genome reference, the frequencies upstream of the 5′ position and downstream of the 3′ position, were calculated using the adjacent read pileup. Several different processes contribute to the compositional biases that are apparent in and around the ends of the reconstructed molecules. These include: post-mortem deamination of cytosines, post-mortem fragmentation, UDG/endoVIII treatment, end repair, and ligation biases. In particular, the C/G (G/C) peak immediately 5′ (3′) of the molecules derive from post-mortem deamination of unmethylated cytosines to uracils in single-stranded overhangs, and the subsequent cleavage of uracil bases by UDG/endoVIII. We note that the overall GC % observed is likely higher than the true GC content of the individual, because short, GC-rich, DNA fragments may be preferentially amplified during PCR.

Chapter 6

Discussion

6.1 Research summary

A variety of approaches exist to investigate demographic and evolutionary processes with genetic data. This can be done from the relationships between extant lineages and using the signatures of ancestral populations that remain in the genomes of their descendants. It is also possible to directly observe past populations, or observe populations more closely related to those of interest, using aDNA. These approaches are complementary and in this thesis I considered both when applying, evaluating, and extending methods to understand past mammal populations. By applying drift-based statistics to modern and ancient data, I detected and quantified gene flow in the ancestors of European bison. I assessed the robustness of two popular methods for estimating past population sizes. I extended methods for genetic sexing of ancient remains and applied this to bison and brown bear specimens, and collated data from online databases to discern sex ratios in large mammal collections. Finally, I designed and implemented new software to process HBS-seq data, which promises to broaden the availability of DNA methylation profiles for ancient samples, thus expanding the potential for demographic and functional analyses.

6.2 Primary outcomes

6.2.1 European bison ancestry

In chapter 2 (Soubrier *et al.*, 2016), we identified that while modern European bison (*Bison bonasus*) are descended from steppe bison (*Bison priscus*), they also derive a non-zero proportion of ancestry from aurochs (*Bos primigenius*), though no more than ∼10 %. A similar result was obtained for a 22 thousand-year-old bison specimen that also possessed a *Bos*-like mitochondrial lineage. This suggests that the gene flow was ancient, and could not have been caused by potential Holocene interactions with domestic cattle. Nor could it be an artefact of 20th century conservation practices following the extinction of European bison in the wild. Further, these results indicate that introgression occurred after the split of European and American bison, but likely predates the split of the two *Bos*-like bison mtDNA lineages, *i.e.* 120–240 thousand years ago. So bison and aurochs must have had range and habitat overlap at times during the Late Pleistocene. This highlights the value of using ancient DNA to investigate gene flow in a severely bottlenecked population, with potentially confounding effects of recent human activities.

6.2.2 PSMC and MSMC with short scaffolds

The papers that introduced PSMC (Li & Durbin, 2011) and MSMC (Schiffels & Durbin, 2014) have a combined total of over 1300 citations (Google Scholar; accessed 16 Oct 2018). Inferring past population sizes is very popular because large-scale demographic changes can often be viewed as an indirect result of changes in environmental factors such as temperature and precipitation. While these tools have undoubtedly already been applied to data comprised of short scaffolds, the accuracy of such results is far from obvious. The SMC model and HMM inference are both non-trivial statistical frameworks, so using simulations to assess the behaviour of PSMC and MSMC is essential in understanding their limitations. In chapter 3, population size inference is shown to be consistent from genomic scaffolds as short as 100 kb when using PSMC, and 1 Mb when using MSMC. Users of either tool can now be confident that simply excluding the shortest scaffolds in their dataset will produce the most robust estimates. If only ultra-short scaffolds are available, perhaps from short-read *de novo* assembly, then PSMC may still be useful, as it can reproduce major demographic shifts despite a divergence from the consensus inferences for sub-100 kb scaffolds.

6.2.3 A male bias is ubiquitous in mammal collections

In chapter 4, I showed that approximately 75 % of bison and brown bear subfossil remains are male, very similar to the ratio observed for mammoth remains (Pečnerová *et al.*, 2017). One possible explanation for this is that male and female bones have a different preservation potential, perhaps due to intrinsic differences in bone density between the sexes. But an assessment of preservation-related attributes for each sample indicated no differences between male and female remains in this respect. The selection of larger *Myotragus balearicus* specimens, on the assumption they would be more likely to yield DNA, resulted in only males being sampled. Deliberate collection of large samples resembles trophy sampling by hunters, where large impressive-looking males may be preferred targets. Notably, mammalian museum collections derive a substantial proportion of samples from individuals that were hunted or trapped in recent centuries. By surveying four large databases of mammal collections, we found that most species are not represented by a 1:1 sex ratio, and when averaged across species, most orders were male biased.

Extreme differences in the number of males versus females are almost certainly caused by ecological or behavioural characteristics. Barnosky (1985) suggested that exclusively male Irish elk samples found at one site resulted from winter deaths in a seasonally sex-segregated population. Similarly, sex segregation has been proposed to explain the overabundance of young male mammoths in some assemblages (Haynes, 2017), the hypothesis also advocated by Pečnerová *et al.* (2017). Extant bison populations have segregated sexes for most of the year too, and with the weight of these examples, it is natural to consider whether this is the ultimate cause of sex-biased observations in all herding mammals. In contrast, brown bears are mostly solitary, and so sex segregation in and of itself is unlikely to be the causative agent.

Genetic sexing of aDNA specimens from shotgun and SNP data is increasingly routine. The methods are effective, simple to apply, and the results can be insightful. The explicit binomial models I developed and used for sex determination are not always necessary, as approximate methods perform well (Skoglund *et al.*, 2013; Mittnik *et al.*, 2016). However, an ad-hoc approach may yield false confidence in a sex assignment, whereas model selection via a likelihood ratio test will indicate when the data are insufficient to confidently distinguish the sexes. My approach can be rigorously applied to samples for which very small numbers of reads have been sequenced, and does not rely on sufficiently large numbers of both sexes in order to obtain a threshold value for sex assignment.

6.2.4 More efficient processing of HBS-seq data

Methylomes of ancient specimens are, in principal, an excellent resource for learning about past populations. Methylation levels at specific loci have been associated with ontological age (Horvath & Raj, 2018), environmental exposure (Bind *et al.*, 2014; Metzger & Schulte, 2017), and nutrition (Gokhman *et al.*, 2017). But DNA methylation remains largely unexplored in aDNA studies. This predominantly reflects the difficulty of obtaining methylomes from ancient samples. A little-used approach for profiling DNA methylation is HBS-seq (Laird *et al.*, 2004; Zhao *et al.*, 2014), which has a distinct advantage for aDNA compared with traditional bisulfite-sequencing protocols. But HBS-seq reads must be preprocessed using specialised software prior to mapping, and the only pipeline currently available, HBS-tools (Sun *et al.*, 2015), was not designed with the limitations of aDNA in mind. PP5mC, a new HBS-seq data processing toolkit, was presented in chapter 5. The toolkit includes a read simulator for HBS-seq reads, regular bisulfite sequencing reads, and artefactual reads that may exist in aDNA HBS-seq libraries. HBS-tools and PP5mC were compared using simulated reads, which showed that PP5mC is: 65× faster than HBS-tools at processing reads; reproduces the original methylation levels with greater accuracy; and excludes a greater proportion of artefactual reads. PP5mC was successfully applied to HBS-seq data generated for an ancient bison specimen. As HBS-seq libraries for aDNA can contain a variety of different molecules, not all desirable, a tool is provided to assist with visually inspecting HBS-seq reads, which may be valuable for optimising HBS-seq protocols.

6.3 Synthesis

The detection of gene flow between the ancestors of European bison and cattle may not be entirely surprising—their hybrid offspring is often fertile, and the mtDNA are discordant with the nuclear phylogeny—but genomic evidence of this had not been confirmed until recently (Soubrier *et al.*, 2016; Gautier *et al.*, 2016). This result adds to a growing body of genomic studies identifying interspecies gene flow in mammals. Besides the high-profile introgression of Neandertal and Denisovan genetic material into non-African humans (Green *et al.*, 2010; Reich *et al.*, 2010), gene flow has also been detected between wild mammals, such as between polar and brown bears (Miller *et al.*, 2012), chimpanzees and bonobos (Manuel *et al.*, 2016), among multiple felids (Li *et al.*, 2015; Figueiró *et al.*, 2017); and during domestication, such as in dogs (Skoglund *et al.*, 2015), pigs (Bosse *et al.*, 2014), goats (Daly *et al.*, 2018), and between many *Bos* species (Wu *et al.*, 2018). Hybridisation is increasingly being recognised as an important force in genome evolution (Sankararaman *et al.*, 2014, 2016; Schumer *et al.*, 2018; Ivancevic *et al.*, 2018; Runemark *et al.*, 2018), because so called *hybrid incompatibilities* can result in very different patterns of ancestry in functional versus non-functional parts of the genome. Incompatibilities need not be the only cause for these patterns, which may also be driven by differences in mutational load of each of the parent lineages (Harris & Nielsen, 2016).

Demographic parameters of a population serve to quantify how a species interacts with its environment. Population size changes that are concomitant with the arrival of humans, the extinction of a prey, or extreme climate fluctuations, are highly suggestive (Shapiro *et al.*, 2004; Campos *et al.*, 2010; Lorenzen *et al.*, 2011; Miller *et al.*, 2012; Cooper *et al.*, 2015). Of course, correlation is not causation, so temporal associations must be interpreted cautiously, and interpretations should be tested using other means if possible (Metcalf *et al.*, 2014). Population structure can also produce signals of population size change (discussed below), which may lead to a very different interpretation on the impact of environmental conditions.

A standing assumption is that populations ancestral to living species behaved similarly to their extant descendents. This assumption is often made by necessity (we can directly observe the behaviour of living animals only), and it is important that the data we obtain for past populations are consistent with living relatives. Male-biased sex ratios in bison and brown bear remains, while initially surprising, are consistent with what we know about sex-segregation, and differences in male and female home ranges, for extant populations.

6.4 Limitations and future directions

6.4.1 Models for the origin of European bison

The detection of aurochs gene flow into the ancestors of European bison does not permit the conclusion that the *Bos*-like mtDNA lineage was introgressed. The mtDNA is a single locus, and incomplete lineage sorting could also produce a discordance with the species tree. This has been advocated as the explanation for this mtDNA discordance (Massilani *et al.*, 2016; Grange *et al.*, 2018), but it remains difficult to distinguish this hypothesis from introgression without obtaining aDNA from aurochs and steppe bison specimens corresponding to the putative period of gene flow.

Grange *et al.* (2018) have criticised the work discussed in chapter 2 (Soubrier *et al.*, 2016). In particular, they show using principal components analysis that our SNP capture data do not cluster with bison sequences from other studies (Gautier *et al.*, 2016; Węcek *et al.*, 2017). This is concerning, and may result from our SNP ascertainment and enrichment strategy, which targeted sites that are polymorphic in domestic cattle, and used RNA baits derived from domestic cattle sequences. Combined with the subsequent mapping of reads to the cattle reference genome (bison and cattle had a common ancestor ∼1.2 million years ago (Wu *et al.*, 2018)), these factors could have introduced a substantial bias towards observing cattle genotypes. Even so, the comparison is not entirely fair, as our data presented in Grange *et al.* (2018) was processed using a different pipeline and a different genotype calling model, compared with the rest of the data, which could also contribute to batch-related effects. Nevertheless, aurochs gene flow into the ancestors of European bison has been confirmed elsewhere (Gautier *et al.*, 2016; Węcek *et al.*, 2017), and shotgun sequencing data for many of the specimens from chapter 2 will soon be available (van Loenen *et al.*, in prep.), which will allow further exploration of the nature and timing of the introgression.

6.4.2 SMC-based population size inference

In chapter 3 I demonstrated that short reference scaffold lengths limit the accuracy of SMC-based analyses only mildly. However, severe limitations on the interpretability of results derived from these methods remain. To interpret the output from PSMC and MSMC, this output must be rescaled to represent real time and effective population size, using the average generation time, and the mutation rate of the organism under study (Li & Durbin, 2011; Schiffels & Durbin, 2014). For rescaling SMC-based output, it is preferred to use a mutation rate corresponding to a long-term average, and aDNA can be used to obtain an estimate (Fu *et al.*, 2014; Skoglund *et al.*, 2015), but there remains an unexplained discordance between mutation rates estimated over long timescales compared with estimates from *de novo* mutations (Scally & Durbin, 2012; Moorjani *et al.*, 2016a). Similarly, the long term average generation time should be used, and has been calculated for humans (Moorjani *et al.*, 2016b), but in absence of this information the generation time for an extant population is typically substituted.

Population structure is the elephant in the room with regard to population size inferences. Irrespective of the method, it is not the census population size which is inferred, but N_e, the *effective* population size (Wright, 1931). This is the size of an idealised Wright-Fisher population, which does not exist in general, and which has the same amount of genetic diversity as the population under study. A reduction in N_e compared with the census population size can arise in a panmictic population due to variation in the number of offspring between individuals or sexes (Wright, 1931). Wahlund's principle suggests that the partial or complete isolation of demes within a population will result in individuals with lower genetic diversity than the population size indicates; in contrast, migration between demes can artificially increase the observed diversity at the population level (Crow & Kimura, 1970, pp. 54–55). As population structure is often cryptic, particularly in unobserved past populations, this can be problematic for population size inferences. The coalescent-based models used by PSMC and MSMC have been shown to produce identical results for real population size changes, and for populations with a fixed size but fluctuating migration between many demes (Leblois *et al.*, 2006; Heller *et al.*, 2013; Mazet *et al.*, 2015, 2016; Chikhi *et al.*, 2018).

It is not only coalescent-based methods that are affected, however, as approaches using site frequency spectra to infer population size (Gutenkunst *et al.*, 2009; Kamm *et al.*, 2018) are also potentially vulnerable to population structure (Städler *et al.*, 2009). Differences in the respective susceptibilities to this problem may, to some extent, drive differences in the results obtained for the two classes of population size inference methodologies (Beichman *et al.*, 2017). Resolving the issue caused by population structure remains an open problem, but in some cases it may be detectable. As migrations between demes are often sex-biased, it may be possible to distinguish between population structure with migrations, and true population size changes, by looking at the concordance between population size inferences from the X chromosome, and inferences from autosomes. However, this is not always possible. For example, I contributed to a project (Feigin *et al.*, 2018) where I performed population size inference for the extinct marsupial wolf (*Thylacinus cynocephalus*), which has no close living relatives. In this case, I conservatively excluded X-linked scaffolds from analysis based on homology to a distantly related X chromosome, but these scaffolds were not a reliable representative of X chromosome data. We were careful to consider population structure as a possible explanation for the population size inferences presented.

6.4.3 Drivers of male-biased sex ratios

The lone male model advocated by Pečnerová *et al.* (2017), and differences in home range between sexes, are related hypotheses that are both consistent with the observed sex ratios. But neither is readily falsifiable with the data at hand. No difference in the spatial distributions of males and females were identified for bison or brown bears in chapter 4, but this is likely due to temporal blurring of the distributions obtained from heterochronous specimens. We might more confidently accept the lone male model for mammoths and bison if age-at-death profiling reveals that a large proportion of male samples are young adults. DNA methylation is a promising source of data to assess ontogenic age, which we anticipate will illuminate future sex-ratio analyses.

Some methylomic loci exhibit changes in their methylation level in an age-dependent manner throughout the life of the individual, which can be used for age determination (Horvath, 2013; De Paoli-Iseppi *et al.*, 2017), and this has been successfully applied to 4000 year-old human hair (Pedersen *et al.*, 2014). As the number of age-informative sites in the methylome is likely small (*e.g.* five sites identified in Koch & Wagner, 2011), shotgun sequencing may be an expensive route to this information. Smith *et al.* (2014) used methyl-binding domain (MBD) enrichment, and bisulfite treatment to investigate methylation levels in Iron Age barley. But MBD-based enrichment produces results that are unacceptably biased towards molecules containing methylated CpGs and longer molecules containing a greater number of CpGs (Seguin-Orlando *et al.*, 2015). Bisulfite-treated PCR amplicon sequencing (Llamas *et al.*, 2012; Smith *et al.*, 2015), or an RNA bait-set comprised of all methylation-state combinations, may be the most effective and cost-efficient ways to survey many samples at a small number of loci.

6.4.4 Prospects for paleo-epigenetics

It remains to be seen whether the use of HBS-seq will be revolutionary for studying paleo-epigenetics, or merely an incremental improvement. Regardless, more work is needed to make full use of HBS-seq data. Post-mortem deamination of a methylated cytosine results in a T/G pair of nucleotides being observed, which is identical to that obtained from bisulfite conversion of an unmethylated cytosine. Hence the inference of methylation levels is confounded with post-mortem damage in degraded samples, resulting in overall hypomethylation compared to pre-mortem levels. The extent of hypomethylation will likely vary depending on the preservation conditions of the sample, such as temperature and humidity. No current method for detecting differential methylation from bisulfite-sequencing data considers such methylome-wide differences. Some work has been done already to translate regional CpG→TpG substitution levels from UDG treated aDNA data into pre-mortem methylation levels (Gokhman *et al.*, 2014), which may be adaptable to HBS-seq data from ancient remains. Alternately, new statistical models will need to be developed to identify differentially methylated regions that account for stochastic hypomethylation.

Ancient DNA extracted from bone is potentially derived from a mixture of distinct cell types (osteoblasts, osteoclasts, and osteocytes). Because DNA methylation is involved in cell differentiation and gene regulation, methylation levels at some loci can differ between cell types (Meissner *et al.*, 2008). It may be possible to determine the extent of cell-type heterogeneity in ancient samples using HBS-seq data (Titus *et al.*, 2017). This would be interesting to understand how different cell types contribute to DNA preservation, but may also be necessary to distinguish within sample, and between sample, variability of methylation levels.

6.4.5 PP5mC computational performance

Like any high-throughput sequencing dataset, HBS-seq datasets can be very large. CPU-time and memory consumption required for processing the data are thus important factors for consideration. The long-running components of PP5mC (`foldreads`, `scanbp`, `mark5mC`) have all been designed to use memory conservatively, and to avoid needless computation. However, the implementations are currently single threaded, which can limit overall throughput when sample-level parallelism is insufficient to use all the available compute resources (*e.g.* when there are fewer samples than CPU cores). Adding multithreading support to `foldreads` ought to be simple and effective, as it processes each read independently and is thus in the class of algorithms known as *embarrass-*

ingly parallel. Langmead *et al.* (2018) provide a recent discussion of strategies for scaling multithreaded applications that process reads independently from FASTQ files. Both `scanbp` and `mark5mC` process indexed BAM files. Thus parallelism could be trivially obtained by working on each chromosome separately, or on fixed-size blocks within chromosomes, then combining the separate results into a single output file. `simhbs` is also single-threaded, with much CPU-time being spent on sampling multinomially distributed quality scores for empirical error profiles. Simulation of 100 000 2x150 bp sequences takes less than seven seconds on an i5-3320M processor, which ought to be sufficient for most purposes. If not, separate `simhbs` invocations can be run independently, one for each CPU-core available, and the output combined into a single file using standard UNIX tools. PP5mC already provides a clear speed advantage over its only competitor, but if there were sufficient interest, multithreading support could be added with relatively little effort.

6.5 Concluding remarks

Detection of gene flow, and the use of coalescent-based inference frameworks, can provide detailed information about how populations are related, how these relationships have changed over time, and the timing of these changes. Sequencing data from ancient specimens are increasingly available, and should be integrated with high-quality modern datasets wherever possible. Shotgun data are preferred over data enrichment for specific loci, unless guarantees can be made that sites targeted for enrichment, and the enrichment process itself, do not produce or contribute to artificial relationships between samples. The robustness of available tools is an important consideration when applying them to data with characteristics not assessed in their original publication, and caution is recommended when applying even established methods, as there may be multiple interpretations that are consistent with the results. Similarly, new tools must be compared against existing software to confirm that genuine advancements are being made. Analysis of the same data using different methods can sometimes produce different results, so complementary investigation of different samples, and new types of data, is recommended to elicit much greater confidence in any conclusions.

6.6 References

Barnosky AD (1985). Taphonomy and herd structure of the extinct Irish elk, *Megaloceros giganteus*. *Science*, **228(4697)**:340–344. http://dx.doi.org/

10.1126/science.228.4697.340

Beichman AC, Phung TN, & Lohmueller KE (2017). Comparison of single genome and allele frequency data reveals discordant demographic histories. *G3 (Bethesda)*, **7(11)**:3605–3620. http://dx.doi.org/10.1534/g3.117.300259

Bind MA, Zanobetti A, Gasparrini A, Peters A, Coull B, Baccarelli A, Tarantini L, Koutrakis P, Vokonas P, & Schwartz J (2014). Effects of temperature and relative humidity on DNA methylation. *Epidemiology*, **25(4)**:561–569. http://dx.doi.org/10.1097/EDE.0000000000000120

Bosse M, Megens HJ, Madsen O, Frantz LAF, Paudel Y, Crooijmans RPMA, & Groenen MAM (2014). Untangling the hybrid nature of modern pig genomes: a mosaic derived from biogeographically distinct and highly divergent *Sus scrofa* populations. *Mol Ecol*, **23(16)**:4089–4102. http://dx.doi.org/10.1111/mec.12807

Campos PF, Willerslev E, Sher A, Orlando L, Axelsson E, Tikhonov A, Aaris-Sørensen K, Greenwood AD, Kahlke RD, Kosintsev P, *et al.* (2010). Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (*Ovibos moschatus*) population dynamics. *Proc Natl Acad Sci U S A*, **107(12)**:5675–5680. http://dx.doi.org/10.1073/pnas.0907189107

Chikhi L, Rodríguez W, Grusea S, Santos P, Boitard S, & Mazet O (2018). The IICR (inverse instantaneous coalescence rate) as a summary of genomic diversity: insights into demographic inference and model choice. *Heredity (Edinb)*, **120(1)**:13–24. http://dx.doi.org/10.1038/s41437-017-0005-6

Cooper A, Turney C, Hughen KA, Brook BW, McDonald HG, & Bradshaw CJA (2015). Abrupt warming events drove Late Pleistocene Holarctic megafaunal turnover. *Science*, **349(6248)**:602–606. http://dx.doi.org/10.1126/science.aac4315

Crow JF & Kimura M (1970). *An introduction to population genetics theory*. Burgess Publishing Company, Minneapolis, Minnesota. ISBN 978-1-932846-12-6

Daly KG, Delser PM, Mullin VE, Scheu A, Mattiangeli V, Teasdale MD, Hare AJ, Burger J, Verdugo MP, Collins MJ, *et al.* (2018). Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science*, **361(6397)**:85–88. http://dx.doi.org/10.1126/science.aas9411

De Paoli-Iseppi R, Deagle BE, McMahon CR, Hindell MA, Dickinson JL, & Jarman SN (2017). Measuring animal age with DNA methylation: from humans to wild animals. *Front Genet*, **8**. http://dx.doi.org/10.3389/fgene.2017.00106

Feigin CY, Newton AH, Doronina L, Schmitz J, Hipsley CA, Mitchell KJ, Gower G, Llamas B, Soubrier J, Heider TN, *et al.* (2018). Genome of the Tasmanian tiger provides insights into the evolution and demography of an extinct marsupial carnivore. *Nat Ecol Evol*, **2(1)**:182–192. http://dx.doi.org/10.1038/s41559-017-0417-y

Figueiró HV, Li G, Trindade FJ, Assis J, Pais F, Fernandes G, Santos SHD, Hughes GM, Komissarov A, Antunes A, *et al.* (2017). Genome-wide signatures of complex introgression and adaptive evolution in the big cats. *Sci Adv*, **3(7)**:e1700299. http://dx.doi.org/10.1126/sciadv.1700299

Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, Johnson PLF, Aximu-Petri A, Prüfer K, de Filippo C, *et al.* (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature*, **514(7523)**:445–449. http://dx.doi.org/10.1038/nature13810

Gautier M, Moazami-Goudarzi K, Levéziel H, Parinello H, Grohs C, Rialle S, Kowalczyk R, & Flori L (2016). Deciphering the wisent demographic and adaptive histories from individual whole-genome sequences. *Mol Biol Evol*, **33(11)**:2801–2814. http://dx.doi.org/10.1093/molbev/msw144

Gokhman D, Lavi E, Prufer K, Fraga MF, Riancho JA, Kelso J, Paabo S, Meshorer E, & Carmel L (2014). Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. *Science*, **344(6183)**:523–527. http://dx.doi.org/10.1126/science.1250368

Gokhman D, Malul A, & Carmel L (2017). Inferring past environments from ancient epigenomes. *Mol Biol Evol*, **34(10)**:2429–2438. http://dx.doi.org/10.1093/molbev/msx211

Grange T, Brugal JP, Flori L, Gautier M, Uzunidis A, Geigl EM, Grange T, Brugal JP, Flori L, Gautier M, *et al.* (2018). The evolution and population diversity of bison in Pleistocene and Holocene Eurasia: sex matters. *Diversity*, **10(3)**:65. http://dx.doi.org/10.3390/d10030065

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MHY, *et al.* (2010). A draft sequence of the Neandertal genome. *Science*, **328(5979)**:710–722. http://dx.doi.org/10.1126/science.1188021

Gutenkunst RN, Hernandez RD, Williamson SH, & Bustamante CD (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet*, **5(10)**:e1000695. http://dx.doi.org/10.1371/journal.pgen.1000695

Harris K & Nielsen R (2016). The genetic cost of Neanderthal introgression. *Genetics*, **203(2)**:881–891. http://dx.doi.org/10.1534/genetics.116.186890

Haynes G (2017). Finding meaning in mammoth age profiles. *Quat Int*, **443**:65–78. http://dx.doi.org/10.1016/j.quaint.2016.04.012

Heller R, Chikhi L, & Siegismund HR (2013). The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. *PLoS One*, **8(5)**:e62992. http://dx.doi.org/10.1371/journal.pone.0062992

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

Horvath S & Raj K (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet*, **19(6)**:371–384. http://dx.doi.org/10.1038/s41576-018-0004-3

Ivancevic AM, Kortschak RD, Bertozzi T, & Adelson DL (2018). Horizontal transfer of BovB and L1 retrotransposons in eukaryotes. *Genome Biol*, **19(1)**:85. http://dx.doi.org/10.1186/s13059-018-1456-7

Kamm JA, Terhorst J, Durbin R, & Song YS (2018). Efficiently inferring the demographic history of many populations with allele count data. *bioRxiv*, p. 287268. http://dx.doi.org/10.1101/287268

Koch CM & Wagner W (2011). Epigenetic-aging-signature to determine age in different tissues. *Aging (Albany NY)*, **3(10)**:1018–1027. http://dx.doi.org/10.18632/aging.100395

Laird CD, Pleasant ND, Clark AD, Sneeden JL, Hassan KMA, Manley NC, Vary JC, Morgan T, Hansen RS, & Stöger R (2004). Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. *Proc Natl Acad Sci U S A*, **101(1)**:204–209. http://dx.doi.org/10.1073/pnas.2536758100

Langmead B, Wilks C, Antonescu V, Charles R, & Hancock J (2018). Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics*, **bty648**:12. http://dx.doi.org/10.1093/bioinformatics/bty648

Leblois R, Estoup A, & Streiff R (2006). Genetics of recent habitat contraction and reduction in population size: does isolation by distance matter? *Mol Ecol*, **15(12)**:3601–3615. http://dx.doi.org/10.1111/j.1365-294X.2006.03046.x

Li G, Davis B, Eizirik E, & Murphy W (2015). Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Res*, p. gr.186668.114. http://dx.doi.org/10.1101/gr.186668.114

Li H & Durbin R (2011). Inference of human population history from individual whole-genome sequences. *Nature*, **475(7357)**:493–496. http://dx.doi.org/10.1038/nature10231

Llamas B, Holland ML, Chen K, Cropley JE, Cooper A, & Suter CM (2012). High-resolution analysis of cytosine methylation in ancient DNA. *PLoS One*, **7(1)**:e30226. http://dx.doi.org/10.1371/journal.pone.0030226

Lorenzen ED, Nogués-Bravo D, Orlando L, Weinstock J, Binladen J, Marske KA, Ugan A, Borregaard MK, Gilbert MTP, Nielsen R, *et al.* (2011). Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, **479(7373)**:359–364. http://dx.doi.org/10.1038/nature10574

Manuel Md, Kuhlwilm M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J, Hernandez-Rodriguez J, Dupanloup I, Lao O, Hallast P, *et al.* (2016). Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science*, **354(6311)**:477–481. http://dx.doi.org/10.1126/science.aag2602

Massilani D, Guimaraes S, Brugal JP, Bennett EA, Tokarska M, Arbogast RM, Baryshnikov G, Boeskorov G, Castel JC, Davydov S, *et al.* (2016). Past climate changes, population dynamics and the origin of Bison in Europe. *BMC Biol*, **14**:93. http://dx.doi.org/10.1186/s12915-016-0317-7

Mazet O, Rodríguez W, & Chikhi L (2015). Demographic inference using genetic data from a single individual: separating population size variation from population structure. *Theor Popul Biol*, **104**:46–58. http://dx.doi.org/10.1016/j.tpb.2015.06.003

Mazet O, Rodríguez W, Grusea S, Boitard S, & Chikhi L (2016). On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference? *Heredity (Edinb)*, **116(4)**:362–371. http://dx.doi.org/10.1038/hdy.2015.104

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, *et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, **454(7205)**:766–770. http://dx.doi.org/10.1038/nature07107

Metcalf JL, Prost S, Nogués-Bravo D, DeChaine EG, Anderson C, Batra P, Araújo MB, Cooper A, & Guralnick RP (2014). Integrating multiple lines of evidence into historical biogeography hypothesis testing: a *Bison bison* case study. *Proc Biol Sci*, **281(1777)**. http://dx.doi.org/10.1098/rspb.2013.2782

Metzger DCH & Schulte PM (2017). Persistent and plastic effects of temperature on DNA methylation across the genome of threespine stickleback (*Gasterosteus aculeatus*). *Proc R Soc B*, **284(1864)**:20171667. http://dx.doi.org/10.1098/rspb.2017.1667

Miller W, Schuster SC, Welch AJ, Ratan A, Bedoya-Reina OC, Zhao F, Kim HL, Burhans RC, Drautz DI, Wittekindt NE, *et al.* (2012). Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proc Natl Acad Sci U S A*, **109(36)**:E2382–E2390. http://dx.doi.org/10.1073/pnas.1210506109

Mittnik A, Wang CC, Svoboda J, & Krause J (2016). A molecular approach to the sexing of the triple burial at the Upper Paleolithic site of Dolní Věstonice. *PLoS One*, **11(10)**:e0163019. http://dx.doi.org/10.1371/journal.pone.0163019

Moorjani P, Gao Z, & Przeworski M (2016a). Human germline mutation and the erratic evolutionary clock. *PLoS Biol*, **14(10)**:e2000744. http://dx.doi.org/10.1371/journal.pbio.2000744

Moorjani P, Sankararaman S, Fu Q, Przeworski M, Patterson N, & Reich D (2016b). A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. *Proc Natl Acad Sci U S A*, **113(20)**:5652–5657. http://dx.doi.org/10.1073/pnas.1514696113

Pečnerová P, Díez-del Molino D, Dussex N, Feuerborn T, von Seth J, van der Plicht J, Nikolskiy P, Tikhonov A, Vartanyan S, & Dalén L (2017). Genome-based sexing provides clues about behavior and social structure in the woolly mammoth. *Curr Biol*, **27(22)**:3505–3510.e3. http://dx.doi.org/10.1016/j.cub.2017.09.064

Pedersen JS, Valen E, Velazquez AMV, Parker BJ, Rasmussen M, Lindgreen S, Lilje B, Tobin DJ, Kelly TK, Vang S, *et al.* (2014). Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res*, **24(3)**:454–466. http://dx.doi.org/10.1101/gr.163592.113

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PLF, *et al.* (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, **468(7327)**:1053–1060. http://dx.doi.org/10.1038/nature09710

Runemark A, Trier CN, Eroukhmanoff F, Hermansen JS, Matschiner M, Ravinet M, Elgvin TO, & Sætre GP (2018). Variation and constraints in hybrid genome formation. *Nat Ecol Evol*, **2(3)**:549–556. http://dx.doi.org/10.1038/s41559-017-0437-7

Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, Patterson N, & Reich D (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, **507(7492)**:354–357. http://dx.doi.org/10.1038/nature12961

Sankararaman S, Mallick S, Patterson N, & Reich D (2016). The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. *Current Biology*, **26(9)**:1241–1247. http://dx.doi.org/10.1016/j.cub.2016.03.037

Scally A & Durbin R (2012). Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet*, **13(10)**:745–753. http://dx.doi.org/10.1038/nrg3295

Schiffels S & Durbin R (2014). Inferring human population size and separation history from multiple genome sequences. *Nat Genet*, **46(8)**:919–925. http://dx.doi.org/10.1038/ng.3015

Schumer M, Xu C, Powell DL, Durvasula A, Skov L, Holland C, Blazier JC, Sankararaman S, Andolfatto P, Rosenthal GG, *et al.* (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, **360(6389)**:656–660. http://dx.doi.org/10.1126/science.aar3684

Seguin-Orlando A, Gamba C, Sarkissian CD, Ermini L, Louvel G, Boulygina E, Sokolov A, Nedoluzhko A, Lorenzen ED, Lopez P, *et al.* (2015). Pros and cons of methylation-based enrichment methods for ancient DNA. *Sci Rep*, **5**:11826. http://dx.doi.org/10.1038/srep11826

Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, Pybus OG, Gilbert MTP, Barnes I, Binladen J, *et al.* (2004). Rise and fall of the Beringian steppe bison. *Science*, **306(5701)**:1561–1565. http://dx.doi.org/10.1126/science.1101074

Skoglund P, Ersmark E, Palkopoulou E, & Dalén L (2015). Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Current Biology*, **25(11)**:1515–1519. http://dx.doi.org/10.1016/j.cub.2015.04.019

Skoglund P, Storå J, Götherström A, & Jakobsson M (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J Archaeol Sci*, **40(12)**:4477–4482. http://dx.doi.org/10.1016/j.jas.2013.07.004

Smith O, Clapham AJ, Rose P, Liu Y, Wang J, & Allaby RG (2014). Genomic methylation patterns in archaeological barley show de-methylation as a time-dependent diagenetic process. *Sci Rep*, **4**:5559. http://dx.doi.org/10.1038/srep05559

Smith RWA, Monroe C, & Bolnick DA (2015). Detection of cytosine methylation in ancient DNA from five Native American populations using bisulfite sequencing. *PLoS One*, **10(5)**:e0125344. http://dx.doi.org/10.1371/journal.pone.0125344

Soubrier J, Gower G, Chen K, Richards SM, Llamas B, Mitchell KJ, Ho SYW, Kosintsev P, Lee MSY, Baryshnikov G, *et al.* (2016). Early cave art and ancient DNA record the origin of European bison. *Nat Commun*, **7**:13158. http://dx.doi.org/10.1038/ncomms13158

Städler T, Haubold B, Merino C, Stephan W, & Pfaffelhuber P (2009). The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. *Genetics*, **182(1)**:205–216. http://dx.doi.org/10.1534/genetics.108.094904

Sun Ma, Velmurugan KR, Keimig D, Xie H, Sun Ma, Velmurugan KR, Keimig D, & Xie H (2015). HBS-tools for hairpin bisulfite sequencing data processing and analysis. *Advances in Bioinformatics*, **2015**:e760423. http://dx.doi.org/10.1155/2015/760423

Titus AJ, Gallimore RM, Salas LA, & Christensen BC (2017). Cell-type deconvolution from DNA methylation: a review of recent applications. *Hum Mol Genet*, **26(R2)**:R216–R224. http://dx.doi.org/10.1093/hmg/ddx275

Węcek K, Hartmann S, Paijmans JLA, Taron U, Xenikoudakis G, Cahill JA, Heintzman PD, Shapiro B, Baryshnikov G, Bunevich AN, *et al.* (2017). Complex admixture preceded and followed the extinction of wisent in the wild. *Mol Biol Evol*. http://dx.doi.org/10.1093/molbev/msw254

Wright S (1931). Evolution in Mendelian populations. *Genetics*, **16(2)**:97–159

Wu DD, Ding XD, Wang S, Wójcik JM, Zhang Y, Tokarska M, Li Y, Wang MS, Faruque O, Nielsen R, *et al.* (2018). Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nat Ecol Evol*, **2(7)**:1139–1145. http://dx.doi.org/10.1038/s41559-018-0562-y

Zhao L, Sun Ma, Li Z, Bai X, Yu M, Wang M, Liang L, Shao X, Arnovitz S, Wang Q, *et al.* (2014). The dynamics of DNA methylation fidelity during mouse embryonic stem cell self-renewal and differentiation. *Genome Res*, p. gr.163147.113. http://dx.doi.org/10.1101/gr.163147.113