The 'next generation' of ancient DNA research: a series of methods and approaches to improve our understanding of the evolutionary history of species in general, and European bison in particular.

Ayla L. van Loenen

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

July 2018

Table of Contents

Thesis Abstract

Understanding the complex and dynamic processes underlying the evolutionary history of species is key for predicting how species evolve and respond to changes in their environment. This is particularly important for the development of appropriate conservation management strategies for species that are currently under threat or predicted to come under threat in the future. While modern DNA may be available, identifying long term patterns of evolutionary history in extant populations can be confounded by more recent population events (such as bottlenecks or local extinctions). Instead, DNA from ancient specimens can be used to directly observe genetic changes through the evolutionary history of a species as well as reconstruct the demographic history of both extinct and extant species or populations. However, the small number of specimens typically included in ancient DNA studies often limits their interpretations to fairly general or location specific conclusions.

In this thesis, I developed approaches for the integration of genetic and non-genetic datasets, specifically combining modern and ancient DNA data, geographic and temporal information, palaeoenvironmental predictions, and historical records to improve our understanding of evolutionary processes. In order to maximise the inclusion of ancient DNA data, I have optimised techniques for sequencing and processing data (Chapters 2 & 3). Collectively these methods significantly improve the recovery of ancient DNA data from larger numbers of poor quality samples. Larger datasets subsequently facilitate the investigation of broad-scale continental patterns of species evolution and response to environmental change. European bison are a particularly good model for studying the responses of megafaunal species to environmental change. Samples are available from both modern and ancient individuals, and European bison have persisted across a broad geographic change throughout periods of dramatic Pleistocene environmental change. In addition, the evolutionary history of European bison remains unclear and relatively few ancient samples have been analysed (Chapter 4). I applied the optimised methods from Chapters 2 & 3 to a large set of European bison samples, significantly increasing the recovery of

ancient genetic data and creating a much larger ancient DNA dataset for subsequent analysis. Using this comprehensive dataset comprising samples from across the temporal and geographic range of the species, I delineated patterns of genetic change through time in both mitochondrial (Chapter 5) and nuclear (Chapter 6) genomes. To identify potential drivers of evolution I combined palaeoenvironmental data (such as palaeoecology and palaeoclimate), and historical records with patterns of genetic change. Using this approach, I found that historical changes in forest cover and anthropogenic impacts were primary drivers of European bison evolution, and shaped their modern diversity and distribution (Chapters 5 & 6). Ultimately the research presented in my thesis contributes towards the 'next-generation' of ancient DNA research, providing a series of approaches to optimise recovery of genetic data from ancient samples, and combining genetic and non-genetic datasets in order to further our understanding of how species evolve and respond to environmental change.

Thesis Declaration

I, Ayla L. van Loenen, 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.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also 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.

....

Ayla L. van Loenen Date

Acknowledgements

Firstly, I would like to thank my supervisors Alan Cooper, Bastien Llamas, and Kieren Mitchell for their scientific critique, support and discussions during my PhD. I am immensely grateful to my collaborators, including: Rafał Kowalczyk, Emilia Hofman-Kamińska, Frédérik Saltré, and Adam "Ben" Rohrlach for their assistance in providing samples, metadata, and participation in valuable discussions about my various projects.

This research was supported by the Polish National Science Centre (2013/11/B/NZ8/00914 and N N304 301940), and through the provision of an Australian Government Research Training Program Scholarship. Attendance at conferences was supported by the Society for Molecular Biology and Evolution and the University of Adelaide.

I would like to thank the members of the Australian Centre of Ancient DNA (ACAD) during my candidature for their support and friendship, and particularly members of the 'Megafaunal' research group for their participation in stimulating discussions. In addition I would like to thank all my friends and colleagues in Australia and abroad for their continued support and friendship. In particular I would like to thank Josie, Steven L., Jo, Tash and Steve M. for their support from near and far throughout my PhD.

Finally I would especially like to thank my parents, Nicoline and Eddo, for always supporting me and encouraging me to follow my dreams, wherever they take me.

Chapter 1

General Introduction

Chapter 1: Introduction

Genetics as a tool for reconstructing evolutionary history

Untangling the complex evolutionary history of species is key for understanding how species evolve and respond to environmental changes. This is particularly important for identifying potential drivers of evolutionary change, and developing conservation management strategies to mitigate species loss due to future environmental changes such as climate change or deforestation (Leonard, 2008). In this thesis I will demonstrate how our understanding of the evolutionary history of species can be significantly improved by applying advanced molecular techniques and novel bioinformatics tools to optimise the recovery of DNA from ancient specimens, and demonstrate the power of integrating ancient DNA data with complementary data types (e.g. specimen metadata, palaeoclimatic reconstructions, historical records). I demonstrate these interdisciplinary approaches in an in-depth study of European bison (*Bison bonasus*), characterising patterns of evolutionary change over time, and identifying potential drivers of those changes.

Previously, evolutionary history was inferred by investigating the fossil record and determining biogeographical patterns of the distribution of species over varying geographical and temporal scales (Nelson, 1969). Relationships between species were defined from morphological characteristics and the geographic and temporal distribution of fossil remains (Donoghue and Yang, 2016). However, the power of inferences about evolutionary processes based solely on palaeontological and biogeographical data is limited (Nelson, 1969). Fossil records are typically incomplete, and can easily be confounded by gaps or by overlapping species that are morphologically indiscernible from each other (e.g. bovid species) (Donoghue and Yang, 2016). As a result, assumptions must be made about hypothetical 'intermediary' species as well as suppositions about their characteristics and constraints (Nelson, 1969; Donoghue and Yang, 2016). This can obscure evolutionary patterns such as population replacement or

population demographic changes, and limits conclusions about patterns of relative evolutionary relationships (Donoghue and Yang, 2016). The advent of Polymerase Chain Reaction (PCR) and Sanger sequencing led to the widespread introduction of genetic data that helped broaden our understanding of evolutionary processes (Brito and Edwards, 2009).

The advent of easily obtainable molecular data led to the formation of a new subdiscipline of biogeography called phylogeography (Avise, 2009). Phylogeography, defined as the study of the "spatial arrangements of genetic lineages, especially within and among closely related species" (Avise, 2009; Futuyma and Agrawal, 2009), allows us to infer the evolutionary history of a species through comparisons of DNA sequences within and among species. DNA sequences are used to construct phylogenetic trees, which can then be used to infer the contemporary and historic forces that resulted in the genetic structure of extant populations (Avise, 2009). Phylogenetic trees can be constructed from a variety of different genetic data, from either the mitochondrial and/or nuclear genomes (Avise, 2009). These trees can then be used to characterise the evolutionary relationships between specimens according to DNA sequence similarity (Bear et al., 2016). Closely related specimens (or species) have high sequence similarity, while less related specimens have low sequence similarity (Bear et al., 2016). The phylogeographic patterns identified in the tree structure can subsequently be used to infer patterns of demographic processes such as population size history or migration (Emerson et al., 2001).

Demographic processes can be inferred via the application of coalescent theory, using phylogenetic data to estimate corresponding population genetic parameters such as effective population size (Emerson et al., 2001). Coalescent theory aims to track genetic variation from a set of samples back to the ancestral 'state', or the most recent common ancestor of the samples, where points of common ancestry represent coalescence events (Fu and Li, 1999; Emerson et al., 2001). The distribution of coalescence events through time among a group of samples is intimately linked to demography. For example, two random individuals within a small population are likely to be more closely related by

chance than two individuals within a large population, and so small populations are characterised by many recent coalescence events. Understanding the patterns and processes of demographic history can provide key insight towards characterising patterns of population change and identifying potential drivers of such (Emerson et al., 2001; Ho and Shapiro, 2011). Phylogenetic analysis of genetic data from extant populations has been proven to be a powerful tool for reconstructing crucial aspects of the evolutionary history of a number of species (Willerslev and Cooper, 2005; Leonard, 2008; Pyron, 2015).

Interpretations of phylogenetic trees based solely on extant/modern populations is complicated by the fact that genetic variation that was present in extinct or historic populations may be absent in extant populations (Weber et al., 2004; Leonard, 2008; Pilot et al., 2014; Chang and Shapiro, 2016). Long-term patterns of evolutionary history that can typically be observed in the DNA of extant populations are usually lost or blurred after bottleneck events (e.g. brown bears or American bison) (Barnes et al., 2002; Shapiro et al., 2004; Chang and Shapiro, 2016). For populations or species that have become extinct (such as mammoth or aurochs), there are no extant specimens available for study. However, there are often subfossil or museum specimens available instead (e.g. bone or teeth). These remains can be analysed using ancient DNA methods in order to reveal genetic change through time, and reconstruct the demographic history of both extinct and extant species or populations (Leonard 2008; Orlando and Cooper, 2014).

History and limitations of ancient DNA research

Since the successful isolation and sequencing of DNA from a 140-year-old quagga skin specimen in 1984, “ancient DNA” has rapidly become a powerful tool for understanding the complex evolutionary history of species (Higuchi et al., 1984; Willerslev and Cooper, 2005; Hagelberg et al., 2015). Ancient DNA extracted from remains of organisms up to hundreds-of-thousands of years old provides a direct window into the past (e.g. Orlando et al. (2013)). Studies including ancient DNA analyses have shown that population histories of a

number of species are much more complex than previous phylogeographic studies (based only on DNA from extant individuals) had initially suggested, with dynamic patterns of replacement, local extinctions, and migrations (Barnes et al., 2002; Shapiro et al., 2004; Willerslev and Cooper, 2005; Hofreiter et al., 2007; Leonard, 2008; Leonard and Wayne, 2008; Kuhn et al., 2010; Haak et al., 2015; Massilani et al., 2016; Soubrier et al., 2016). However, DNA from ancient specimens is typically of relatively poor-quality, meaning that ancient DNA research is much more technically challenging than modern DNA research. As a result, many initial attempts to sequence DNA from ancient specimens were unsuccessful (Hagelberg et al., 2015; Sproul and Maddison, 2017).

Ancient DNA is characterised by extensive post-mortem damage (PMD) over time (including fragmentation and nucleotide degradation), and is typically present in relatively low quantities and often contaminated with more abundant modern DNA (Gilbert et al., 2005, 2007; Leonard, 2008; Rizzi et al., 2012; Sproul and Maddison, 2017). The most commonly observed form of PMD is hydrolytic deamination of cytosine to uracil (subsequently replicated as a thymine during DNA amplification) and observed in sequencing data as a series of C to T and complementary G to A transitions (Hofreiter et al., 2001; Gilbert et al., 2003a, 2007; Orlando et al., 2015). Both PMD and contamination from modern DNA can present as base modifications in sequencing data, which can affect downstream analyses (Krings et al., 1997). For example, haplotype characterisation may be affected by PMD, as potential PMD hotspots may overlap with diagnostic sites (e.g. single nucleotide polymorphisms (SNPs)). In order to reduce the potential effects of contamination and PMD, ancient specimens need to be carefully processed in specially designed 'clean lab' facilities with a number of stringent checks and protocols to ensure authenticity of the results (Cooper and Poinar, 2000; Hofreiter and Shapiro, 2012).

Over the last decade significant advances in both laboratory techniques (including library preparation and hybridisation enrichment) and High-Throughput Sequencing (HTS) technologies have revolutionised the ancient DNA field (Meyer and Kircher, 2010; Hofreiter and Shapiro, 2012; Orlando et

al., 2015; Rohland et al., 2015; Hagelberg et al., 2015; García-García et al., 2016; Llamas et al., 2017; Nieves-Colón et al., 2018). HTS technologies have made it possible to efficiently and accurately sequence billions of DNA molecules at a comparatively low cost per base (Van Dijk et al., 2014; Hofreiter et al., 2015; Orlando et al., 2015). Additionally, with HTS data it is possible to identify and remove cloned sequences from sequence alignments to ensure only unique reads are included in subsequent analysis (known as de-duplication). As a result, it is now possible to characterise PMD in individual sequencing reads as well as confidently identify contaminant reads. Methods have been developed that allow aspects of PMD to be measured and accounted for HTS data (e.g. mapDamage2.0, Jónsson et al. (2013)), but these methods generally rely on summary statistics and ignore the regional context of the observed sequence damage. While summary methods are appropriate for some applications, there has so far been less investigation into understanding the distribution of damage across target genomic regions. Some researchers have suggested that particular regions of the genomes may be especially prone to damage or protected from damage (i.e. hot/cold spots; Gilbert et al. (2003b)), which may be important for haplotype and haplogroup characterisations. This is because potential damage hotspots may overlap with diagnostic SNPs, compromising sample classification and phylogenetic inference (e.g. Neparáczki et al. (2017)). Fortunately, large population-level ancient DNA HTS datasets are now making the study of localised damage patterns possible.

While the development and refinement of HTS technologies have helped to detect and mitigate the effects of damage and contamination, its cost-effectiveness is still limited by the generally low proportion of target DNA molecules in ancient DNA extractions. A popular solution to this limitation in the ancient DNA field is ‘hybridisation enrichment’, which facilitates a relative increase in the proportion of target DNA prior to sequencing, even if that target DNA is significantly degraded (Carpenter et al., 2013; Hofreiter et al., 2015; Haak et al., 2015; Soubrier et al., 2016). Hybridisation enrichment methods use short DNA or RNA baits designed to ‘capture’ the regions of interest (e.g.

species-specific SNPs, or genomic regions) (Horn, 2012; Carpenter et al., 2013). After capturing target DNA fragments matching the regions of interest the baits can be immobilised, allowing any undesirable 'off-target' DNA fragments to be washed away. This protocol reduces the proportion of off-target or contaminant DNA and increases the cost-effectiveness of HTS. As a result, specimens can be analysed that would otherwise not contain high enough proportions of target DNA for sequencing (Horn, 2012; Carpenter et al., 2013; Orlando et al., 2015; Chang and Shapiro, 2016).

With the significant methodological improvements provided by hybridisation enrichment and HTS, the ancient DNA field has shifted towards a 'genomic era', with more studies moving to include whole mitochondrial genomes (WMGs) and nuclear genome data (Orlando et al., 2015; Grealy et al., 2017). Increased cost-effectiveness has made HTS data accessible to more research groups, and expanded the scope of studies using ancient DNA in terms of both number and geographical breadth of samples. These broader-scale studies have uncovered further information about the evolutionary history of species that was cryptic to previous studies (Haak et al., 2015; Schroeder et al., 2015; Llamas et al., 2016; Tobler et al., 2017; Nagle et al., 2017). For example, the study by Haak et al. (2015) compiled 69 ancient European genomes and identified a previously unknown mass migration from the steppe into Central Europe during the late Neolithic. However, most of these studies focussed on human genomes, and many studies of non-human species remain restricted in the geographical and temporal range of specimens included. This limits the potential resolution of their analyses and conclusions, as local evolutionary changes do not always reflect the global patterns and processes of evolutionary history of a species (Witman et al., 2004; Urban et al., 2012; Gilbert and O'Connor, 2013; Hidasi-Neto et al., 2015; Chang and Shapiro, 2016). In order to answer large-scale phylogeographic questions, it is clear that researchers need to generate and analyse large datasets of mitochondrial and nuclear DNA covering the geographical and temporal range of a species. However, these same researchers are now beginning to encounter an

additional problem: increasing professional competition and the increased demand for limited ancient specimens (Makarewicz et al., 2017).

As ancient DNA research grew in popularity over the last three decades, large numbers of specimens were sampled from museum collections across the world. For some species there are numerous specimens available (e.g. American bison, horse, or mammoth), but this is not the case for all species (e.g. European bison, humans, or many birds). Extensive prior sampling over the past 30 years of ancient DNA research, plus the development of improved methodologies has lead to an increased demand for subfossil specimens (Makarewicz et al., 2017). For example, the bones of Denisovans – recently discovered archaic humans - are especially rare (and small) and many of the specimens found to date have already been destructively sampled for analysis (Reich et al., 2010; Meyer et al., 2012; Stringer and Barnes, 2015; Brown et al., 2016). Even when samples can be obtained, researchers must carefully decide what kind of genetic data to generate from finite starting material (e.g. mitochondrial genomes, nuclear SNPs, RADseq, or nuclear genomes). In addition, researchers may wish to retain some sample for other types of analysis (e.g. radiocarbon date, isotope analysis and morphological characterisation). This metadata can provide key contextual information that may help identify the evolutionary patterns and processes driving any observed genetic changes identified by ancient DNA analyses. Consequently, a key aim of ancient DNA researchers is to successfully sequence ancient DNA data from ever decreasing amounts of starting material, while including the generation and collation of relevant specimen metadata (e.g. Korlević et al. (2018)).

Integrating complementary data for more powerful hypothesis testing

Analysing specimen metadata and complementary data types (e.g. palaeoclimatic reconstructions) alongside genetic data allows researchers to explicitly link the evolutionary history as described by genetic changes with potential drivers of evolution (Futuyma and Agrawal, 2009). As an example of how non-genetic datasets can supplement current research projects, recently

Gautier et al. (2016) identified a number of potential genes under differential selection between European bison and bovine lineages. These included genes involved in coat thickness and growth, potentially as an adaptation to environmental change. Gautier et al. (2016) suggested that selection acting on these genes provided evidence that European bison had adapted to cold environments. However, given that the study only included modern bison genomes, it is difficult to determine precisely when these adaptations were selected for (beyond that it occurred after the split between the European bison and bovine lineages). The addition of ancient European bison genomes would provide information about whether these genes were selected for recently (i.e. the observed variation is only found in modern bison) or a long time ago. In addition, by including supplemental non-genetic datasets (such as the palaeoecology or palaeoclimate the ancient bison specimens were found in), we can test such adaptation hypotheses directly. For example, if palaeoclimatic reconstructions support the survival of European bison in cold environments (matching the genetic patterns of selection), the proposed hypothesis that European bison were cold adapted and retained that adaptation through till present day has much stronger support.

The integration of complementary non-genetic datasets is of particular importance in the field of ancient DNA, where there are often limited number of specimens and a limited amount of data that can be generated from those specimens (due to accumulation of DNA damage over time). By supplementing the genetic data with additional specimen metadata and other non-genetic datasets, we can provide much stronger support for any conclusions or hypotheses. Metadata included in previous genetic studies was typically restricted to the location of specimen collection, and estimated specimen ages from archaeological context that have high rates of error. This led to a lack of context for interpretations of evolutionary history of species, and consequently geneticists, morphologists, archaeologists and other research teams often produced contradicting hypotheses that limit progress (Bocherens et al., 2015; Zou and Zhang, 2016). However, as techniques in these associated fields

improve, specific metadata is more often available. These include more accurate sample ages, isotope information related to the diet and environment of the specimen during its lifetime, environmental reconstructions such as climate and ecology (encompassing both spatial and temporal variation), and morphological reconstructions.

Estimated archaeological ages (or associated dates) of specimens typically have high rates of error. With direct dating of specimens, much more accurate measures of specimen age can be obtained, which provides important temporal context for genetic data. There are multiple methods to directly date specimens, including: radiocarbon dating, luminescence dating (e.g. optically stimulated luminescence or OSL dating), and amino acid racemisation (Taylor, 2009; Duller and Wintle, 2012; Duller, 2012; Taylor, 2018). The most commonly used method of direct dating for ancient DNA studies is radiocarbon dating. After death, the amount of radioactive ^{14}C found in a specimen's remains (e.g. bone or tooth fragments) – which begins in equilibrium with reservoir carbon levels – begins to decrease as the ^{14}C undergoes radioactive decay and becomes ^{12}C, such that the ratio of ^{14}C to ^{12}C decreases with increasing sample age (Bowman, 1990; Taylor, 2009; Taylor, 2018). Accelerator mass spectrometry (AMS) is used to measure the ratio of ^{14}C to ^{12}C, and this ratio is used to calculate the 'radiocarbon age' of the specimen, which can be calibrated into calendar years using absolutely dated records of reservoir carbon isotope ratios (Hedges and Gowlett, 1986; Taylor, 2009). While the relatively short half-life of ^{14}C means that radiocarbon dating is generally limited to specimens no more than 50,000 years old (Taylor, 2018), this is usually not a problem for ancient DNA studies, which tend to focus on specimens <50,000 years old (as these are more abundant and generally better preserved). Along with radiocarbon dates, stable isotope information can also be obtained from the specimen using mass spectrometry (Brock et al., 2010).

Stable isotope data can be used to identify characteristics of diet and environment of the specimen during its lifetime (i.e. where it lives and what it was eating), which provides additional context to the genetic data and spatiotemporal distribution of specimens. For example, the ratio of stable

isotopes of carbon and nitrogen in a to their radioactive isotopes specimen ($\delta^{13}C$ to $\delta^{12}C$ or $\delta^{15}N$ to $\delta^{14}N$ respectively) can be used to characterise a diet of C_3 or C_4 plants and aridity. Stable isotope data can also be used to identify periods of poor quality diet (Julien et al., 2012). This can inform our understanding of a species adaptability for dietary change, determining if the species may be restricted to a specific diet which they track across the landscape following habitat changes, or if the species may have some dietary flexibility to remain in a single location and subsist on poor quality diet where necessary (Julien et al., 2012; Craine et al., 2016). The context provided by isotope data can be further supported by environmental and morphological reconstructions.

Environmental reconstructions include both climate and ecological reconstructions, which can be generated either from historic proxy data (e.g. ice cores and pollen records), or from predictive climate models. Palaeoenviromental and palaeoclimatic models are progressively being developed for predictions of ecological and environmental change over broad geographical and temporal scales. For example, the Hadley Centre Coupled Model version 3 (hadcm3) (Gordon et al., 2000; Pope et al., 2000), has been used to infer the locally dominant biome as characterized by plant functional types across the globe, and then classify areas of 'forest' and 'non-forest' vegetation types for a series of palaeoenviromental snapshots from 120kyr to present (Kaplan et al., 2003; Kaplan et al., 2016; Binney et al., 2017). Additionally, prior palaeoclimate simulations produced fragmented time slices separated by thousands of years or even millennia. Recently, Fordham et al. (2017) generated continuous global climate projections covering the last 21,000 years until present. The finer time scale presented in Fordham et al. (2017) allows for the investigation of short-term climate variability as opposed to solely long-term gradual changes, providing a more accurate measure of palaeoclimatic instability and change over time. By using a combination of palaeoclimatic reconstructions and stable isotope data, researchers can provide key environmental context to any observed genetic or spatiotemporal distribution variation. Predictions of environment and

diet ascertained from stable isotope data and environmental reconstructions can be further supported by morphological analysis.

Morphological analysis can provide clues to the form and function of species characteristics. For example, European bison have characteristic jaw shape and tooth structure that suggests that they are adapted for grazing in an open environment (Bocherens et al., 2015). However, morphological data is not always the most accurate predictor of feeding strategies (Merceron et al., 2014; Bocherens et al., 2015). Instead, combining morphological datasets with the additional contextual information about diet and the environment an organism lived in (from stable isotope data and palaeoenvironmental models) is likely to result in more accurate and powerful conclusions. Those conclusions can subsequently be used to answer key questions about the drivers of evolutionary change acting on a species.

In short, metadata can be used to identify a broad range of characteristics about each specimen beyond solely DNA data. Through mass spectrometry both radiocarbon dates and stable isotope ratios can be measured, and given the date and specimen location the genetic data can be placed in a spatiotemporal distribution. The spatiotemporal distribution can be used to identify the particular environment the specimen was likely residing in from palaeoenvironmental models (which can be further supported by the diet predictions from stable isotope data, and morphological reconstructions). When used in combination with a broad spatiotemporal distribution of specimens across a species' range, such datasets could also be compared to global palaeoclimate models to give a broader context to the genetic and environmental patterns observed. In particular, the global changes in climate over thousands of years may drive ecological or isotopic changes that in turn may play a role in driving the genetic changes observed using ancient DNA. It is the combination of these traditionally separate datasets and research fields that is key to developing our understanding of the evolutionary history of species. The collective specimen information and complementary non-genetic data sets can then be used in what has been termed a

'total-evidence' approach, which typically leads to much more accurate representations of the evolutionary history of species (Pyron, 2015).

Individually, each of the datasets mentioned above (genetics, stable isotopes, palaeoenvironmental reconstructions, historical records, etc) is not usually sufficient evidence to identify evolutionary drivers. However, when the clues provided by each dataset are looked at as a whole (like putting together a jigsaw puzzle), they can be used to identify the key evolutionary patterns and processes acting on the species as a whole. Researchers have begun to integrate various datasets, but usually with a limited number of specimens, and broad scale examples are rare. For example, a handful of directly dated specimens and regional predictions of environmental change were included in a recent study of the evolutionary history of European bison to provide context to observed patterns of mtDNA replacement (Soubrier et al., 2016). Due to the geographic range of the bison specimens included, and limits of accurate palaeoenviromental modelling, the environmental reconstruction was restricted to the Urals Mountains (Russia). The continuous refinement of palaeoclimate and palaeoecological models discussed above improves the identification of global patterns of environmental and climatic changes through time. These patterns of environmental change can subsequently be directly linked with concurrent genetic changes in species, facilitating the characterisation of direct relationships between environmental and genetic changes through time. In combination with a much broader geographic and temporal range of specimens across the species range, a more detailed characterisation of the evolutionary history of species (and European bison in particular) could be identified.

With recent advances in molecular methods and HTS, generating genomic DNA data to understand the evolutionary history of species is possible for an increasing number of research groups. Concurrent advances in various related specimen metadata and non-genetic datasets discussed above (including dating, isotope analysis, palaeoenvironmental models), can be used to provide context to the genomic DNA data. Consequently, it is now possible to undertake truly comprehensive continental or global studies of the evolutionary history of a

species across its temporal and geographic range and characterise potential drivers. This approach would be especially useful for species for which there is a wide range of ancient specimens available, but whose evolutionary history remains poorly understood.

European bison (*Bison bonasus*) as a model species for phylogeography

Despite an extensive fossil record throughout Eurasia, the evolutionary history of European bison (also known as wisent or *Bison bonasus*, Linnaeus 1758) remains poorly understood. Specimens show extensive morphological diversity throughout the fossil record, with many distinctive forms appearing contemporaneously with no clear temporal or spatial pattern. In addition, European bison are interesting as they are one of the few Holarctic megafaunal species to have survived the Late Pleistocene/early Holocene megafaunal extinctions, though they then became extinct in the wild during the 1920s and suffered an extreme population bottleneck (54 individuals, all descended from only 12 founders; Pucek, 2004). Such an extreme bottleneck coupled with founder effects likely masks long-term patterns of evolutionary history from DNA analyses of extant specimens (Chang and Shapiro, 2016). Broadly speaking, due to the abundant fossil record, large-scale geographical and ecological distribution of European bison through time, and their apparent dynamic population history (Soubrier et al., 2016; Massilani et al., 2016), they represent an excellent model species to further investigate the application of ancient DNA techniques to understand the evolutionary history of species.

Thesis outline

Summary

Understanding the complex and dynamic patterns and processes underlying the evolutionary history of a species in an environmental context is key to our understanding of how species evolve. This in turn is key for predicting how species may continue to evolve in future, and to develop appropriate conservation management strategies for species that are currently under threat or

predicted to come under threat in the future. Phylogeography, defined as the study of the "spatial arrangements of genetic lineages, especially within and among closely related species", allows us to infer the evolutionary history of a species through comparisons of DNA sequences within and among species (Avise, 2009; Futuyma and Agrawal, 2009). DNA from ancient specimens (ancient DNA) can be used to directly observe genetic changes through the evolutionary history of the species as well as reconstruct the demographic history of both extinct and extant species and populations (Leonard, 2008; Orlando and Cooper, 2014; Chang and Shapiro, 2016). The field of ancient DNA research is rapidly expanding into the 'genomics era', with recent advances in molecular methods and HTS allowing for the generation of increasingly large amounts of ancient DNA data. In combination with concurrent advances in related palaeoenviromental disciplines (e.g. palaeoecology and palaeoclimate), it is now possible to undertake truly comprehensive global/continental scale studies of the evolutionary history of a species across it's former temporal and geographic range, and provide a palaeoenviromental context to the patterns of genetic change observed. This is especially useful for species for which there is a wide range of ancient specimens available, but whose evolutionary history remains poorly understood.

In the subsequent chapters of this thesis I will demonstrate a series of approaches to improve the sequencing and analysis of ancient DNA data. I will then apply these improvements to investigate the dynamic and complex evolutionary history of the European bison (*Bison bonasus*) as a case study demonstrating the power of integrating genetic data with other complementary data types.

Chapter 2: Characterisation of localised post-mortem DNA damage profiles with High-Throughput Sequencing

In this chapter I develop a novel bioinformatics method to characterise post-mortem DNA damage (PMD) in HTS data. I use this method to generate PMD profiles for 32 human specimens sequenced using HTS technology, and compare

these profiles within and between specimens to identify shared characteristics of PMD. The distribution and characterisation of PMD throughout the mitochondrial genome is of key importance for ensuring appropriate filtering and consensus thresholds for aDNA data. This is especially important when performing broad scale analyses that include either a combination of modern and ancient specimens, or specimens with varying levels of PMD (e.g. ancient specimens covering a wide age range). Incorrectly accounting for PMD across such datasets can result in the misidentification of specimens, or confound the results by misinterpreting PMD as 'true' variation between specimens. Previous research into PMD patterns suggested that PMD hotspots and regions protected from PMD could be found in the mitochondrial HVR1 region. However, these results have not been replicated using current high-throughput sequencing (HTS) techniques, and may not accurately reflect potential PMD trends that can be observed in HTS data.

Chapter 3: Improved recovery of ancient DNA from archived extracts

For many ancient specimens extracted over the last few decades, limitations of prior techniques restricted sequencing success to fragments of the mitochondrial HVRs. In addition, many specimens failed to yield any usable genetic data at all. With recent advances in molecular methods and sequencing technologies, many of these specimens may now be amenable to DNA sequencing, expanding the geographic and temporal coverage of specimens included in future evolutionary research. Further, the DNA extracted from many of these specimens still exists in the archives of laboratories around the world. However, there are a number of potential pitfalls with archived DNA extracts: the solvent may evaporate from the tube, or the DNA may have degraded further or adhered to the sides of the tube over time. Consequently, researchers may be hesitant to work with archived extracts, instead opting to re-sample the specimen when available (further damaging the specimen) and perform new DNA extractions. Here I test a number of approaches for optimizing the sequencing success of old DNA extracts, and demonstrate that mitochondrial genomes and nuclear DNA data can be obtained

from archived DNA extracts that have been stored for >10 years without the need for re-sampling and re-extraction.

Chapter 4: Review of the proposed hypotheses for the evolutionary history of European bison (Bison bonasus)

In this chapter I critically review the current published research into the evolutionary history of European bison (*Bison bonasus*), as well as discuss the various alternative hypotheses proposed to explain the evolution of European bison to the present day and identify key limitations in current research. In particular, I address the validity of recent claims surrounding the continued survival of *Bison schoetensaki* into the Late Pleistocene, and suggest expanding the geographic and temporal coverage of the sampling, and including more detailed supplementary non-genetic datasets in order to help improve our understanding. In the subsequent chapters (5 and 6) of my thesis I address a number of the limitations identified as part of this review.

Chapter 5: Untangling the evolutionary history of European bison (Bison bonasus)

Bison are one of the few species to have survived the mass megafaunal extinctions during the Pleistocene/Holocene transition (12-9ky BP), and therefore have the potential to inform us about the response of megafaunal populations to periods of rapid environmental change. Nonetheless, the evolutionary history of the European bison (*Bison bonasus*) remains contentious and cryptic despite several recent genetic studies. Ancient DNA provides a unique opportunity to directly observe genetic evolution by investigating the changes in genetic structure of species and populations in real time. Previous studies with fragmentary data from a limited number of samples from restricted geographic ranges have hinted at a dynamic series of changes through time potentially correlated with environmental change (Soubrier et al., 2016; Massilani et al., 2016). In this chapter I generate high-resolution mitochondrial sequencing data from over 140 ancient European bison samples from across western Eurasia,

revealing geographic and temporal distribution shifts in response to environmental change.

Chapter 6: Low genomic diversity in European bison (Bison bonasus) prior to their near-extinction during the 20th Century

Untangling long-term patterns of evolutionary history is particularly complicated when researching extant populations after extreme bottlenecks. European bison (*Bison bonasus*) were initially found throughout Europe during the Late Pleistocene/early Holocene until suffering an extreme bottleneck during the 1920s, to the point that they became extinct in the wild. The current day population was recovered through a captive breeding program of the 12 remaining European bison. This severe bottleneck was thought to have been the cause of the observed low genetic diversity present in extant European bison. Prior research has focused on historic and modern genomes only, and consequently was unable to accurately characterise the lost during the 1920s bottleneck. In this chapter I present the first ancient European bison nuclear genomes, and describe patterns of genomic diversity (heterozygosity) through time between ancient, historic, and modern European bison.

References

Avise, J. C. (2009). Phylogeography: retrospect and prospect. *Journal of Biogeography*, 36(1):3–15.

Barnes, I., Matheus, P., Shapiro, B., Jensen, D., and Cooper, A. (2002). Dynamics of pleistocene population extinctions in beringian brown bears. *Science*, 295(5563):2267–2270.

Bear, R., Rintoul, D., Snyder, B., Smith-Caldas, M., Herren, C., and Horne, E. (2016). *Principles of Biology*. OpenStax CNX.

Binney, H., Edwards, M., Macias-Fauria, M., Lozhkin, A., Anderson, P., Kaplan, J. O., Andreev, A., Bezrukova, E., Blyakharchuk, T., Jankovska, V., Khazina, I., Krivonogov, S., Kremenetski, K., Nield, J., Novenko, E., Ryabogina, N., Solovieva, N., Willis, K., and Zernitskaya, V. (2017). Vegetation of Eurasia from the last glacial maximum to present: Key biogeographic patterns. *Quaternary Science Reviews*, 157(Supplement C):80–97.

Bocherens, H., Hofman-Kamińska, E., Drucker, D. G., Schmölcke, U., and Kowalczyk, R. (2015). European Bison as a Refugee Species? Evidence from Isotopic Data on Early Holocene Bison and Other Large Herbivores in Northern Europe. *PLOS ONE*, 10(2):e0115090.

Bowman, S. (1990). *Radiocarbon dating*, volume 1. University of California Press.

Brito, P. H. and Edwards, S. V. (2009). Multilocus phylogeography and phylogenetics using sequence based markers. *Genetica*, 135(3):439–455.

Brock, F., Higham, T., Ditchfield, P., and Ramsey, C. B. (2010). Current pretreatment methods for AMS radiocarbon dating at the Oxford Radiocarbon Accelerator Unit (ORAU). *Radiocarbon*, 52(1):103–112.

Brown, S., Higham, T., Slon, V., Pääbo, S., Meyer, M., Douka, K., Brock, F., Comeskey, D., Procopio, N., Shunkov, M., et al. (2016). Identification of a new

hominin bone from Denisova cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Scientific Reports*, 6:23559.

Carpenter, M. L., Buenrostro, J. D., Valdiosera, C., Schroeder, H., Allentoft, M. E., Sikora, M., Rasmussen, M., Gravel, S., Guillén, S., Nekhrizov, G., Leshtakov, K., Dimitrova, D., Theodossiev, N., Pettener, D., Luiselli, D., Sandoval, K., Moreno-Estrada, A., Li, Y., Wang, J., Gilbert, M. T. P., Willerslev, E., Greenleaf, W. J., and Bustamante, C. D. (2013). Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. *American Journal of Human Genetics*, 93(5):852–864.

Chang, D. and Shapiro, B. (2016). Using ancient DNA and coalescent-based methods to infer extinction. *Biology Letters*, 12(2):20150822.

Cooper, A. and Poinar, H. N. (2000). Ancient DNA: do it right or not at all. *Science*, 289(5482):1139–1139.

Craine, J. M., Angerer, J. P., Elmore, A., and Fierer, N. (2016). Continental-Scale Patterns Reveal Potential for Warming-Induced Shifts in Cattle Diet. *PLOS ONE*, 11(8):e0161511.

Donoghue, P. C. and Yang, Z. (2016). The evolution of methods for establishing evolutionary timescales. *Phil. Trans. R. Soc. B*, 371(1699):20160020.

Duller, G. (2012). Improving the accuracy and precision of equivalent doses determined using the optically stimulated luminescence signal from single grains of quartz. *Radiation Measurements*, 47(9):770–777.

Duller, G. A. and Wintle, A. G. (2012). A review of the thermally transferred optically stimulated luminescence signal from quartz for dating sediments. *Quaternary Geochronology*, 7:6–20.

Emerson, B. C., Paradis, E., and Thébaud, C. (2001). Revealing the demographic histories of species using DNA sequences. *Trends in Ecology & Evolution*, 16(12):707–716.

Fordham, D. A., Saltré, F., Haythorne, S., Wigley, T. M., Otto-Bliesner, B. L., Chan, K. C., and Brook, B. W. (2017). Paleoview: a tool for generating continuous climate projections spanning the last 21 000 years at regional and global scales. *Ecography*, 40(11):1348–1358.

Fu, Y.-X. and Li, W.-H. (1999). Coalescing into the 21st century: an overview and prospects of coalescent theory. *Theoretical Population Biology*, 56(1):1–10.

Futuyma, D. J. and Agrawal, A. A. (2009). Evolutionary history and species interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 106(43):18043–18044.

García-García, G., Baux, D., Faugère, V., Moclyn, M., Koenig, M., Claustres, M., and Roux, A.-F. (2016). Assessment of the latest NGS enrichment capture methods in clinical context. *Scientific Reports*, 6.

Gautier, M., Moazami-Goudarzi, K., Levéziel, H., Parinello, H., Grohs, C., Rialle, S., Kowalczyk, R., and Flori, L. (2016). Deciphering the Wisent Demographic and Adaptive Histories from Individual Whole-Genome Sequences. *Molecular Biology and Evolution*, 33(11):2801–2814.

Gilbert, B. and O'Connor, M. I. (2013). Climate change and species interactions: beyond local communities. *Annals of the New York Academy of Sciences*, 1297(1):98–111.

Gilbert, M. T. P., Bandelt, H.-J., Hofreiter, M., and Barnes, I. (2005). Assessing ancient DNA studies. *Trends in Ecology & Evolution*, 20(10):541–544.

Gilbert, M. T. P., Binladen, J., Miller, W., Wiuf, C., Willerslev, E., Poinar, H., Carlson, J. E., Leebens-Mack, J. H., and Schuster, S. C. (2007). Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Research*, 35(1):1–10.

Gilbert, M. T. P., Hansen, A. J., Willerslev, E., Rudbeck, L., Barnes, I., Lynnerup, N., and Cooper, A. (2003a). Characterization of Genetic Miscoding

Lesions Caused by Postmortem Damage. *The American Journal of Human Genetics*, 72(1):48–61.Gilbert, M. T. P., Willerslev, E., Hansen, A. J., Barnes, I., Rudbeck, L., Lynnerup, N., and Cooper, A.

(2003b). Distribution Patterns of Postmortem Damage in Human Mitochondrial DNA. *The American Journal of Human Genetics*, 72(1):32–47.

Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., Mitchell, J. F. B., and Wood, R. A. (2000). The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, 16(2-3):147–168.

Grealy, A., Rawlence, N. J., and Bunce, M. (2017). Time to Spread Your Wings: A Review of the Avian Ancient DNA Field. *Genes*, 8(7):184.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., Fu, Q., Mittnik, A., Bánffy, E., Economou, C., Francken, M., Friederich, S., Pena, R. G., Hallgren, F., Khartanovich, V., Khokhlov, A., Kunst, M., Kuznetsov, P., Meller, H., Mochalov, O., Moiseyev, V., Nicklisch, N., Pichler, S. L., Risch, R., Guerra, M. A. R., Roth, C., Szécsényi-Nagy, A., Wahl, J., Meyer, M., Krause, J., Brown, D., Anthony, D., Cooper, A., Alt, K. W., and Reich, D. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, 522(7555):207–211.

Hagelberg, E., Hofreiter, M., and Keyser, C. (2015). Ancient DNA: the first three decades. *Phil. Trans. R. Soc. B*, 370 (1660): 20130371

Hedges, R. E. and Gowlett, J. A. (1986). Radiocarbon dating by accelerator mass spectrometry. *Scientific American*, 254(1):100–107.

Hidasi-Neto, J., Loyola, R., and Cianciaruso, M. V. (2015). Global and local evolutionary and ecological distinctiveness of terrestrial mammals: identifying priorities across scales. *Diversity and Distributions*, 21(5):548–559.

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A., and Wilson, A. C. (1984). DNA sequences from the quagga, an extinct member of the horse family. *Nature*, 312(5991):282–284.

Ho, S. Y. W. and Shapiro, B. (2011). Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular Ecology Resources*, 11(3):423–434.

Hofreiter, M., Jaenicke, V., Serre, D., Haeseler, A. v., and Pääbo, S. (2001). DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research*, 29(23):4793–4799.

Hofreiter, M., Münzel, S., Conard, N. J., Pollack, J., Slatkin, M., Weiss, G., and Pääbo, S. (2007). Sudden replacement of cave bear mitochondrial DNA in the late Pleistocene. *Current Biology*, 17(4):R122–R123.

Hofreiter, M., Paijmans, J. L. A., Goodchild, H., Speller, C. F., Barlow, A., Fortes, G. G., Thomas, J. A., Ludwig, A., and Collins, M. J. (2015). The future of ancient DNA: Technical advances and conceptual shifts. *BioEssays*, 37(3):284–293.

Hofreiter, M. and Shapiro, B. (2012). *Ancient DNA: Methods and Protocols*. Methods in Molecular Biology. Humana Press Incorporated.

Horn, S. (2012). Target enrichment via DNA hybridization capture. *Methods in Molecular Biology (Clifton, N.J.)*, 840:177–188.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L., and Orlando, L. (2013). mapdamage2. 0: fast approximate bayesian estimates of ancient dna damage parameters. *Bioinformatics*, 29(13):1682–1684.

Julien, M.-A., Bocherens, H., Burke, A., Drucker, D. G., Patou-Mathis, M., Krotova, O., and Péan, S. (2012). Were European steppe bison migratory? ^{18}O, ^{13}C and Sr intra-tooth isotopic variations applied to a palaeoethological reconstruction. *Quaternary International*, 271(Supplement C):106–119.

Kaplan, J. O., Bigelow, N. H., Prentice, I. C., Harrison, S. P., Bartlein, P. J., Christensen, T. R., Cramer, W., Matveyeva, N. V., McGuire, A. D., Murray, D. F., Razzhivin, V. Y., Smith, B., Walker, D. A., Anderson, P. M., Andreev, A. A., Brubaker, L. B., Edwards, M. E., and Lozhkin, A. V. (2003). Climate change and Arctic ecosystems: 2. Modeling, paleodata-model comparisons, and future projections. *Journal of Geophysical Research: Atmospheres*, 108(D19).

Kaplan, J. O., Pfeiffer, M., Kolen, J. C. A., and Davis, B. A. S. (2016). Large Scale Anthropogenic Reduction of Forest Cover in Last Glacial Maximum Europe. *PLOS ONE*, 11(11):e0166726.

Korlevic, P., Talamo, S., and Meyer, M. (2018). A combined method for DNA analysis and radiocarbon dating from a single sample. *Scientific Reports*, 8(1):4127.

Krings, M., Stone, A., Schmitz, R. W., Krainitzki, H., Stoneking, M., and Pääbo, S. (1997). Neandertal DNA sequences and the origin of modern humans. *Cell*, 90(1):19–30.

Kuhn, T. S., Mcfarlane, K. A., Groves, P., Mooers, A. Ø., and Shapiro, B. (2010). Modern and ancient DNA reveal recent partial replacement of caribou in the southwest Yukon. *Molecular Ecology*, 19(7):1312–1323.

Leonard, J. A. and Wayne, R. K. (2008). Native great lakes wolves were not restored. *Biology Letters*, 4(1):95–98.

Leonard, J.A. (2008). Ancient DNA applications for wildlife conservation. *Molecular Ecology*, 17(19):4186–4196.

Llamas, B., Fehren-Schmitz, L., Valverde, G., Soubrier, J., Mallick, S., Rohland, N., Nordenfelt, S., Valdiosera, C., Richards, S. M., Rohrlach, A., Romero, M. I. B., Espinoza, I. F., Cagigao, E. T., Jiménez, L. W., Makowski, K., Reyna, I. S. L., Lory, J. M., Torrez, J. A. B., Rivera, M. A., Burger, R. L., Ceruti, M. C. Reinhard, J., Wells, R. S., Politis, G., Santoro, C. M., Standen, V. G., Smith, C.,

Reich, D., Ho, S. Y. W., Cooper, A., Haak, W. (2016). Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Science Advances*, 2(4):e1501385.

Llamas, B., Valverde, G., Fehren-Schmitz, L., Weyrich, L. S., Cooper, A., and Haak, W. (2017). From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *STAR: Science & Technology of Archaeological Research*, 3(1):1–14.

Makarewicz, C., Marom, N., and Bar-Oz, G. (2017). Palaeobiology: Ensure equal access to ancient DNA. *Nature*, 548(7666):158–158.

Massilani, D., Guimaraes, S., Brugal, J.-P., Bennett, E. A., Tokarska, M., Arbogast, R.-M., Baryshnikov, G., Boeskorov, G., Castel, J.-C., Davydov, S., Madelaine, S., Putelat, O., Spasskaya, N. N., Uerpmann, H.-P., Grange, T., and Geigl, E.-M. (2016). Past climate changes, population dynamics and the origin of Bison in Europe. *BMC Biology*, 14(1):93.

Merceron, G., Hofman-Kamińska, E., and Kowalczyk, R. (2014). 3d dental microwear texture analysis of feeding habits of sympatric ruminants in the Białowieża primeval forest, Poland. *Forest Ecology and Management*, 328:262–269.

Meyer, M. and Kircher, M. (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols*, 2010(6):pdb.prot5448.

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J. G., Jay, F., Prüfer, K., De Filippo, C., et al. (2012). A high-coverage genome sequence from an archaic denisovan individual. *Science*, 338(6104):222–226.

Nagle, N., Oven, M., Wilcox, S., Pellekaan, S. H., Tyler-Smith, C., Xue, Y.,

Ballantyne, K. N., Wilcox, L., Papac, L., Cooke, K., Oorschot, R. A. H., McAllister, P., Williams, L., Kayser, M., Mitchell, R. J., Consortium, T. G., Adhikarla, S., Adler, C. J., Balanovska, E., Balanovsky, O., Bertranpetit, J., Clarke, A. C., Comas, D., Cooper, A., Sarkissian, C. S. I. D., Dulik, M. C., Gaieski, J. B., GaneshPrasad, A., Haak, W., Haber, M., Hobbs, A., Javed, A., Jin, L., Kaplan, M. E., Li, S., Martínez-Cruz, B., Matisoo-Smith, E. A., Melé, M., Merchant, N. C., Owings, A. C., Parida, L., Pitchappan, R., Platt, D. E., Quintana-Murci, L., Renfrew, C., Royyuru, A. K., Santhakumari, A. V., Santos, F. R., Schurr, T. G., Soodyall, H., Hernanz, D. F. S., Swamikrishnan, P., Vilar, M. G., Wells, R. S., Zalloua, P. A., and Ziegle, J. S. (2017). Aboriginal Australian mitochondrial genome variation – an increased understanding of population antiquity and diversity. *Scientific Reports*, 7:43041.

Nelson, G. J. (1969). The problem of historical biogeography. *Systematic Zoology*, 18(2):243–246.

Neparáczki, E., Kocsy, K., Tóth, G. E., Maróti, Z., Kalmár, T., Bihari, P., Nagy, I., Pálfi, G., Molnár, E., Raskó, I., and Török, T. (2017). Revising mtDNA haplotypes of the ancient Hungarian conquerors with next generation sequencing. *PLOS ONE*, 12(4):e0174886.

Nieves-Colón, M. A., Ozga, A. T., Pestle, W. J., Cucina, A., Tiesler, V., Stanton, T. W., and Stone, A. C. (2018). Comparison of two ancient DNA extraction protocols for skeletal remains from tropical environments. *American Journal of Physical Anthropology*.

Orlando, L. and Cooper, A. (2014). Using Ancient DNA to Understand Evolutionary and Ecological Processes. *Annual Review of Ecology, Evolution, and Systematics*, 45(1):573–598.

Orlando, L., Gilbert, M. T. P., and Willerslev, E. (2015). Reconstructing ancient genomes and epigenomes. *Nature Reviews Genetics*, 16(7):395.

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M.,

Schubert, M., Cappellini, E., Petersen, B., Moltke, I., Johnson, P., Fumagalli, M., Vilstrup, J., Raghavan, M., Korneliussen, T., Malaspinas, A.-S., Vogt, J., Szklarczyk, D., Kelstrup, C., Vinther, J., Dolocan, A., Stenderup, J., Velazquez, A., Cahill, J., Rasmussen, M., Wang, X., Min, J., Zazula, G., Seguin-Orlando, A., Mortensen, C., Magnussen, K., Thompson, J., Weinstock, J., Gregersen, K., Røed, K., Eisenmann, V., Rubin, C., Miller, D., Antczak, D., Bertelsen, M., Brunak, S., Al-Rasheid, K., Ryder, O., Andersson, L., Mundy, J., Krogh, A., Gilbert, M., Kjær, K., Sicheritz-Ponten, T., Jensen, L., Olsen, J., Hofreiter, M., Nielsen, R., Shapiro, B., Wang, J., and Willerslev, E. (2013). Recalibrating equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, 499(7456):74–78.

Pilot, M., Greco, C., von Holdt, B. M., Jędrzejewska, B., Randi, E., Jędrzejewski, W., Sidorovich, V. E., Ostrander, E. A., and Wayne, R. K. (2014). Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. *Heredity*, 112(4):428–442.

Pope, V. D., Gallani, M. L., Rowntree, P. R., and Stratton, R. A. (2000). The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3. *Climate Dynamics*, 16(2-3):123–146.

Pucek, Z. (2004). *European Bison (Bison Bonasus): Current State of the Species and Strategy for Its Conservation*. Council of Europe Publishing.

Pyron, R. A. (2015). Post-molecular systematics and the future of phylogenetics. *Trends in Ecology & Evolution*, 30(7):384–389.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L. F., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., Stoneking, M., Richards, M., Talamo, S., Shunkov, M. V., Derevianko, A. P., Hublin, J.-J., Kelso, J., Slatkin, M., and Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, 468(7327):1053–

1060.

Rizzi, E., Lari, M., Gigli, E., De Bellis, G., and Caramelli, D. (2012). Ancient DNA studies: new perspectives on old samples. *Genetics Selection Evolution*, 44:21.

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Phil. Trans. R. Soc. B*, 370(1660):20130624.

Schroeder, H., Ávila-Arcos, M. C., Malaspinas, A.-S., Poznik, G. D., Sandoval-Velasco, M., Carpenter, M. L., Moreno-Mayar, J. V., Sikora, M., Johnson, P. L. F., Allentoft, M. E., Samaniego, J. A., Haviser, J. B., Dee, M. W., Stafford, T. W., Salas, A., Orlando, L., Willerslev, E., Bustamante, C. D., and Gilbert, M. T. P. (2015). Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean. *Proceedings of the National Academy of Sciences of the United States of America*, 112(12):3669–3673.

Shapiro, B., Drummond, A. J., Rambaut, A., Wilson, M. C., Matheus, P. E., Sher, A. V., Pybus, O. G., Gilbert, M. T. P., Barnes, I., Binladen, J., Willerslev, E., Hansen, A. J., Baryshnikov, G. F., Burns, J. A., Davydov, S., Driver, J. C., Froese, D. G., Harington, C. R., Keddie, G., Kosintsev, P., Kunz, M. L., Martin, L. D., Stephenson, R. O., Storer, J., Tedford, R., Zimov, S., and Cooper, A. (2004). Rise and Fall of the Beringian Steppe Bison. *Science*, 306(5701):1561–1565.

Soubrier, J., Gower, G., Chen, K., Richards, S. M., Llamas, B., Mitchell, K. J., Ho, S. Y. W., Kosintsev, P., Lee, M. S. Y., Baryshnikov, G., Bollongino, R., Bover, P., Burger, J., Chivall, D., Crégut-Bonnoure, E., Decker, J. E., Doronichev, V. B., Douka, K., Fordham, D. A., Fontana, F., Fritz, C., Glimmerveen, J., Golovanova, L. V., Groves, C., Guerreschi, A., Haak, W., Higham, T., Hofman-Kamińska, E., Immel, A., Julien, M.-A., Krause, J., Krotova, O., Langbein, F., Larson, G., Rohrlach, A., Scheu, A., Schnabel, R. D.,

Taylor, J. F., Tokarska, M., Tosello, G., van der Plicht, J., van Loenen, A., Vigne, J.-D., Wooley, O., Orlando, L., Kowalczyk, R., Shapiro, B., and Cooper, A. (2016). Early cave art and ancient DNA record the origin of European bison. *Nature Communications*, 7:13158.

Sproul, J. S. and Maddison, D. R. (2017). Sequencing historical specimens: successful preparation of small specimens with low amounts of degraded DNA. *Molecular Ecology Resources*.

Stringer, C. B. and Barnes, I. (2015). Deciphering the denisovans. *Proceedings of the National Academy of Sciences*, 112(51):15542–15543.

Taylor, R. E. (2009). *Encyclopedia of Paleoclimatology and Ancient Environments*. Springer Netherlands. Taylor, R. E. (2018). *Encyclopedia of Global Archaeology*. Springer International Publishing.

Tobler, R., Rohrlach, A., Soubrier, J., Bover, P., Llamas, B., Tuke, J., Bean, N., Abdullah-Highfold, A., Agius, S., O'Donoghue, A., O'Loughlin, I., Sutton, P., Zilio, F., Walshe, K., Williams, A. N., Turney, C. S. M., Williams, M., Richards, S. M., Mitchell, R. J., Kowal, E., Stephen, J. R., Williams, L., Haak, W., and Cooper, A. (2017). Aboriginal mitogenomes reveal 50,000 years of regionalism in Australia. *Nature*, 544(7649):180.

Urban, M. C., De Meester, L., Vellend, M., Stoks, R., and Vanoverbeke, J. (2012). A crucial step toward realism: responses to climate change from an evolving metacommunity perspective. *Evolutionary Applications*, 5(2):154–167.

Van Dijk, E. L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics*, 30(9):418–426.

Weber, D. S., Stewart, B. S., and Lehman, N. (2004). Genetic Consequences of a Severe Population Bottleneck in the Guadalupe Fur Seal (Arctocephalus townsendi). *Journal of Heredity*, 95(2):144–153.

Willerslev, E. and Cooper, A. (2005). Ancient DNA. *Proceedings of the Royal*

Society of London B: Biological Sciences, 272(1558):3–16.

Witman, J. D., Etter, R. J., and Smith, F. (2004). The relationship between regional and local species diversity in marine benthic communities: A global perspective. *Proceedings of the National Academy of Sciences*, 101(44):15664–15669.

Zou, Z. and Zhang, J. (2016). Morphological and molecular convergences in mammalian phylogenetics. *Nature Communications*, 7:12758.

Chapter 2

Characterisation of localised post-mortem DNA damage profiles with high throughput sequencing

Manuscript prepared for submission

Statement of Authorship

Title of Paper	Characterisation of localised post-mortem DNA damage (PMD) profiles with high throughput sequencing.
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☒ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Manuscript prepared in publication format

Principal Author

Name of Principal Author (Candidate)	Ayla L. van Loenen
Contribution to the Paper	Designed experiments, compiled sequencing data and performed bioinformatics analyses. Analysed and interpreted results. Wrote the paper with comments from co-authors.
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 3/7/18

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, interpreted results, made comments on the manuscript		
Signature		Date	3/7/18

Name of Co-Author	Bastien Llamas		
Contribution to the Paper	Provided raw published sequencing data, interpreted results, made comments on the manuscript.		
Signature		Date	3/7/18

Name of Co-Author	Kieren Mitchell		
Contribution to the Paper	Interpreted results, made comments on the manuscript		
Signature		Date	3/7/18

Name of Co-Author	A.B. Rohrlach		
Contribution to the Paper	Designed experiments, performed bioinformatics analyses, interpreted results, made comments on the manuscript		
Signature		Date	12/07/18

Characterisation of localised post-mortem DNA damage profiles with high-throughput sequencing.

Ayla L. van Loenen[1], A.B. Rohrlach[2], Kieren J. Mitchell[1], Bastien Llamas[1], and Alan Cooper[1]

[1]**Australian Centre for Ancient DNA (ACAD), University of Adelaide, Australia**
[2]**School of Mathematical Sciences, University of Adelaide, Australia**

ABSTRACT

Ancient DNA (aDNA) is characterised by extensive post-mortem damage (PMD), typically observed as C to T transitions in sequencing data. Characterising and monitoring the distribution of PMD throughout the mitochondrial genome is of key importance for ensuring appropriate quality filtering and consensus thresholds for aDNA analyses. This is especially important when performing analyses of datasets that include either a combination of modern and ancient specimens, or other specimens with varying levels of PMD (e.g. ancient specimens covering a wide age range). Failure to account for PMD across such datasets can result in the misattribution of specimens or confound estimates of genetic divergence by misinterpreting PMD as 'true' variation between specimens. Previous research into PMD patterns in the mitochondrial HVR1 region suggested that both hotspots and regions protected from PMD could be found. However, these results have not been replicated using high-throughput sequencing techniques which provide far greater sequencing depth and resolution. Here we present a novel method for characterising the distribution of PMD across the genome using HTS data and use this method to demonstrate that PMD appears to occurs randomly across the mitochondrial genome (with no positional biases, hotspots or regions protected from PMD shared between specimens). This method can also be used to identify appropriate thresholds for consensus sequence calling in order to ensure accurate SNP calling for downstream analyses.

Keywords: Ancient DNA (aDNA), Post-mortem DNA damage (PMD)

INTRODUCTION

Ancient DNA (aDNA) is characterized by extensive post-mortem DNA damage (PMD), including DNA strand breaks, oxidative nucleotide modification, depurination, and hydrolytic deamination (Dabney et al., 2013; Gilbert et al., 2003a; Pääbo, 1989). Some of these damage types may cause an incorrect nucleotide to be incorporated during *in vitro* DNA amplification (such as PCR—Polymerase Chain Reaction), presenting as base modifications in sequencing data (Krings et al., 1997). While methods have been developed that allow aspects of PMD to be measured and accounted for in High-Throughput Sequencing (HTS) data (e.g. mapDamage2.0; Jónsson et al. (2013)), these methods generally rely on summary statistics and ignore the regional context of the observed sequence damage. However, the distribution of PMD across the target genomic region may be important, because the characterisation of any potential hotspots or protected regions is key to understanding how PMD may affect the results of downstream analysis (Gilbert et al., 2003b). In particular, haplotype characterisation may be affected by PMD, as potential PMD hotspots may overlap with diagnostic sites (e.g. single nucleotide polymorphisms (SNPs)). This is particularly important for closely related sequences such as human mitochondrial haplotypes which are sometimes defined on the basis of a single SNP (van Oven and Kayser, 2009). Understanding the distribution of potential PMD hotspots or regions protected from PMD is key to ensuring appropriate filtering for downstream analysis (especially when combining data from modern and ancient specimens).

The most commonly observed form of PMD is the hydrolytic deamination of cytosine to uracil (Hofreiter et al., 2001; Gilbert et al., 2003a, 2007; Dabney et al., 2013; Orlando et al., 2015). Uracil is subsequently replicated as a thymine during *in vitro* DNA amplification, and PMD is consequently observed in sequencing data as a series to C to T (and complementary G to A) transitions (Hofreiter et al., 2001; Gilbert et al., 2003a, 2007; Dabney et al., 2013; Orlando et al., 2015). In addition, these C to T transitions are more commonly found

towards the ends of DNA fragment rather than the middle, and have been shown to accumulate much faster in single-stranded DNA than in double-stranded DNA (Gilbert et al., 2007; Orlando et al., 2015). This suggests that most PMD occurs in single-stranded overhangs at the ends of ancient DNA fragments, formed after nicks in the DNA backbone cause fragmentation of the DNA (Dabney et al., 2013; Orlando et al., 2011). As a result, the presence of cytosine deamination is commonly used to confirm that DNA reads are truly ancient, and not modern contamination (Rohland et al., 2015).

Characterising patterns of PMD is key to ensuring the generation of accurate DNA data from ancient or degraded specimens, and to understanding any potential biases that may be introduced into downstream analyses. This is especially important given the advent of HTS technologies, which allow sequencing data to be obtained from more poorly preserved specimens (that may exhibit more severe PMD) compared to traditional PCR amplification and Sanger sequencing. In addition, when small fragments were routinely amplified through PCR and Sanger sequenced it was relatively simple to check sample alignments by eye (and visually check the results of downstream analyses). However, with the large quantities of data generated by HTS technologies, visual inspection of all the reads per specimen (particularly when sequencing full nuclear genomes or large scale SNP panels) rapidly becomes impossible. A variety of methods are now used to analyse large-scale genomic sequencing data and collectively generate consensus sequences or variant calls for downstream analyses (e.g. Schubert et al. (2014); Renaud et al. (2015); Poplin et al. (2017)). These consensus sequences and variant calls are not typically checked by eye, as the data sets involved are too large for this to be efficient. Instead, various library preparation methods for HTS are often used to minimise the introduction of potential biases from PMD, including full or partial PMD repair methods (Meyer and Kircher, 2010; Briggs et al., 2010; Rohland et al., 2015). However, these methods mostly involve trimming single-stranded overhangs from molecules, resulting in the shortening of fragments and loss of information. If PMD was better understood and could be more accurately accounted for, these methods may not be necessary and more information may be extracted from ancient molecules.

PMD patterns were previously investigated by Gilbert et al. (2003b) using pooled Sanger sequencing data from PCR clones for regions of the human mitochondrial genome (HVR1 and COIII) from 37 archaeological human specimens. Gilbert et al. (2003b) identified potential PMD 'cold-spots' (indicating regions that did not show any PMD) and 'hot-spots' (regions that were typically more frequently subject to PMD than others), suggesting that both PMD hotspots and regions protected from PMD exist in human mitochondrial HVR1. This may suggest that diagnostic SNPs found in or near PMD hotspots are more likely to be incorrectly called than diagnostics SNPs in or near a region protected from PMD, which would introduce biases into downstream analyses. However, the relative rate for PMD calculated in Gilbert et al. (2003b) includes all (SNPs) from the consensus reference sequence. Consequently their estimation of PMD would have included 'true' SNP variation between the 37 archaeological human specimens as well as potential sequencing errors in the calculation of PMD rate per site, which could significantly affect the resulting patterns of PMD. Subsequent research has identified Type 2 transitions (C to T and G to A) as the typical signature of PMD, with Type 1 transitions (T to C and A to G) and transversions more frequently identified as the result of sequencing errors instead of PMD (Hofreiter et al., 2001; Binladen et al., 2006; Gilbert et al., 2007; Orlando et al., 2015). Due to the limitations of Sanger sequencing from clones, previous research by Gilbert et al. (2003b) was also restricted to investigating a single pooled set of specimens for a small region of the mitochondrial genome and were unable to distinguish cloned sequences originating from the same template molecule. With HTS technologies, it is now possible to characterise PMD patterns across the mitochondrial genome within individual specimens and identify molecules representing independent template molecules to ensure only these unique molecules are included in subsequent analyses (known as de-duplication). Importantly, de-duplication ensures accuracy of proportion estimates between damaged positions and undamaged positions.

Here we develop a novel bioinformatics method to characterise site-specific PMD profiles from HTS data, and use this method to identify PMD patterns in 32 human mitochondrial genomes. In direct contrast to previous research findings from Sanger sequencing data, we demonstrate that PMD occurs randomly throughout the mitochondrial genome, with no regions that are either 'protected' or especially prone to PMD, and no correlation with phylogenetically informative sites. Further, this method can be used to characterise PMD in specimens and identify potential thresholds to generate accurate consensus sequences from ancient specimens.

METHODS AND MATERIALS

Samples

We downloaded BAM files for 171 previously published human mitochondrial genome sequences from Llamas et al. (2016) and Schuenemann et al. (2017). The libraries used to produce these BAM files were prepared using two protocols:

- **No DNA repair.** The libraries were prepared according to Meyer and Kircher (2010), with no DNA repair treatment.

- **Partial-UDG repair.** USER treatment excises uracil bases (UDG activity) derived from hydrolytic deamination of cytosines, and cleaves on both sides of the resulting abasic sites (Endo VIII activity) (Briggs et al., 2010). A modification of this treatment, termed 'partial UDG treatment' involves inhibiting UDG through the addition of UGI (UDG inhibitor, New England Biolabs) (Rohland et al., 2015) to ensure the majority of the sequencing read no longer contains uracil bases, but the most terminal bases remain unrepaired. This allows sequencing artefacts caused by PMD to be reduced, while also permitting authentication of aDNA fragments by examining the damage profiles of the terminal bases of each fragment (Rohland et al., 2015; Hofreiter et al., 2001; Pääbo, 1989).

171 specimens had libraries prepared according to the no DNA repair protocol, and eight of those specimens had libraries prepared using both treatments, allowing us to compare the impacts of partial UDG treatment of PMD patterns.

Defining PMD across the mitochondrial genome

We defined read depth at a given site i as:

$$Depth_i = \sum A_i + T_i + C_i + G_i$$

From each BAM file, the number of nucleotides for each site i and the total depth of reads at each site were collated to calculate the proportion of each base at each site i in the mitochondrial genome per library. Any insertions or deletions were excluded from downstream analysis, as these are likely to be sequencing error rather than PMD. For comparison of PMD on 'heavy' and 'light' strands of the human mitochondrial genome, BAM files were sorted for reads mapping to either strand using samtools v1.3.1 (Li, 2011; Li et al., 2009), and then proportions of each nucleotide per site were generated as described above.

An estimate of the relative rate of PMD per site was calculated as follows: for each site, the number of A and T bases (putative 'damage' observations from a C to T and complementary G to A transitions) were added together and the total was divided by the total read depth at that site providing a proportion $\%AT_i$ per

$$\%AT_i = \frac{\sum A_i + T_i}{Depth_i}$$

Sites where $\%AT_i$ was greater than 0.75 were considered true A/T observations to be included in the consensus sequence, and these sites were excluded from downstream analysis. A threshold of 75% was chosen to be relatively conservative. Sites where $\%AT_i = 0$ were considered true C/G observations to be included in the consensus sequence, and these sites were excluded from downstream analysis. PMD proportions were calculated as described above for all sites in all specimens individually and for a combined sum of all sites from all specimens used (i.e. the PMD proportions were calculated for each sample individually, and for the combination of all specimens summed together). This was done in order to look for patterns of PMD across multiple specimens as well as each individual sample.

Validation of PMD profiling method

Predictions of 'true SNP' or 'damage' were made from the PMD profile of the untreated library, and compared to the observed PMD profile of the partial-UDG treated library. Sites with a PMD proportion less than 0.75 and greater than 0 were predicted to be 'damage' and sites with a PMD proportion greater than 0.75 or equal to 0 were predicted to be 'true SNPs'. Predictions were compared to the sequencing results from the matching partial UDG treated library for each specimen to ascertain accuracy of the prediction, and consequently validate the method used to profile PMD.

Correlation between PMD proportion and depth

The proportion of PMD per site as characterised by methods developed in this study was significantly correlated with low sequencing depth in a number of specimens ($p<0.05$). Consequently a minimum depth requirement of 20 reads per position was chosen in order to reduce the introduction of potential biases into the calculation of PMD proportions due to variation in sequencing coverage between specimens. With a minimum of 20 reads per position, the effect of a single stochastic observation of PMD is reduced to $\sim$0.05% and consequently unlikely to significantly affect the overall PMD proportion at the site. Of the original 171 mitochondrial genomes sequenced between the Llamas et al. (2016) and Schuenemann et al. (2017) papers, 139 failed the minimum depth requirement of less than 50 sites with less than 20 reads per site, leaving 32 specimens with sufficient depth for further investigation (Table 1).

Profiling PMD proportions across the mitochondrial genome

Global patterns of PMD across the mitochondrial genome were investigated using a mixture model of k exponential distributions. Let site i, have sequencing depth N_i, of which X_i were damaged. Let the proportion of damaged reads be

$$Y_i = \frac{X_i}{N_i}.$$

Quantile plots indicate that for $Y_i > 0$, the data appears to follow a gamma distribution. To investigate if there was a mixture of two gamma distributions, each sample had non-zero damage site hard classified to a class of "low" and "high" damage via an EM-algorithm using the `mixtools` package (Benaglia et al., 2009) in `R` (R Development Core Team, 2017). For each position i, we find

$$\boldsymbol{n}_i = \left(n_{1,i}, n_{2,i}\right),$$

where $n_{1,i}$ and $n_{2,i}$ are the number of times that site i was hard classified as "low" and "high" damage, respectively. To test if the assignments of site to low and high were equally likely (i.e. whether there were no true low and high classes) we assumed a null distribution of

$$\boldsymbol{n}_i \sim Bin(N_i, 1/2).$$

This is the classical binomial test (Clopper and Pearson, 1934), and we obtain a p-value for each site, which we adjust using the Holm-Bonferroni method (Holm, 1979), to account for multiple hypotheses. Using an adjusted p-value of $\alpha = 0.05$, the only site classified as damage for which the null hypothesis could be rejected was site 4137, indicating no global patterns of PMD.

The proportions of PMD were compared between specimens, annotated gene regions, codons, and phylogenetically informative sites within specimens using a Wilcox Signed Rank test and between specimens using a χ^2 test of independence within the `stats` package in `R` (R Development Core Team, 2017). Gene regions, codons, and sites used to define haplogroups (i.e. phylogenetically informative sites) were defined as per the annotated RSRS human mitochondrial reference and the PhyloTree tree build 17 (Table 2) (van Oven and Kayser, 2009; Behar et al., 2012).

RESULTS

Validation of PMD profiling method

Predictions of 'true SNP' or 'damage' positions from each PMD profile were compared to the PMD profile generated from the sequencing results of matching partial-UDG treated libraries in order to validate the PMD profiling method described above. At all positions with a minimum depth of 20 reads, the distribution of 'true SNP' or 'damage' as predicted by the PMD profiling method developed in this study was accurate for all specimens when compared to the distribution of 'true SNPs' observed in the matching partial-UDG treated library for the specimen (sixteen sequenced libraries total; eight untreated libraries and eight corresponding partial-UDG treated libraries).

No Heavy or Light strand bias towards PMD

Both the individual and the summed specimen PMD profiles demonstrated no strand bias. That is, neither the heavy nor the light strand was significantly more damaged than the other. This holds true both globally across the mitochondrial genome, for each individual gene region, and for codons within the gene regions (Wilcox

SR test, $p>0.05$ for all comparisons). Further, the distribution of PMD on one strand is well predicted by the distribution of PMD on the other strand (i.e. heavy strand light strand, and light strand heavy strand, $p<0.05$ for all specimens). Consequently reads mapping to both strands were combined in all subsequent analyses.

PMD occurs randomly across the mitochondrial genome

PMD profiles showed no evidence for global PMD hot-spots or cold-spots across the mitochondrial genome. There were no sites across the genome that are always subject to PMD, nor regions that are never subject to PMD. Additionally, there are no significant differences in the proportion of PMD sites across the mitochondrial genome, between gene regions, codon positions, or at phylogenetically informative sites (SNP positions used to define haplogroups) (χ^2, $p>0.05$ for all comparisons). However, comparing the proportion of PMD positions between different specimens showed that each sample has a unique proportion of PMD positions, likely relative to the level of PMD exposure unique to each individual sample (Table 3). PMD profiles are also unique to each individual sample, and the distribution of PMD in one sample cannot be used to predict the distribution in another sample (χ^2, $p>0.05$ for all comparisons).

Partial UDG treatment

The proportion of PMD sites across the mitochondrial genome was compared across gene regions and codon positions between untreated and partial UDG treated libraries of the same specimen (e.g. Figure 1). As per the previous results from untreated libraries, there were no gene region or codon biases observed across partial UDG treated libraries (Wilcox SR test, $p>0.05$ for all comparisons). In addition, a grouped comparison of all PMD counts for gene regions and codon positions across all partial UDG treated libraries showed no significant difference in the proportion of PMD observed between specimens ((χ^2 $p>0.05$), in contrast to the observed highly significant variation in proportion of PMD observed between all untreated libraries of the same specimens (χ^2 $p<0.05$).

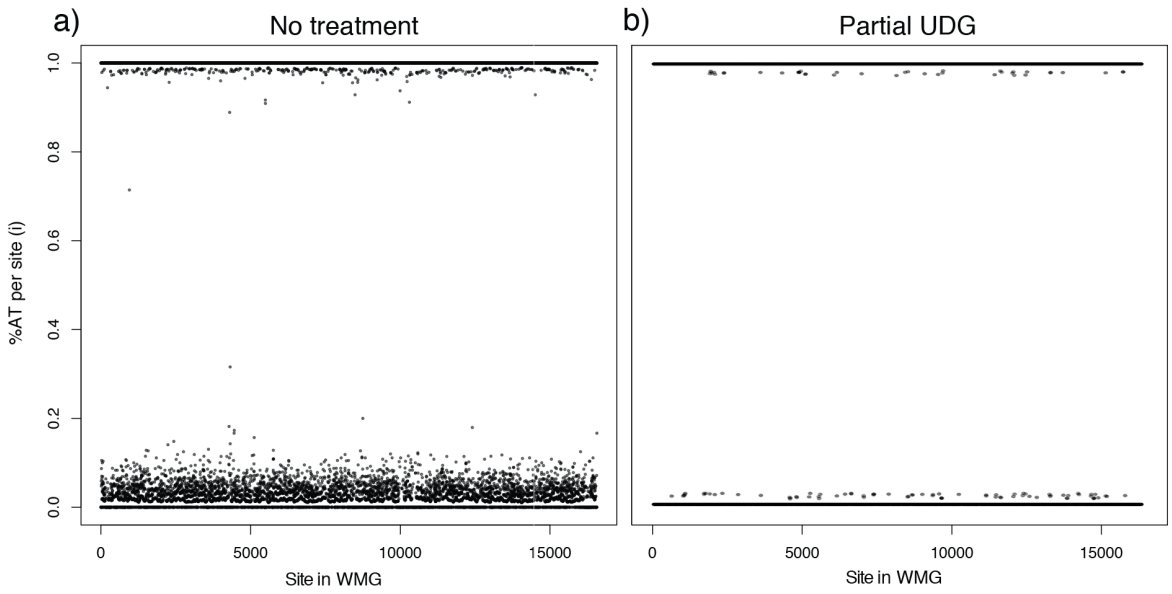


Figure 1. An example of the PMD distribution (and clear reduction in sites with PMD) between a library prepared with no treatment (a) or partial UDG treatment (b) for the same specimen (A10725)

DISCUSSION

In this study we developed a novel bioinformatics method to characterise PMD at the individual nucleotide level using HTS data. In contrast to Sanger sequencing data, with HTS data we were able to accurately characterise the proportion of PMD at each site within the mitochondrial genome and predict whether sites should be considered 'true SNPs' or 'damaged'. With this method we demonstrated that PMD appears to occur

randomly throughout the mitochondrial genome, and that there are no regions that are consistently 'protected' from PMD, in contrast to previous Sanger sequencing studies of mitochondrial DNA (Gilbert et al., 2003b). However, those Sanger sequencing data were generated from a series of pooled clones, whereas with HTS data "cloned" reads are de-duplicated and only the unique reads from each individual specimen are analysed. Therefore, the results from Gilbert et al. (2003b) may be due to a high number of clones representing only a handful of 'original' ancient DNA molecules leading to overestimation of the actual sequencing depth at each site. As we demonstrated during method validation, decreased depth is shown to increase the stochasticity of PMD profiles (and consequently results in inaccurate predictions of PMD).

Our results highlight the limitations of Sanger sequencing to distinguish between 'true SNPs' and 'damaged' sites from cloned DNA molecules. This is further supported by a recent re-analysis of mitochondrial haplogroup calls from a collection of ancient Hungarian specimens (Neparáczki et al., 2017). The original haplogroup calls of three specimens from Tömöry et al. (2007) based on Sanger sequencing of clones were shown to be incorrect when re-analysed with HTS data (Neparáczki et al., 2017). As a result, inferences of haplogroup assignment based on Sanger sequencing should be treated with caution (and confirmed with HTS data where possible). This is especially crucial for species where haplogroups can be assigned from a single SNP (like humans).

Highly damaged specimens with low sequencing depth can be biased towards incorrect calls of PMD as 'true SNPs' in the consensus sequence, causing biases in downstream phylogenetic analyses (especially when specimens with varying levels of PMD are included). For example, ancient specimens with high levels of PMD would appear to be much more distantly related to modern specimens than they actually are (increased PMD resulting in more SNPs between specimens). By characterising the proportions of PMD across the genome using our method, researchers can identify appropriate thresholds for consensus calling for each specimen. For example, it may be appropriate to use a lower or higher consensus threshold instead of 0.75 defined from the specimen specific PMD profile in order to reduce the number of ambiguous base calls due to PMD. This approach can also be used to identify potential sites subject to ambiguous base calls (that are likely a result of contamination or PMD). For example, if the consensus threshold is 0.75 then positions with a proportion of PMD between 0.25 and 0.75 will be ambiguously called for that specimen and may benefit from further checking by eye.

Ultimately, we have demonstrated that PMD is randomly distributed across the mitochondrial genome. This means that (at least for mitochondrial data) a model that takes the position of individual nucleotides into account will not improve our ability to account for PMD. Thus, researchers are left with two options: a) ensure sufficiently deep sequencing of the sample to remove stochastic biases towards potential PMD (i.e. identify an appropriate minimum depth to call consensus for each specimen), or b) repair the extract in order to excise the majority of PMD bases from ancient DNA molecules before sequencing. We have shown that partial UDG treatment significantly reduced the global highly variable proportion of PMD unique to each specimen, while still allowing ancient DNA authentication. This in turn reduces potential for biases to be introduced when comparing a single highly damaged specimen with a group of undamaged or low-damage specimens. In order to further reduce the number of PMD bases in downstream analyses of libraries repaired with the partial UDG treatment, the terminal bases from each sequencing read (i.e. the bases likely to remain damaged) can be clipped during preliminary read processing. This makes treatment with the partial UDG method the best approach for mitochondrial studies where obtaining high read depth is not possible (due either to cost or limitations in the number of template molecules).

REFERENCES

Behar, D. M., Van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.-L., Silva, N. M., Kivisild, T., Torroni, A., and Villems, R. (2012). A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *The American Journal of Human Genetics*, 90(4):675–684.

Benaglia, T., Chauveau, D., Hunter, D., and Young, D. (2009). mixtools: An r package for analyzing finite mixture models. *Journal of Statistical Software*, 32(6):1–29.

Binladen, J., Wiuf, C., Gilbert, M. T. P., Bunce, M., Barnett, R., Larson, G., Greenwood, A. D., Haile, J., Ho, S. Y., Hansen, A. J., et al. (2006). Assessing the fidelity of ancient DNA sequences amplified from nuclear genes. *Genetics*, 172(2):733–741.

Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., and Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Research*, 38(6):e87.

Clopper, C. J. and Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413.

Dabney, J., Meyer, M., and Pääbo, S. (2013). Ancient DNA damage. *Cold Spring Harbor Perspectives in Biology*, page a012567.

Gilbert, M. T. P., Hansen, A. J., Willerslev, E., Rudbeck, L., Barnes, I., Lynnerup, N., and Cooper, A. (2003a). Characterization of Genetic Miscoding Lesions Caused by Postmortem Damage. *The American Journal of Human Genetics*, 72(1):48–61.

Gilbert, M. T. P., Tomsho, L. P., Rendulic, S., Packard, M., Drautz, D. I., Sher, A., Tikhonov, A., Dalén, L., Kuznetsova, T., Kosintsev, P., Campos, P. F., Higham, T., Collins, M. J., Wilson, A. S., Shidlovskiy, F., Buigues, B., Ericson, P. G. P., Germonpré, M., Götherström, A., Iacumin, P., Nikolaev, V., Nowak-Kemp, M., Willerslev, E., Knight, J. R., Irzyk, G. P., Perbost, C. S., Fredrikson, K. M., Harkins, T. T., Sheridan, S., Miller, W., and Schuster, S. C. (2007). Whole-Genome Shotgun Sequencing of Mitochondria from Ancient Hair Shafts. *Science*, 317(5846):1927–1930.

Gilbert, M. T. P., Willerslev, E., Hansen, A. J., Barnes, I., Rudbeck, L., Lynnerup, N., and Cooper, A. (2003b). Distribution Patterns of Postmortem Damage in Human Mitochondrial DNA. *The American Journal of Human Genetics*, 72(1):32–47.

Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M., and Pääbo, S. (2001). Ancient DNA. *Nature Reviews Genetics*, 2(5):353.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L., and Orlando, L. (2013). mapdamage2. 0: fast approximate bayesian estimates of ancient dna damage parameters. *Bioinformatics*, 29(13):1682–1684.

Krings, M., Stone, A., Schmitz, R. W., Krainitzki, H., Stoneking, M., and Pääbo, S. (1997). Neandertal dna sequences and the origin of modern humans. *Cell*, 90(1):19–30.

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.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

Llamas, B., Fehren-Schmitz, L., Valverde, G., Soubrier, J., Mallick, S., Rohland, N., Nordenfelt, S., Valdiosera, C., Richards, S. M., Rohrlach, A., et al. (2016). Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Science Advances*, 2(4):e1501385.

Meyer, M. and Kircher, M. (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols*, 2010(6):pdb.prot5448.

Neparáczki, E., Kocsy, K., Tóth, G. E., Maróti, Z., Kalmár, T., Bihari, P., Nagy, I., Pálfi, G., Molnár, E., Raskó, I., et al. (2017). Revising mtDNA haplotypes of the ancient Hungarian conquerors with next generation sequencing. *PlOS ONE*, 12(4):e0174886.

Orlando, L., Gilbert, M. T. P., and Willerslev, E. (2015). Reconstructing ancient genomes and epigenomes. *Nature Reviews Genetics*, 16(7):395.

Orlando, L., Ginolhac, A., Raghavan, M., Vilstrup, J., Rasmussen, M., Magnussen, K., Steinmann, K. E., Kapranov, P., Thompson, J. F., Zazula, G., Froese, D., Moltke, I., Shapiro, B., Hofreiter, M., Al-Rasheid, K. A. S., Gilbert, M. T. P., and Willerslev, E. (2011). True single-molecule DNA sequencing of a Pleistocene horse bone. *Genome Research*, 21(10):1705–1719.

Pääbo, S. (1989). Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences*, 86(6):1939–1943.

Poplin, R., Newburger, D., Dijamco, J., Nguyen, N., Loy, D., Gross, S. S., McLean, C. Y., and DePristo, M. A. (2017). Creating a universal SNP and small indel variant caller with deep neural networks. *BioRxiv*, page 092890.

R Development Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Renaud, G., Slon, V., Duggan, A. T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biology*, 16(1):224.

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 370(1660):20130624.

Schubert, M., Ermini, L., Sarkissian, C. D., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M. D., Fernández, R., Kircher, M., McCue, M., Willerslev, E., and Orlando, L. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature Protocols*,

9(5):1056–1082.

Schuenemann, V. J., Peltzer, A., Welte, B., van Pelt, W. P., Molak, M., Wang, C.-C., Furtwängler, A., Urban, C., Reiter, E., Nieselt, K., et al. (2017). Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. *Nature Communications*, 8:15694.

Tömöry, G., Csányi, B., Bogácsi-Szabó, E., Kalmár, T., Czibula, Á., Csősz, A., Priskin, K., Mende, B., Langó, P., Downes, C. S., et al. (2007). Comparison of maternal lineage and biogeographic analyses of ancient and modern hungarian populations. *American Journal of Physical Anthropology*, 134(3):354–368.

van Oven, M. and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, 30(2):E386–E394.

Table 1. Sample information

Sample ID	GenBank ID	Source	DNA repair	Country	Haplogroup
A10709	KU523313	Llamas et al., 2016	None	Peru	C1b
A10717	KU523314	Llamas et al., 2016	None	Peru	C1b
A10720	KU523277	Llamas et al., 2016	None	Peru	B2b
A10722	KU523342	Llamas et al., 2016	None	Peru	D1
A10725	KU523315	Llamas et al., 2016	None	Peru	C1b
A10731	KU523316	Llamas et al., 2016	None	Peru	C1b
A10732	KU523343	Llamas et al., 2016	None	Peru	D*
A10789	KU523283	Llamas et al., 2016	None	Peru	B2b
A10791	KU523266	Llamas et al., 2016	None	Peru	A2
A10800	KU523285	Llamas et al., 2016	None	Peru	B2b
A10805	KU523344	Llamas et al., 2016	None	Peru	D1
A10814	KU523287	Llamas et al., 2016	None	Peru	B2
A10817	KU523324	Llamas et al., 2016	None	Peru	C1b
A11116	KU523289	Llamas et al., 2016	None	Chile	B2
A11121	KU523291	Llamas et al., 2016	None	Chile	B2
A11165	KU523292	Llamas et al., 2016	None	Peru	B2
A11168	KU523293	Llamas et al., 2016	None	Peru	B2
A11176	KU523295	Llamas et al., 2016	None	Peru	B2b
A11205	KU523330	Llamas et al., 2016	None	Peru	C1b
A11209	KU523331	Llamas et al., 2016	None	Peru	C1c
A13237	KU523299	Llamas et al., 2016	None	Bolivia	B2
A13272	KU523302	Llamas et al., 2016	None	Bolivia	B2
A13989	KU523304	Llamas et al., 2016	None	Mexico	B2
JK2127	SAMEA4461502	Schuenemann et al., 2017	None	Egypt	W6
JK2130	SAMEA4461504	Schuenemann et al., 2017	None	Egypt	M1a1
JK2131	SAMEA4461505	Schuenemann et al., 2017	None	Egypt	U3b
JK2133	SAMEA4461507	Schuenemann et al., 2017	None	Egypt	X
JK2134	SAMEA4461508	Schuenemann et al., 2017	None	Egypt	J1d
JK2136	SAMEA4461510	Schuenemann et al., 2017	None	Egypt	R0a2
JK2139	SAMEA4461512	Schuenemann et al., 2017	None	Egypt	K1a
JK2866	SAMEA4461522	Schuenemann et al., 2017	None	Egypt	R0a2
JK2879	SAMEA4461530	Schuenemann et al., 2017	None	Egypt	T1a7
A10709*	KU523313	Llamas et al., 2016	partial UDG	Peru	C1b
A10722*	KU523342	Llamas et al., 2016	partial UDG	Peru	D1
A10725*	KU523315	Llamas et al., 2016	partial UDG	Peru	C1b
A10800*	KU523285	Llamas et al., 2016	partial UDG	Peru	B2b
A10805*	KU523344	Llamas et al., 2016	partial UDG	Peru	D1
A10814*	KU523287	Llamas et al., 2016	partial UDG	Peru	B2
A10817*	KU523324	Llamas et al., 2016	partial UDG	Peru	C1b
A13272*	KU523302	Llamas et al., 2016	partial UDG	Bolivia	B2

Table 2. Mitochondrial gene regions as defined by (Behar et al., 2012)

Gene	Start	End
D-loop (1)	1	576
12S RNA	648	1601
16sRNA	1671	3229
ND1	3307	4262
ND2	4470	5511
COX1	5904	7445
COX2	7586	8269
ATP8	8366	8572
ATP6	8527	9207
COX3	9207	9990
ND3	10059	10404
ND4L	10470	10766
ND4	10760	12137
ND5	12337	14148
ND6	14149	14673
CYTB	14747	15887
D-loop (2)	16024	16569

Table 3. Average proportion of PMD sites in the mitochondrial genome for all specimens included in this study

Sample ID	Average proportion of PMD sites
A10709	0.422
A10717	0.364
A10720	0.427
A10722	0.423
A10725	0.429
A10731	0.396
A10732	0.421
A10789	0.339
A10791	0.293
A10800	0.413
A10805	0.425
A10814	0.429
A10817	0.407
A11116	0.289
A11121	0.334
A11165	0.432
A11168	0.436
A11176	0.400
A11205	0.361
A11209	0.414
A13237	0.406
A13272	0.427
A13989	0.255
JK2127	0.437
JK2130	0.407
JK2131	0.442
JK2133	0.437
JK2134	0.425
JK2136	0.400
JK2139	0.360
JK2866	0.420
JK2879	0.416

Chapter 3

Improved recovery of ancient DNA from archived extracts

Manuscript submitted to *BMC Research Notes*

Statement of Authorship

Title of Paper	Improved recovery of ancient DNA from archived extracts
Publication Status	☐ Published ☐ Accepted for Publication ☒ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Manuscript submitted to *BMC Research Notes*

Principal Author

Name of Principal Author (Candidate)	Ayla L. van Loenen		
Contribution to the Paper	Designed experiments, performed laboratory experiments, compiled sequencing data and performed bioinformatics analyses. Analysed and interpreted results. Wrote the paper with comments from co-authors.		
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	3/7/18

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, interpreted results, made comments on the manuscript.
Signature	Date: 3/7/18

Name of Co-Author	Bastien Llamas
Contribution to the Paper	Designed experiments, interpreted results, made comments on the manuscript.
Signature	Date: 3/7/18

Name of Co-Author	Kieren Mitchell
Contribution to the Paper	Interpreted results, made comments on the manuscript.
Signature	Date: 3/7/18

RESEARCH

Improved recovery of ancient DNA from archived extracts.

Ayla L. van Loenen*, Kieren J. Mitchell, Bastien Llamas and Alan Cooper

*Correspondence:
aylalore22@gmail.com,
kieren.mitchell@adelaide.edu.au,
bastien.llamas@adelaide.edu.au,
alan.cooper@adelaide.edu.au
Australian Centre for Ancient DNA (ACAD), University of Adelaide, Adelaide, Australia
Full list of author information is available at the end of the article

Abstract

Objective: Ancient DNA (aDNA) research has provided key information about the evolutionary history of a wide range of species. However, the limitations of early aDNA techniques meant that many initial attempts to sequence DNA from ancient specimens were unsuccessful. Recent advances in laboratory methods and the advent of High-Throughput Sequencing technologies have made it possible to analyse aDNA from a far broader range of specimens, including those that previously yielded negative results. There are a number of potential pitfalls with archived DNA extracts including: evaporation of the solution, DNA degradation, or adsorption to plastic tube walls over time. Consequently, researchers can be hesitant to use archived extracts, instead opting to re-sample the specimen and perform new DNA extractions. *Results*: Here we trial several methods to optimise the recovery and re-use of archived aDNA extracts, and demonstrate these extracts can be successfully sequenced using current laboratory methods and High-Throughput Sequencing technologies without the need for re-extraction.

Keywords: Ancient DNA; evolutionary genetics; bison

Introduction

Understanding the evolutionary history of species is critical for predicting how species may respond in future to factors such as climate and environmental change, and to ensure the development of appropriate conservation management strategies for current and future threatened species. When applied to humans, past evolutionary patterns of migration and adaptation to various environments or diets have major implications for understanding human health [1]. However, interpretations of evolutionary history based solely on extant populations are complicated by the fact that long-term genetic patterns of evolutionary history are often lost following population bottlenecks or genetic admixture (i.e. interbreeding between different populations or species) [2, 3, 4]. Fortunately, the remains of individuals from extinct populations or species (such as subfossil bones or museum skin specimens) can be analysed using ancient DNA (aDNA) in order to directly observe genetic change through time [3, 5].

Studies incorporating aDNA have identified complex and dynamic patterns of population change over time that were not previously easily observable in analyses of extant specimens (e.g. [6, 7, 8, 3]). However aDNA is characterised by extensive post-mortem damage that can impede genetic analysis, including DNA strand breaks, oxidative damage, depurination, and hydrolytic deamination [9]. As a result of this damage and the limitations of early aDNA techniques, initial success with aDNA recovery was limited to especially well-preserved specimens typically

found in cool/dry regions or permafrost [10, 8, 11, 12]. Recent advances in laboratory methods such as genomic library preparation for High-Throughput Sequencing (HTS) and hybridisation capture techniques [13, 14, 15, 16, 17] have made it possible to sequence mitochondrial genomes and nuclear DNA from a far broader range of ancient specimens (e.g. [18, 19, 20]). The improved recovery of short fragments has even allowed successful DNA sequencing of samples that were unsuccessful with earlier techniques [8, 11, 12, 21].

Ancient DNA studies are not only limited by our ability to sequence DNA from a specimen, but also by the availability and abundance of specimens for sampling. The number of suitably preserved ancient specimens varies widely between species, varying from thousands of North American Pleistocene bison, to just a handful of specimens for some archaic hominids. Extensive sampling over the past 30 years of aDNA research, plus the development of improved methodologies, has led to an increased demand for subfossil specimens [22]. In particular, permission to resample specimens from museum collections has become increasingly difficult. For particularly rare or small specimens, there may be no bone remaining to resample [23]. For example, Denisovan bones and teeth are especially rare and much of the specimens found to date have already been destroyed for analysis [24, 25, 26, 27].

For most of these previously sampled ancient specimens, the original aDNA extracts have been archived for long-term storage. However, there are a number of potential pitfalls with the long-term storage of DNA extracts such as evaporation of the solvent (typically an elution buffer or H_2O) or DNA binding to the walls of tubes over time [28, 29, 30]. In the present study we evaluate the potential to sequence genomic data from archived DNA extracts and trial several methods for optimizing the recovery and re-use of such archived aDNA.

Main Text

Methods

Sixteen steppe bison *(Bison priscus)* DNA extracts were used in this study (Supplementary Table 1). These extracts were initially prepared in the early 2000s using a phenol-cholorophorm extraction method [7], and have been stored in 500μL polypropylene tubes at -20°C since then (Supplementary Table 1). For this study, we diluted each extract with ddH_2O to a final volume of 100μL ($\sim$75μL added to each sample) to ensure sufficient extract remained to trial a number of different treatments. All aDNA work was performed in the purpose built isolated aDNA facility at the Australian Centre for Ancient DNA (ACAD) at the University of Adelaide following previously published guidelines [31, 15].

Four potential DNA recovery treatment methods were tested on the sixteen extracts; No treatment, Tween-20, Tween-20 + heat + agitation, and Sodium Dodecyl Sulfate (SDS) + agitation (described further in Box 1). Tween-20 is a non-ionic surfactant, which should help to reduce surface tension and increase hydrophilicity, enabling the solution to more effectively contact the walls of the tube and potentially allow bound DNA to be re-suspended [32]. SDS is also a surfactant that reduces the surface tension of the extract, but unlike Tween-20 it is anionic (negatively charged in solution) [32]. Anionic surfactants such as SDS may facilitate the release of DNA from compact complexes into solution more effectively than Tween-20, potentially

allowing re-suspension of DNA adhering to tube walls [33, 34]. Both surfactants used in these treatments (SDS and Tween-20) have also been used in other studies to remove PCR inhibitors (such as phenol or salts carried over from DNA extraction), and may have similar effect in this study (reducing PCR inhibition rather than improving the recovery of DNA fragments). [28] has also suggested that other contaminants or PCR inhibitors present in the elute could affect DNA adsorption.

Box 1 - Description of treatments.

1. **No treatment.** A 25μL aliquot of each 100μL archived aDNA extract was removed from each sample before any specialized treatment as a control for amplification and sequencing analysis, leaving 75μL for subsequent treatments.
2. **Tween-20.** For eight of the 75μL extracts we added Tween-20 (Sigma-Aldrich) at a final concentration of 0.05%, then mixed thoroughly by pipetting and incubated the tubes at 4°C overnight (as used by [35]). We removed 25μL for analysis.
3. **Tween-20 + heat + agitation.** For the same eight samples subjected to Treatment 2, we incubated the remaining 50μL volume in the tubes at 40°C for 30 min in a thermoshaker (Grant Bio) at 500RPM. The aim of this step was to further improve the solubility and subsequent recovery of DNA.
4. **Sodium Dodecyl Sulfate (SDS) + agitation.** For the second set of eight extracts (i.e. those not subjected to Treatments 2 and 3) we added Sodium Dodecyl Sulfate (SDS; Sigma-Aldrich) at a final concentration of 0.05%, mixed thoroughly by pipetting, and then kept the tubes on the thermoshaker (Grant Bio) at 500RPM overnight at room temperature.

Double-stranded Illumina libraries were built following the protocol of [14] using 25μl of extract from each recovery treatment, along with P5 and P7 7-mer internal barcode sequences as per [36]. The libraries were pooled and sequenced on an Illumina NextSeq 500 (2x75 paired-end chemistry). Sequencing reads were demultiplexed, quality-filtered, and mapped to a published mitochondrial genome of American bison (NC012346) using sabre (https://github.com/najoshi/sabre), mapDamage v2.0.6 [37], BWA v0.7.15 [38], AdapterRemoval v2.2.1 [39], Genome Analysis ToolKit [40], Picard Tools (http://broadinstitute.github.io/picard), and SAMtools v1.3.1 [41, 42], as previously published [36]. The summary sequencing values for each library from each treatment method were compiled from the PALEOMIX pipeline output (Supplementary Tables 2 and 3), and damage patterns were consistent with aDNA (Supplementary Table 4). As the number of reads sequenced per library is relative to the proportion of endogenous DNA in each sample, the sequencing summary results were compared as paired measures for each specimen using a Wilcox Signed Rank test within the `stats` package in `R` [43]. Due to the wide range of raw values observed between samples (a consequence of varying levels of post-mortem DNA damage and proportions of contaminant to endogenous DNA among specimens), the raw values for number of reads sequenced (i.e. RetainedReads, HitsRaw,

HitsUniq) were normalized to re-scale to the unit interval (0-1). The normalized values were plotted using the `ggplot2` package in `R` [43, 44]. The predictive power of various factors (detailed in Table 2) on the resulting 'normalized' summary sequencing values was tested using the Generalized Linear Models function within the `lme4` package in `R` [43, 45].

Results

We were able to obtain reads mapping to the bison mitochondrial genome with low coverage shotgun sequencing for fifteen out of the sixteen specimens included in this study (Supplementary Table 2). For nine out of those fifteen specimens, we were even able to obtain mitochondrial genomes with coverage greater than 0.5x. This includes three out of four specimens for which previous PCR-based methods were either unsuccessful or partially successful (Supplementary Table 1). Increasing the sequencing effort, or including hybridization capture methods would likely increase the coverage and depth per sample.

Comparison of normalized sequencing summary values from the treated and untreated libraries prepared from the same samples showed that 6 of 8 samples treated with Tween-20 yielded more uniquely mapping reads (3 of 8 for Treatment 2, 5 of 8 for Treatment 3) than the control (Treatment 1) (Figure 1, Supplementary Table 2). In contrast, only 3 of 8 samples treated with SDS (Treatment 4) yielded more uniquely mapping reads than the control (Figure 1, Supplementary Table 2). However, there were no significant differences observed in sequencing summary values between libraries made from the treated and untreated (control) groups (Tables 1). In particular, there were no significant changes in the number of reads sequenced (total or unique), average DNA read length, or endogenous content (Tables 1).

Further, testing the predictive power of various factors on the normalized sequencing summary values using generalized linear models showed that the factor with the lowest AIC score (and consequently the strongest individual predictive factor) in all instances was 'sample ID' and not treatment type or other potential predictive factors (such as DNA read length, sample age or endogenous content) (Table 2). Sample ID refers to the unique identifier given to each extract, so this result means that each specimen was unique and there was no single factor (from those tested in this study) that could predict sequencing success *a priori*. Combining multiple factors together (e.g. treatment type + sample ID) did not tend to improve the predictive power (Table 2).

Discussion

Collectively, our results suggest that none of the tested treatments were successful in broadly improving the recovery of aDNA, as none demonstrated significantly improved results compared to the control (Treatment 1). Initial observation of plots of the normalized results (Figure 1) suggested that there may be a positive correlation between the addition of Tween-20 and the recovery of additional unique DNA sequences (Treatments 2 and 3), and a potential negative correlation with the addition of SDS (Treatment 4). However, this pattern is not consistent across all specimens, nor is the effect of the treatment significantly correlated with our results. The stochasticity of the results likely reflects the fact that each sample is unique,

and a treatment that may improve DNA recovery (or reduce PCR inhibition) for one sample may not necessarily work for another. Given that both DNA adsorption and PCR inhibition is known to be extremely stochastic (5-95% adsorption within the same batch of tubes [28]), it is also possible (though unlikely) that by chance the tubes of all the DNA extracts used in this experiment had low levels of adsorption. The stochasticity of DNA adsorption may also explain the lack of significance in the observed trends in the plots. Specifically, by chance many of the archived extracts did not have significant proportions of DNA adsorption, and as a consequence Tween-20 treatments did not appear to have a statistically significant effect. Regardless, we were able to sequence molecules mapping to the American bison mitochondrial genome for fifteen of the sixteen old DNA extracts in this study, regardless of whether or not they were previously successfully analysed with PCR and Sanger sequencing (Supplementary Table 1).

Re-sequencing existing archived aDNA extracts is likely the fastest and most cost-effective way to rapidly increase the temporal and geographical range of specimens for broad continental-scale studies of evolutionary history of species. Consequently, the first option for expanding any previous aDNA research project into the 'genomics age' should be to identify previously extracted samples and re-sequence them using the most optimised laboratory and sequencing methods available. There was no significant benefit observed from performing additional treatments to the archived DNA extracts to improve sequencing success.

Limitations

The main limitation in this study is the relatively small sample size, which may be concealing minor associations in the data. However, if there was a true trend across treatment types and recovery of archived DNA, we would expect the results to be consistent between treatment types and not sample dependent as demonstrated here. Larger scale analyses would be necessary to untangle the stochastic effect of DNA adsorption from different treatments to improve the recovery of archived DNA.

Supplementary files
Supplementary Table 1 - Sample Information. A table containing all available sample information for the extracts used in this study. Supplementary Table 2 - PALEOMIX summary values. A table containing the summary values from the PALEOMIX pipeline output. Supplementary Table 3 - PALEOMIX definitions. A table defining the PALEOMIX variables used in subsequent comparisons. Supplementary Table 4 - MapDamage values. A table containing the MapDamage values for each of the libraries generated per extract.

Declarations
Abbreviations
ancient DNA (aDNA) High-Throughput Sequencing (HTS) Sodium Dodecyl Sulfate (SDS)

Competing interests
The authors declare that they have no competing interests.

Availability of Data and Materials
All data analysed during this study are included in this published article [and its supplementary information files]. Raw data files generated during this study are available from Zenodo (doi:10.5281/zenodo.1476377).

Ethics approval and consent to participate
Not applicable

Consent to publish
Not applicable

Author's contributions

ALvL, BL and AC designed experiments, ALvL, KJM, BL and AC interpreted results, ALvL wrote the manuscript with comments from all co-authors. All authors read and approved the manuscript.

Acknowledgements

We thank the reviewers for their constructive feedback on this manuscript.

Funding

ALvL was supported by an Australian Government Research Training Program (RTP) Scholarship. K.J.M, B.L., and A.C. are supported by the Australian Research Council.

References

1. Marciniak, S., Perry, G.H.: Harnessing ancient genomes to study the history of human adaptation. Nature Reviews Genetics **18**(11), 659 (2017)
2. Ramakrishnan, U., Hadly, E.A., Mountain, J.L.: Detecting past population bottlenecks using temporal genetic data. Molecular Ecology **14**(10), 2915–2922 (2005)
3. Leonard, J.A.: Ancient DNA applications for wildlife conservation. Molecular Ecology **17**(19), 4186–4196 (2008)
4. Chang, D., Shapiro, B.: Using ancient DNA and coalescent-based methods to infer extinction. Biology Letters **12**(2), 20150822 (2016)
5. Orlando, L., Cooper, A.: Using Ancient DNA to Understand Evolutionary and Ecological Processes. Annual Review of Ecology, Evolution, and Systematics **45**(1), 573–598 (2014)
6. Barnes, I., Matheus, P., Shapiro, B., Jensen, D., Cooper, A.: Dynamics of Pleistocene population extinctions in Beringian brown bears. Science **295**(5563), 2267–2270 (2002)
7. Shapiro, B., Drummond, A.J., Rambaut, A., Wilson, M.C., Matheus, P.E., Sher, A.V., Pybus, O.G., Gilbert, M.T.P., Barnes, I., Binladen, J., Willerslev, E., Hansen, A.J., Baryshnikov, G.F., Burns, J.A., Davydov, S., Driver, J.C., Froese, D.G., Harington, C.R., Keddie, G., Kosintsev, P., Kunz, M.L., Martin, L.D., Stephenson, R.O., Storer, J., Tedford, R., Zimov, S., Cooper, A.: Rise and Fall of the Beringian Steppe Bison. Science **306**(5701), 1561–1565 (2004)
8. Hagelberg, E., Hofreiter, M., Keyser, C.: Ancient DNA: the first three decades. Philosophical Transactions of the Royal Society B: Biological Sciences **370**(1660), 20130371 (2015)
9. Gilbert, M.T.P., Hansen, A.J., Willerslev, E., Rudbeck, L., Barnes, I., Lynnerup, N., Cooper, A.: Characterization of Genetic Miscoding Lesions Caused by Postmortem Damage. The American Journal of Human Genetics **72**(1), 48–61 (2003)
10. Hofreiter, M., Serre, D., Poinar, H.N., Kuch, M., Pääbo, S.: Ancient DNA. Nature Reviews Genetics **2**(5), 353 (2001)
11. Orlando, L., Gilbert, M.T.P., Willerslev, E.: Reconstructing ancient genomes and epigenomes. Nature Reviews Genetics **16**(7), 395 (2015)
12. Sproul, J.S., Maddison, D.R.: Sequencing historical specimens: successful preparation of small specimens with low amounts of degraded DNA. Molecular Ecology Resources, (2017)
13. Meyer, M., Kircher, M.: Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. Cold Spring Harbor Protocols **2010**(6), 5448 (2010)
14. 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 **370**(1660), 20130624 (2015)
15. Hofreiter, M., Shapiro, B. (eds.): Ancient DNA: Methods and Protocols. Methods in Molecular Biology, Humana Press Incorporated, New York (2012)
16. García-García, G., Baux, D., Faugère, V., Moclyn, M., Koenig, M., Claustres, M., Roux, A.-F.: Assessment of the latest NGS enrichment capture methods in clinical context. Scientific Reports **6** (2016)
17. Nieves-Colón, M.A., Ozga, A.T., Pestle, W.J., Cucina, A., Tiesler, V., Stanton, T.W., Stone, A.C.: Comparison of two ancient dna extraction protocols for skeletal remains from tropical environments. American Journal of Physical Anthropology (2018)
18. Posth, C., Renaud, G., Mittnik, A., Drucker, D.G., Rougier, H., Cupillard, C., Valentin, F., Thevenet, C., Furtwängler, A., Wißing, C., *et al.*: Pleistocene mitochondrial genomes suggest a single major dispersal of non-Africans and a Late Glacial population turnover in Europe. Current Biology **26**(6), 827–833 (2016)
19. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., *et al.*: Massive migration from the steppe was a source for indo-european languages in europe. Nature **522**(7555), 207 (2015)
20. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., *et al.*: The genetic history of ice age europe. Nature **534**(7606), 200 (2016)
21. Llamas, B., Valverde, G., Fehren-Schmitz, L., Weyrich, L.S., Cooper, A., Haak, W.: From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. STAR: Science & Technology of Archaeological Research **3**(1), 1–14 (2017)
22. Makarewicz, C., Marom, N., Bar-Oz, G.: Palaeobiology: Ensure equal access to ancient DNA. Nature **548**(7666), 158–158 (2017)
23. Korlević, P., Talamo, S., Meyer, M.: A combined method for DNA analysis and radiocarbon dating from a single sample. Scientific Reports **8**(1), 4127 (2018)
24. Brown, S., Higham, T., Slon, V., Pääbo, S., Meyer, M., Douka, K., Brock, F., Comeskey, D., Procopio, N., Shunkov, M., *et al.*: Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. Scientific Reports **6**, 23559 (2016)
25. Stringer, C.B., Barnes, I.: Deciphering the Denisovans. Proceedings of the National Academy of Sciences **112**(51), 15542–15543 (2015)

26. Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., De Filippo, C., *et al.*: A high-coverage genome sequence from an archaic Denisovan individual. Science **338**(6104), 222–226 (2012)
27. Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L.F., *et al.*: Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature **468**(7327), 1053–1060 (2010)
28. Gaillard, C., Strauss, F.: Avoiding adsorption of DNA to polypropylene tubes and denaturation of short DNA fragments. Technical Tips Online **T01392**, 1–4 (1998)
29. Gaillard, C., Strauss, F.: Eliminating DNA loss and denaturation during storage in plastic microtubes. International Biotechnology Laboratory **18**(13), 6 (2001)
30. Lee, S.B., Crouse, C.A., Kline, M.C.: Optimizing Storage and Handling of DNA Extracts. Forensic Science Review **22**(2), 131–144 (2010)
31. Cooper, A., Poinar, H.N.: Ancient DNA: do it right or not at all. Science **289**(5482), 1139–1139 (2000)
32. Pan, H., Xia, Y., Qin, M., Cao, Y., Wang, W.: A simple procedure to improve the surface passivation for single molecule fluorescence studies. Physical Biology **12**(4), 045006 (2015)
33. Dias, R., Antunes, F., Miguel, M., Lindman, S., Lindman, B.: DNA-lipid systems: A physical chemistry study. Brazilian Journal of Medical and Biological Research **35**(5), 509–522 (2002)
34. Dias, R.S., Pais, A.A.C.C., Miguel, M.G., Lindman, B.: DNA and surfactants in bulk and at interfaces. Colloids and Surfaces A: Physicochemical and Engineering Aspects **250**(1), 115–131 (2004)
35. Mitchell, K.J., Llamas, B., Soubrier, J., Rawlence, N.J., Worthy, T.H., Wood, J., Lee, M.S., Cooper, A.: Ancient dna reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. Science **344**(6186), 898–900 (2014)
36. Soubrier, J., Gower, G., Chen, K., Richards, S.M., Llamas, B., Mitchell, K.J., Ho, S.Y.W., Kosintsev, P., Lee, M.S.Y., Baryshnikov, G., *et al.*: Early cave art and ancient DNA record the origin of European bison. Nature Communications **7**, 13158 (2016)
37. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L., Orlando, L.: mapdamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics **29**(13), 1682–1684 (2013)
38. Li, H., Durbin, R.: Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics **25**(14), 1754–1760 (2009)
39. Lindgreen, S.: AdapterRemoval: easy cleaning of next-generation sequencing reads. BMC Research Notes **5**(1), 337 (2012)
40. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., *et al.*: The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research **20**, 1297–1303 (2010)
41. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.: The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**(16), 2078–2079 (2009)
42. Li, H.: A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics **27**(21), 2987–2993 (2011)
43. R Development Core Team: R: A Language and Environment for Statistical Computing. Vienna, Austria (2017). http://www.R-project.org
44. Wickham, H.: ggplot2: elegant graphics for data analysis. J Stat Softw **35**(1), 65–88 (2010)
45. Bates, D., Mächler, M., Bolker, B., Walker, S.: Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software **67**, 1–48 (2015)

Figure

Figure 1 Normalised HitsUniq values from PALEOMIX summary statistics for each specimen coloured by treatment. Values are compared as relative measures (i.e. scaled relative to the sequencing effort for each treatment, and then re-scaled to the unit interval 0-1). Treatment 1 (control) is in blue, treatment 2 (Tween-20) in green, treatment 3 (Tween plus heat and agitation) in yellow, and treatment 4 (SDS + agitation) in red. The first eight specimens were subjected to treatments 1, 2, and 3, while the second eight specimens were subjected to treatments 1 and 4.

Tables

Table 1 P values from pairwise tests on normalised PALEOMIX values

Variable		Treatment 1	Treatment 2	Treatment 3	Treatment 4
RetainedReads	Treatment 1	-	0.6742	0.6742	0.07399
	Treatment 2	0.6742	-	0.7525	-
	Treatment 3	0.6742	0.7525	-	-
	Treatment 4	0.07399	-	-	-
HitsRaw	Treatment 1	-	0.8335	0.5992	0.5283
	Treatment 2	0.8335	-	0.9163	-
	Treatment 3	0.5992	0.9163	-	-
	Treatment 4	0.5283	-	-	-
HitsUniq	Treatment 1	-	0.8335	0.5992	0.5283
	Treatment 2	0.8335	-	0.9163	-
	Treatment 3	0.5992	0.9163	-	-
	Treatment 4	0.5283	-	-	-
Endogenous	Treatment 1	-	1	0.8335	0.8335
	Treatment 2	1	-	0.6742	-
	Treatment 3	0.8335	0.6742	-	-
	Treatment 4	0.8335	-	-	-
LengthStd_Mit	Treatment 1	-	0.6733	0.5271	0.8335
	Treatment 2	0.6733	-	0.4587	-
	Treatment 3	0.5271	0.4587	-	-
	Treatment 4	0.8335	-	-	-
LengthStd_Nuc	Treatment 1	-	0.2933	0.05852	0.9163
	Treatment 2	0.2933	-	0.4619	-
	Treatment 3	0.05852	0.4619	-	-
	Treatment 4	0.9163	-	-	-

Table 2 AIC scores from glm tests

Explanatory variable	Coverage_Mit	Endogenous	HitsUniq	RetainedReads	LengthStd_Mit	LengthStd_Nuc
Treatment	39.71	30.93	1043	-25.99	0.4761	-37.28
Sample ID	-29.13	-52.01	-1599	-31.76	-16.55	-54.71
Clonality	26.02	20.73	1039	-30.31	-20.22	-38.61
SampleAge (calibrated)	7.467	-11.33	270.2	291.7	66.01	71.82
SampleAge (MIS)	10.89	-29.43	621.9	662.1	151.6	150.8
LengthStd_Mit	37.35	21.39	889.3	-30.97	-	-51.04
LengthStd_Nuc	38.92	26.16	1043	-28.25	-14.68	-
Treatment+Sample ID	-29.47	-53.56	-1578	-31.35	-18.38	-67.2
Treatment+Sample ID +Clonality	-27.56	-51.64	-1578	-29.39	-16.43	-82.72

Supplementary Information

Supplementary Table 1. Sample Information

Sample ID	PCR success (per [7])	Coverage mtgenome	RetainedReads	Treatments	Species ID	Country	Location	Museum accession	Museum
A1019	Y (all)	0.99	2238590	1, 2, 3	Bison	Russia	Kolyma lowland, Stanchikovsky Yar, Siberia	CRS-SY-11	ChNRS
A1026	Y (all)	0.99	2502417	1, 2, 3	Bison	Russia	Kolyma lowland, Bol. KhomusYuryakh R., Siberia	CRS-IC-18	ChNRS
A1027	N	0.09	2172826	1, 2, 3	Bison	Russia	Kolyma lowland, Chersky, Siberia	CRS-CB-1	ChNRS
A1032	N	0.99	2381464	1, 2, 3	Bison	Canada	Oro Grande Cr., Dawson City, YT	CMN 47476	CMN
A1037	Y (all)	0.76	2467288	1, 2, 3	Bison	USA	Sheep Cr., Fairbanks, AK	V-37-30	UAF
A938	Y (all)	0	288251	1, 2, 3	Bison	Canada	Charlie Lake Cave, Peace R., BC	1848	SFU
A943	Y (all)	0.99	2179180	1, 2, 3	Bison	USA	Natural Trap Cave (KU-WY-90), WY	KU 43556	UKans
A944	N	NA	639168	1, 2, 3	Bison	USA	Natural Trap Cave (KU-WY-90), WY	KU 61712	UKans
A1046	Y (all)	0.02	302456	1, 4	Bison	USA	Natural Trap Cave (KU-WY-90), WY	KU 51275	UKans
A1048	Y (>2/3)	0.34	3693898	1, 4	Bison	Canada	Maisy May Cr., Dawson City, YT	CMN 45517	CMN
A2991	Y (all)	0.03	632072	1, 4	Bison	USA	Black River, Yukon Flats, Alaska	RS9202	UAF
A6396	Y (all)	0.9	585156	1, 4	Bison	Canada	Evergreen Cr., Dawson City, YT	110.19	ABC
A939	Y (all)	0.03	433815	1, 4	Bison	Canada	Charlie Lake Cave, Peace R., BC	2294	SFU
BS326	Y (all)	0.99	671162	1, 4	Bison	Russia	Kolyma lowland, Bol. KhomusYuryakh R., Siberia	CRS-IC-29	ChNRS
BS340	Y (all)	0.99	683205	1, 4	Bison	USA	Ikpikpuk R., North Slope, AK	IK-98-302	BLM
BS408	Y (all)	0.54	509414	1, 4	Bison	Russia	Yana-Indigirka lowland, Siberia	StP-1	SIN

Supplementary Table 2. PALEOMIX summary values

SampleID	RetainedReads	HitsRaw	HitsUniq	Endogenous	Clonal	Raw_Mit	Uniq_Mit	Coverage_Mit	Length_Mit	LengthStd_Mit	Length_Nuc	LengthStd_Nuc
BS310-T1	734672	99123	97684	0.135	0.015	780	744	0.87	55.53	20.09	59.64	18.31
BS310-T2	669672	98239	96752	0.147	0.015	781	738	0.88	52.89	16.42	56.35	16.06
BS310-T3	822366	108232	106059	0.132	0.020	532	525	0.79	63.96	23.18	60.13	18.57
BS299-T1	709890	297471	292634	0.419	0.016	1666	842	0.93	56.53	17.31	59.13	18.51
BS299-T2	842313	356362	350579	0.423	0.016	1822	979	0.94	57.92	17.44	61.12	19.12
BS299-T3	934722	421327	409654	0.451	0.028	2714	1107	0.96	58.58	18.88	61	20.21
BS298-T1	589361	1035	1010	0.002	0.024	37	24	0.05	45.83	13.02	42.42	15.41
BS298-T2	819461	1546	1446	0.002	0.065	101	18	0.04	47.78	16.27	47.87	14.56
BS298-T3	743257	1494	1469	0.002	0.017	41	30	0.05	44.77	8.65	42.75	14.06
BS295-T1	791646	125967	124380	0.159	0.013	740	728	0.9	62.5	18.89	64.07	17.69
BS295-T2	866601	182820	180662	0.211	0.012	1217	1176	0.95	52.55	15.53	55.13	15
BS295-T3	703819	126387	124757	0.180	0.013	493	447	0.69	52.14	14.57	50.43	13.33
BS285-T1	729947	88130	86712	0.121	0.016	94	93	0.39	88.85	38.6	94.02	36.05
BS285-T2	865385	89746	88644	0.104	0.012	80	77	0.26	72.25	28.05	71.24	24.87
BS285-T3	863107	116508	115050	0.135	0.013	162	145	0.43	63.59	21.98	66.52	23.13
BS337-T1	66786	769	620	0.012	0.194	2	2	0	48	19.8	52.54	16.32
BS337-T2	353	2	2	0.006	0.000	0	0	NA	NA	NA	25.5	0.71
BS337-T3	212369	22	13	0.000	0.409	0	0	NA	NA	NA	46.23	13.37
BS343-T1	513550	150721	147855	0.293	0.019	649	633	0.86	58.88	22.45	69.54	23.72
BS343-T2	756992	241347	237242	0.319	0.017	1421	1367	0.93	43.33	11.02	47.93	12.79
BS343-T3	882909	295328	291507	0.334	0.013	1233	1216	0.96	55.32	16.92	58.88	17.73
BS344-T1	139674	1889	1403	0.014	0.257	0	0	NA	NA	NA	65.02	21.9
BS344-T2	11584	89	31	0.008	0.652	0	0	NA	NA	NA	47.84	20.62
BS344-T3	482497	425	162	0.001	0.619	0	0	NA	NA	NA	59.81	23.7
BS326-T1	375133	136727	134470	0.364	0.017	1592	1556	0.99	57.63	20.53	64.97	19.25
BS326-T4	264082	25968	25200	0.098	0.030	411	401	0.72	54.45	17.91	64.24	19
BS340-T1	372005	153769	151654	0.413	0.014	967	954	0.96	63.86	20.29	68.85	20.1
BS340-T4	298664	86697	84666	0.290	0.023	709	683	0.85	54.72	17.21	63.35	18.97
BS342-T1	286267	3119	2842	0.011	0.089	134	16	0.03	42.94	24.07	49.45	14.77
BS342-T4	137631	1032	803	0.007	0.222	28	4	0	25.5	0.58	42.29	17.25
BS356-T1	280980	1873	1849	0.007	0.013	14	14	0.03	42.07	13.46	51.97	18.01
BS356-T4	3385061	24919	23465	0.007	0.058	293	134	0.31	48.83	15.1	45.19	15.72
BS357-T1	2312	22	19	0.010	0.136	1	1	0	26		46.72	11.53
BS357-T4	289532	3266	2693	0.011	0.175	282	12	0.02	36.75	13.05	45.29	14.12
BS408-T1	247232	35088	34608	0.142	0.014	147	143	0.42	64.54	18.49	65.03	18.82
BS408-T4	252025	60931	59986	0.242	0.016	81	79	0.2	55.3	15.97	55.98	15.78
BS414-T1	302604	3444	3287	0.011	0.046	5	5	0.02	75.8	7.82	54.73	17.69
BS414-T4	314239	2610	2506	0.008	0.040	19	6	0	37	20.49	41.25	16.7
BS478-T1	320382	42496	41913	0.133	0.014	714	689	0.87	50.94	15.66	55.03	15.32
BS478-T4	233711	10491	9786	0.045	0.067	164	103	0.26	49.02	15.66	51.98	15.43

Supplementary Table 3. Definitions for PALEOMIS summary statistics

Variable	Definition
RetainedReads	Number of sequenced reads per sample
HitsRaw	Number of sequenced reads mapping to reference genome
HitsUniq	Number of unique reads mapping to the reference genome
Endogenous	Proportion of HitsRaw/RetainedReads
Clonal	Proportion of HitsUniq/HitsRaw
LengthStd_Mit	Average length of reads mapping to the mitochondrial reference genome
LengthStd_Nuc	Average length of reads mapping to the nuclear reference genome

Supplementary Table 4. MapDamage values

SampleID	3pGtoA_freq at position 1	5pCtoT_freq at position 1
BS310-T1	0.102	0.076
BS310-T2	0.096	0.093
BS310-T3	0.105	0.061
BS299-T1	0.063	0.064
BS299-T2	0.050	0.058
BS299-T3	0.023	0.042
BS298-T1	0.099	0.079
BS298-T2	0.101	0.104
BS298-T3	0.088	0.153
BS295-T1	0.144	0.121
BS295-T2	0.090	0.136
BS295-T3	0.149	0.096
BS285-T1	0.069	0.062
BS285-T2	0.064	0.055
BS285-T3	0.054	0.071
BS337-T1	0.087	0.151
BS337-T2	0.000	0.000
BS337-T3	0.000	0.333
BS343-T1	0.071	0.104
BS343-T2	0.092	0.067
BS343-T3	0.070	0.085
BS344-T1	0.020	0.040
BS344-T2	0.222	0.100
BS344-T3	0.033	0.094
BS357-T1	0.000	0.000
BS357-T4	0.189	0.249
BS356-T1	0.239	0.170
BS356-T4	0.329	0.321
BS414-T1	0.033	0.060
BS414-T4	0.107	0.068
BS478-T1	0.176	0.176
BS478-T4	0.196	0.283
BS342-T1	0.175	0.171
BS342-T4	0.268	0.213
BS326-T1	0.095	0.097
BS326-T4	0.179	0.189
BS340-T1	0.065	0.059
BS340-T4	0.176	0.173
BS408-T1	0.071	0.071
BS408-T4	0.095	0.095

Chapter 4

Review of the proposed hypothesis for the evolutionary history of European bison (*Bison bonasus*)

Manuscript prepared for submission

Statement of Authorship

Title of Paper	Review of the proposed hypothesis for the evolutionary history of European bison (Bison bonasus)
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☒ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Manuscript prepared in publication format

Principal Author

Name of Principal Author (Candidate)	Ayla L. van Loenen		
Contribution to the Paper	Compiled and reviewed data. Wrote the paper with comments from co-authors.		
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	3/7/18

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	Made comments on the manuscript		
Signature		Date	3/7/18

Name of Co-Author	Bastien Llamas		
Contribution to the Paper	Made comments on the manuscript		
Signature		Date	3/7/18

Name of Co-Author	Kieren Mitchell		
Contribution to the Paper	Made comments on the manuscript		
Signature		Date	3/7/18

Review of the proposed hypothesis for the evolutionary history of European bison (*Bison bonasus*)

Ayla L. van Loenen[1], **Kieren J. Mitchell**[1], **Bastien Llamas**[1], **and Alan Cooper**[1]

[1]**Australian Centre for Ancient DNA (ACAD), University of Adelaide, Australia**

ABSTRACT

Despite a number of recent research papers on the topic, the evolutionary history of European bison remains poorly understood. Various alternative species designations, and evolutionary relationships between them have been defined, but thus far no consensus has been reached. The lack of a consensus stems primarily from the extremely severe and recent population bottleneck experienced by living bison populations, which has eroded much of the evolutionary history recorded in their genomes. However, signatures of a dynamic history are evident even among modern European bison, which possess a mitochondrial lineage closely related to that of the aurochs despite strong evidence from morphology and the nuclear genome that European bison are the sister-species to American bison. Several studies have used ancient DNA to delve deeper into the history of European bison, and in the process identified a genetically distinct but now extinct mitochondrial clade of European bison. These studies further suggested population replacement and environmental change at a regional level in two specific locations (Urals mountains, Russia and France) during the Pleistocene. However, an explanation for the relationships between these two recognised mitochondrial clades of European bison, and the origin of European bison as a species, remain contentious. Conflicting hypotheses have been proposed for the origin of European bison: either incomplete lineage sorting between *Bos* and *Bison* mitochondrial lineages or hybridisation between aurochs and steppe bison. In this review, we explore the debated aspects of European bison evolutionary history, highlight outstanding questions, and suggest approaches to reconcile and expand on existing data.

Keywords: European bison (*Bison bonasus*), evolutionary history, ancient DNA

Palaeontological history of bison in Eurasia

A variety of bison species and subspecies have been described from across Eurasia, spanning the last 1.8 million years (Figure 1) (Maniakas and Kostopoulos, 2017). Three bovina species in particular are widely recognized through the Late Pleistocene and early Holocene fossil record of Eurasia: aurochs (*Bos primigenius*, Bojanus, 1827), the ancestor of modern cattle; steppe bison (*Bison priscus*, Bojanus 1827); and the European bison (*Bison bonasus*, Linnaeus 1758) (Kurtèn, 1968; Benecke, 2005; Bocherens et al., 2015). Of these, European bison is the only extant species. Morphologically, European bison are most similar to steppe bison and it is generally accepted that they are descendants of steppe bison, either directly or via another species such as *Bison schoetensacki* (Kurtèn, 1968; Vercoutère and Guérin, 2010; Krasińska and Krasiński, 2013; Marsolier-Kergoat et al., 2015; Massilani et al., 2016; Soubrier et al., 2016). Initially found across Eurasia during the Late Pleistocene and Holocene, by the end of the 19th century European bison were restricted to two populations. These populations were defined as two distinct subspecies (Pucek, 2004; Krasińska and Krasiński, 2013; Węcek et al., 2016): one in Białowieża forest (*Bison bonasus bonasus*) and the other in the Caucasus Mountains (*Bison bonasus caucasus*). During the 1920s, both populations were hunted to extinction in the wild, with only a handful of bison remaining in captivity (Pucek, 2004; Krasińska and Krasiński, 2013; Węcek et al., 2016). Steppe bison were distributed throughout Eurasia, Beringia and across the Bering land bridge, ranging as far east as western Canada where they were also the ancestor of modern American bison (*Bison bison*, Linnaeus 1758) (Shapiro et al., 2004). *Bison schoetensaki* was thought to have been distributed throughout Eurasia during the middle Pleistocene, and subsequently went extinct (Kurtèn, 1968; Verkaar et al., 2004; Palacio et al., 2017). However, the description of *Bison schoetensaki* bones is contentious, and their distribution (and existence) after the middle Pleistocene remains under debate (Guérin and Philippe, 1971; Brugal, 1999; Palacio et al., 2017). Despite an extensive fossil record throughout the late Pleistocene and early Holocene to present day, the specific origins,

evolutionary history, taxonomy, and palaeobiogeography of bison throughout Eurasia remains relatively unknown and widely debated (Krasińska and Krasiński, 2013; Massilani et al., 2016; Soubrier et al., 2016).

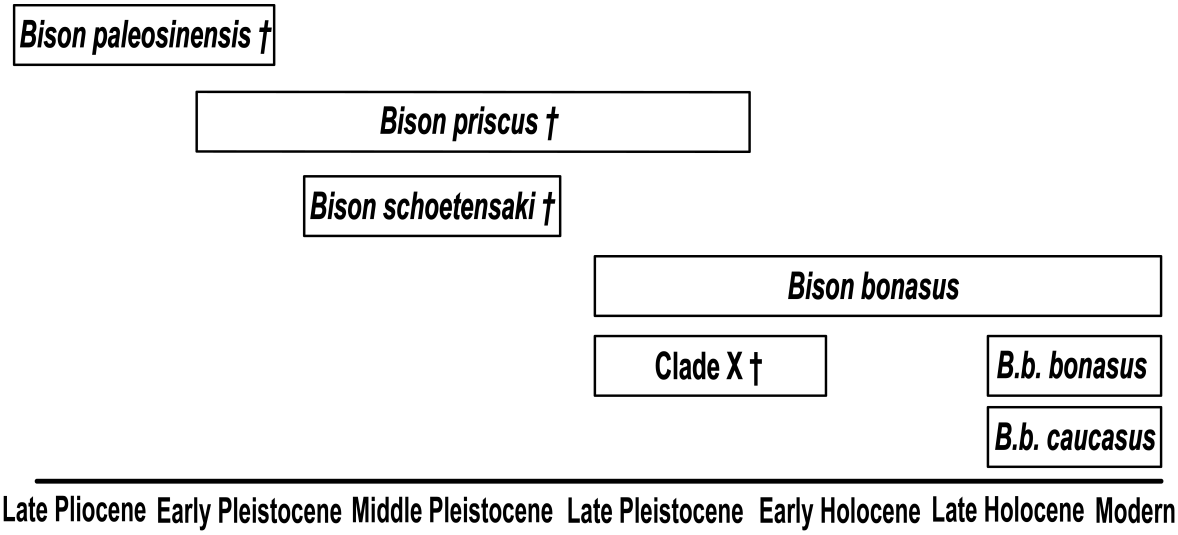


Figure 1. Temporal distribution of *Bison* taxa in Eurasia from the Late Pliocene to present based on palaeontological and genetic data (Kurtèn, 1968; Vercoutère and Guérin, 2010; Verkaar et al., 2004; Benecke, 2005; Krasińska and Krasiński, 2013; Maniakas and Kostopoulos, 2017).

Living bison in Eurasia

Modern European bison are currently listed as vulnerable by the IUCN red list (IUCN, 2008) following their extinction in the wild during the 1920s. The species was reduced to only 54 individuals in captivity (29 males; 25 females) with pedigrees information suggesting they were descended from only 12 founders (Pucek, 2004; Krasińska and Krasiński, 2013). This near-extinction event is thought to have been the result of a combination of habitat fragmentation and anthropogenic pressures, with increased agricultural activity and hunting during the late Holocene until the last known wild European bison was shot in 1927 (Pucek, 2004). Through a captive breeding program (van de Vlasakker, 2014) the species was 'resurrected' and as numbers grew, populations were released back into the wild; initially in Białowieża forest (Poland/Ukraine) and then to additional forest parks across Europe as part of the ReWilding Europe initiative (van de Vlasakker, 2014; EBPB, 2016). As of 2015, there were 6,573 European bison living in a number of locations across Europe, with 1,706 in captivity and 4,867 in free or semi-free ranging herds (EBPB, 2016). However, the extreme population bottleneck from the 1920s caused a massive reduction in genetic diversity across the species that may have had significant consequences on the continued survival of the extant species (Węcek et al., 2016).

Further exacerbating the effects of the population bottleneck in the 1920s, the offspring of the 12 founders in the captive breeding program were divided and managed separately as two genetic lines: the Lowland line (comprised of the offspring of seven 'pure' Lowland founders which were exclusively *B. b. bonasus*), and the Lowland-Caucasian line (comprised of the remaining offspring from all 11 Lowland bison (B.b. bonasus) and one Caucasian bull (*B. b. caucasus*) which did not qualify for inclusion in the Lowland line) (Pucek, 2004; Olech, 2006; Krasińska and Krasiński, 2013). The Lowland line is a 'closed line', which means that only registered offspring of the founders can be classified as a part of this line in order to preserve the lineage (Krasińska and Krasiński, 2013; Wojciechowska et al., 2017). Both genetic lines are maintained through to present day (IUCN, 2008; Krasińska and Krasiński, 2013; Wojciechowska et al., 2017). Compounding the loss of genetic diversity in extant populations of European bison is the fact that not all of the founding individuals contributed equally to subsequent generations. Tokarska et al. (2011) discovered that around 80% of the ancestry of modern European bison could be traced back to just two of the original 12 founders.

A range of genetic markers have been used to study extant European bison populations (Gralak et al., 2004; Luenser et al., 2005; Radwan et al., 2007; Wójcik et al., 2009; Babik et al., 2012; Tokarska et al., 2009a,b, 2011). Several studies have attempted to resolve the phylogenetic relationships between bovine species and have consistently identified an unexpected disagreement between mitochondrial and nuclear data with regards to the

affinities of the European bison (Verkaar et al., 2004). Nuclear DNA markers tend to suggest a sister-species relationship between European and American bison, but mitochondrial DNA markers suggest that European bison are most closely related to *Bos* species, such as cattle (Verkaar et al., 2004; Bibi, 2013). Morphological and cross-fertility studies also support a closer relationship between European and American bison (Verkaar et al., 2004), suggesting that the mitochondrial data do not reflect the true species tree. There are two possible explanations for the mitochondrial discordance: incomplete lineage sorting (ILS) or an ancestral hybridization event (Figure 2). Distinguishing between these two hypotheses has proven challenging based on data from extant populations, and no consensus exists on this issue.

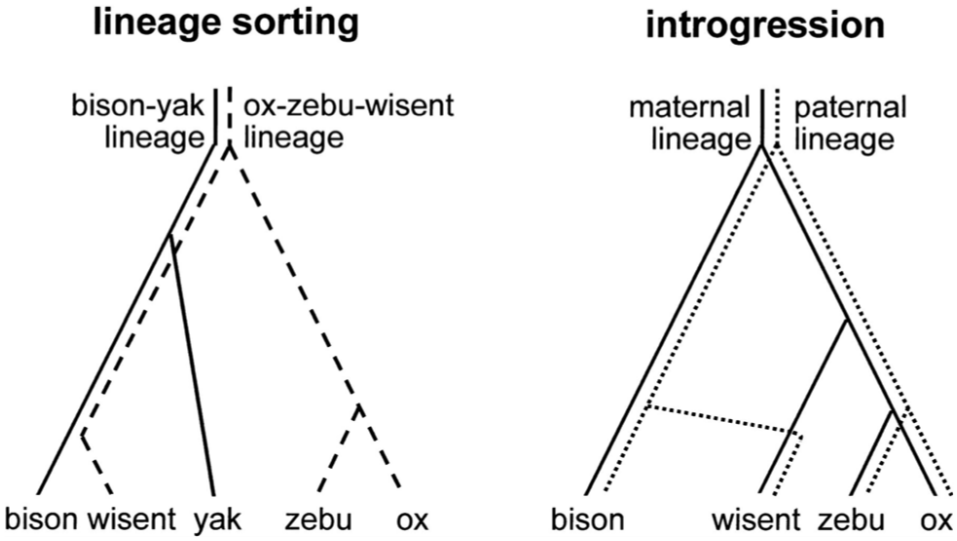


Figure 2. From Verkaar et al. (2004) illustrating both proposed explanations for the mitochondrial divergence between European bison (termed wisent in this figure) and American bison (termed bison in this figure).

Whole genome comparisons between several low-coverage genomes from modern and historical European bison have revealed multiple instances of admixture between the two late Holocene subspecies (*B. b. caucasus* and *B. b. bonasus*), as well as previous admixture with the cattle/aurochs lineage (*Bos taurus/Bos primigenius*) (Węcek et al., 2016; Gautier et al., 2016). The delineation of a specific source of admixture from within the cattle/aurochs lineage into European bison is not possible with the single aurochs genome used in Węcek et al. (2016), and requires additional data. Gautier et al. (2016) also characterised a potential decline in population size following the Last Glacial Maximum (LGM) around 20 thousand years ago (kya), and identified a number of genes differentially selected between bison and cattle lineages. However, Gautier et al. (2016) were restricted to two modern genomes (BBO3569 and BBO3574), making it difficult to identify directional patterns of selection between lineages as they lacked genomic information from ancestral lineages. In addition, the founder/son pair of specimens included by Węcek et al. (2016) in order to include historic genomic information are still very recent examples of diversity in European bison (early 1900s) and it is known from the historical record that European bison populations were already significantly reduced by that time (Krasińska and Krasiński, 2013). This may complicate genetic analyses, as long term patterns of evolutionary history are often lost after such bottleneck events (Chang and Shapiro, 2016). A better understanding of the source and frequency of admixture between the two subspecies of European bison and between the bison and cattle/aurochs lineages will further our understanding of the evolutionary history of European bison, and aid the development of appropriate conservation management strategies.

The massive reduction in genetic diversity among extant European bison means that any signal of evolutionary processes prior to the bottleneck event (including events such as population mixing or separation) will be cryptic to phylogenetic analyses of those extant populations (Leonard, J. A., 2008; Chang and Shapiro, 2016). However, it is precisely those changes in genetic variation of ancient and historic populations that are the clues necessary for a more complete understanding of the evolutionary patterns and processes underlying the genetic structure of extant populations (Leonard, J. A., 2008; Leonard and Wayne, 2008). In an ideal situation, we would be able to ‘go back in time’ and study the extinct populations of European bison directly to compare them to extant

populations. While this is not possible, there are ancient subfossil and historical bone/skin specimens of European bison that can be analysed in order to give an indication of what ancient genetic diversity may have been like (Leonard, J. A., 2008; Steeves et al., 2010). With the analysis of ancient DNA (aDNA) samples, a more complete picture of the evolutionary processes underlying the genetic structure of the extinct and extant populations of European bison can be uncovered.

Ancient DNA: bridging the gap between palaeontology and genomics of living bison

Recently, two aDNA studies have been published regarding the evolutionary history of European bison and potential ancestral species (Soubrier et al., 2016; Massilani et al., 2016). Both studies included a small number of samples across a restricted geographic range (France and the Urals Mountains respectively), predominantly with mitochondrial D-loop DNA data and a subset of samples with whole mitochondrial genomes (WMGs). From these data, both groups independently identified the presence of a previously unknown mitochondrial clade of European bison distinct from – but closely related to – modern European bison. As a result, European bison specimens could be split into two clades, termed 'wisent' or 'Bb2', and 'CladeX' or 'Bb1' by Soubrier et al. (2016) and Massilani et al. (2016) respectively. Hereafter the two clades are referred to as wisent and CladeX, with the term European bison referring collectively to both clades. Soubrier et al. (2016) showed that representatives of this unknown clade demonstrated a pattern of population replacement between steppe bison and European bison over time in the Urals mountains that could be correlated with regional environmental changes from predominantly open steppe to forest, and suggested that this clade represented an 'ecotype' of ancient European bison. Specifically the previously unidentified clade (CladeX) appeared to share a preference for cold tundra-steppe with steppe bison, while wisent appeared to prefer open woodland environments (Soubrier et al., 2016; Massilani et al., 2016). However, neither study drew firm conclusions about the taxonomic identity of CladeX and wisent.

While both Soubrier et al. (2016) and Massilani et al. (2016) agreed on the identification of a new clade of European bison in the Late Pleistocene from D-loop and limited WMG data, each paper came to alternate conclusions regarding the origins of European bison. Massilani et al. (2016) were limited to mitochondrial information only and favoured ILS between *Bos* and *Bison* to explain the phylogenetic relationship of the European bison mitochondrial lineage to other bovid species. They cited overlapping mitochondrial divergence dates between European bison and *Bos taurus*, and between the *Bos* and *Bison* lineages they observed in their dataset, as evidence for their interpretation (Massilani et al., 2016). In contrast, Soubrier et al. (2016) included a subset of nuclear SNPs in their analyses, which suggested that European bison appeared most similar to American bison (*Bison bison*). In this respect, these ancient DNA studies recapitulated the results of previous studies using genetic data from modern populations. However, Soubrier et al. (2016) also detected a weak nuclear signature of introgression from aurochs into the European bison genome. Consequently, Soubrier et al. (2016) favoured a hybrid origin for European bison: where European bison are a result of introgression between aurochs and steppe bison with a 'cow-like' mitochondrial genome from the aurochs introgression but vast proportion of nuclear DNA remaining 'steppe-bison like' due to repeated back-crosses with 'pure' steppe bison males. This process may have led to a new species that had bison-like morphology and nuclear DNA markers, which retained a Bos-like mitochondrial genome from their maternal ancestor (Verkaar et al., 2004).

Some authors have proposed that CladeX actually represents the extinct bison species, *Bison schoetensacki* (Palacio et al., 2017). If true, this would provide evidence of *Bison schoetensacki* survival into the Late Pleistocene (not only the Middle Pleistocene as previously thought) and reinforce the theory that *Bison schoetensacki* was the direct ancestor of modern European bison instead of steppe bison (Palacio et al., 2017). Palacio et al. (2017) assembled a mitochondrial genome sequence phylogenetically similar to CladeX bison from DNA extracted from a cave hyena coprolite. The bison sequence from the hyena coprolite also matched a short 300bp fragment from a cannon bone morphologically identified as *Bison schoetensacki* from a nearby cave ($\sim$30km away). Because the short DNA fragment obtained from the cannon bone matched the sequence obtained from the cave hyena coprolite, researchers concluded that this sequence must therefore also be from *Bison schoetensacki* and consequently all other samples from the same phylogenetic clade should be re-classified as *Bison schoetensacki* (Palacio et al., 2017). However, cannon (ankle) bones are notoriously difficult to identify morphologically, and none of the samples identified as CladeX in either paper were considered morphologically representative of *Bison schoetensacki* (Soubrier et al., 2016; Massilani et al., 2016). Further, the morphological identity of this particular cannon bone remains contentious, and has been alternatively identified as *Bison schoetensacki* and a novel subspecies of *Bison priscus* (Guérin and Philippe, 1971; Brugal, 1999; Palacio et al., 2017). While these data certainly shows that the cave hyena coprolite contained bison DNA from a member of CladeX, without corresponding sequence data from a conclusively identified specimen of *Bison schoetensacki* it is impossible to

determine if CladeX should actually be considered *Bison schoetensacki*. Until otherwise demonstrated, CladeX should not be considered *Bison schoetensacki* and remain a clade of European bison.

Conclusions

Broadly, there are three key outstanding questions to further our understanding of the evolutionary history of European bison:

1. What is the relationship between CladeX and wisent?
2. Do the patterns of replacement between European bison and steppe bison hold across a broad geographic and temporal range?
3. What was the effect of the demographic bottleneck in the 1920s on genetic diversity of European bison?

The ancient DNA research on European bison described above was predominantly focused on mitochondrial D-loop data, supplemented with a handful of mitochondrial genomes. In addition, prior studies were temporally and geographically restricted, with Massilani et al. (2016) focused on regional France, and Soubrier et al. (2016) predominantly focused on samples predominantly from the Urals Mountains in Russia. Combining both datasets and further expanding the temporal and geographical range covered by the samples is important in order to ascertain if the patterns observed hold true across the Eurasian continent, or are region-specific. In parallel with expanding the sample coverage geographically and temporally, expanding the coverage of ecological and climate information that can be linked to the sample distribution is also important, and advances in palaeoclimate modelling allow us to predict environmental change between 120kyr and the present (Gordon et al., 2000; Pope et al., 2000; Kaplan et al., 2003, 2016; Binney et al., 2017; Fordham et al., 2017). By using specimens that are directly dated, explicit links can be drawn between genetic changes and local environmental and ecological changes such that regional and continental patterns can be observed.

In addition, including complete mitochondrial genomes (as opposed to mitochondrial D-loop regions) or nuclear DNA data will further improve the resolution of subsequent analyses over those from previous datasets. The small subset of nuclear SNPs used in Soubrier et al. (2016) are potentially biased due to their ascertainment from a cow SNP panel, but for the purposes of distinguishing between steppe-like and cow-like genomic patterns they are sufficient. However, an unbiased SNP panel, or full genomic level information would provide better support for introgression from a hybrid origin. The limited number of samples used for nuclear SNP analysis and their geographic and temporal restrictions can impact the results obtained, and additional genomic data in particular is required in order to untangle potential hybridisation or ILS origins for European bison. Finally, the addition of nuclear data prior to the 1920s bottleneck would allow characterisation of the true diversity loss during the near-extinction event, and consequently inform conservation management of the extant populations. In summary, the evolutionary history of European bison remains debated. Only further research expanding the geographical and temporal range of samples studied, as well as including additional mitochondrial and whole genomes for further analyses will allow researchers untangle the true evolutionary history of European bison.

REFERENCES

Babik, W., Kawałko, A., Wójcik, J. M., and Radwan, J. (2012). Low major histocompatibility complex class I (MHC I) variation in the European bison (Bison bonasus). *Journal of Heredity*, 103(3):349–359.

Benecke, N. (2005). The Holocene distribution of European bison, the archaeozoological record. (57):421–428.

Bibi, F. (2013). A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. *BMC Evolutionary Biology*, 13(1):166.

Binney, H., Edwards, M., Macias-Fauria, M., Lozhkin, A., Anderson, P., Kaplan, J. O., Andreev, A., Bezrukova, E., Blyakharchuk, T., Jankovska, V., Khazina, I., Krivonogov, S., Kremenetski, K., Nield, J., Novenko, E., Ryabogina, N., Solovieva, N., Willis, K., and Zernitskaya, V. (2017). Vegetation of Eurasia from the last glacial maximum to present: Key biogeographic patterns. *Quaternary Science Reviews*, 157(Supplement C):80–97.

Bocherens, H., Hofman-Kamińska, E., Drucker, D. G., Schmölcke, U., and Kowalczyk, R. (2015). European Bison as a Refugee Species? Evidence from Isotopic Data on Early Holocene Bison and Other Large Herbivores in Northern Europe. *PLOS ONE*, 10(2):e0115090.

Brugal, J. P. (1999). Étude de populations de grands Bovidés européens: intérêt pour la connaissance des comportements humains au Paléolithique. *Le bison: gibier et moyen de subsistance des hommes du Paléolithique aux Paléoindiens des Grandes Plaines. Antibes: Éditions APDCA*.

Chang, D. and Shapiro, B. (2016). Using ancient DNA and coalescent-based methods to infer extinction. *Biology Letters*, 12(2):20150822.

EBPB (2016). *European Bison Pedigree Book*. Białowieża National Park.

Fordham, D. A., Saltré, F., Haythorne, S., Wigley, T. M., Otto-Bliesner, B. L., Chan, K. C., and Brook, B. W. (2017). Paleoview: a tool for generating continuous climate projections spanning the last 21 000 years at regional and global scales. *Ecography*, 40(11):1348–1358.

Gautier, M., Moazami-Goudarzi, K., Levéziel, H., Parinello, H., Grohs, C., Rialle, S., Kowalczyk, R., and Flori, L. (2016). Deciphering the Wisent Demographic and Adaptive Histories from Individual Whole-Genome Sequences. *Molecular Biology and Evolution*, 33(11):2801–2814.

Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., Mitchell, J. F. B., and Wood, R. A. (2000). The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, 16(2-3):147–168.

Gralak, B., Krasińska, M., Niemczewski, C., Krasiński, Z. A., and Żurkowski, M. (2004). Polymorphism of bovine microsatellite DNA sequences in the lowland European bison. *Acta Theriologica*, 49(4):449–456.

Guérin, C. and Philippe, M. (1971). Les gisements de vertébrés pléistocènes du Causse de Martel. *Bull Soc Hist Archéol*, 93:31–46.

IUCN (2008). The IUCN Red List of Threatened Species. http://www.iucnredlist.org/ Accessed: 2018-02-01.

Kaplan, J. O., Bigelow, N. H., Prentice, I. C., Harrison, S. P., Bartlein, P. J., Christensen, T. R., Cramer, W., Matveyeva, N. V., McGuire, A. D., Murray, D. F., Razzhivin, V. Y., Smith, B., Walker, D. A., Anderson, P. M., Andreev, A. A., Brubaker, L. B., Edwards, M. E., and Lozhkin, A. V. (2003). Climate change and Arctic ecosystems: 2. Modeling, paleodata-model comparisons, and future projections. *Journal of Geophysical Research: Atmospheres*, 108(D19).

Kaplan, J. O., Pfeiffer, M., Kolen, J. C. A., and Davis, B. A. S. (2016). Large Scale Anthropogenic Reduction of Forest Cover in Last Glacial Maximum Europe. *PLOS ONE*, 11(11):e0166726.

Krasińska, M. and Krasiński, Z. (2013). *European Bison: The Nature Monograph*. Springer Berlin Heidelberg.

Kurtèn, B. (1968). *Pleistocene Mammals of Europe*. Routledge.

Leonard, J. A. and Wayne, R. K. (2008). Native Great Lakes wolves were not restored. *Biology Letters*, 4(1):95–98.

Leonard, J. A. (2008). Ancient DNA applications for wildlife conservation. *Molecular Ecology*, 17(19):4186–4196.

Luenser, K., Fickel, J., Lehnen, A., Speck, S., and Ludwig, A. (2005). Low level of genetic variability in European bisons (Bison bonasus) from the Białowieża national park in Poland. *European Journal of Wildlife Research*, 51(2):84–87.

Maniakas, I. and Kostopoulos, D. S. (2017). Assessing astragalar morphology and biomechanics in western Palaearctic Bison populations with geometric morphometrics. *Comptes Rendus Palevol*, 16(7):783–794.

Marsolier-Kergoat, M.-C., Palacio, P., Berthonaud, V., Maksud, F., Stafford, T., Bégouën, R., and Elalouf, J.-M. (2015). Hunting the Extinct Steppe Bison (Bison priscus) Mitochondrial Genome in the Trois-Frères Paleolithic Painted Cave. *PLOS ONE*, 10(6).

Massilani, D., Guimaraes, S., Brugal, J.-P., Bennett, E. A., Tokarska, M., Arbogast, R.-M., Baryshnikov, G., Boeskorov, G., Castel, J.-C., Davydov, S., Madelaine, S., Putelat, O., Spasskaya, N. N., Uerpmann, H.-P., Grange, T., and Geigl, E.-M. (2016). Past climate changes, population dynamics and the origin of Bison in Europe. *BMC Biology*, 14(1):93.

Olech, W. (2006). The influence of inbreeding on European bison sex ratio. *Animals, Zoo and Conservation. E. Zgrabczynska, P. Cwiertnia, J. Ziomek eds. Published by The Zoological Garden in Poznañ*, pages 29–33.

Palacio, P., Berthonaud, V., Guérin, C., Lambourdière, J., Maksud, F., Philippe, M., Plaire, D., Stafford, T., Marsolier-Kergoat, M.-C., and Elalouf, J.-M. (2017). Genome data on the extinct Bison schoetensacki establish it as a sister species of the extant European bison (Bison bonasus). *BMC Evolutionary Biology*, 17(1):48.

Pope, V. D., Gallani, M. L., Rowntree, P. R., and Stratton, R. A. (2000). The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3. *Climate Dynamics*, 16(2-3):123–146.

Pucek, Z. (2004). *European Bison (Bison Bonasus): Current State of the Species and Strategy for Its Conservation*. Council of Europe Publishing.

Radwan, J., Kawałko, A., Wojcik, J., and Babik, W. (2007). MHC-DRB3 variation in a free-living population of the European bison, Bison bonasus. *Molecular Ecology*, 16(3):531–540.

Shapiro, B., Drummond, A. J., Rambaut, A., Wilson, M. C., Matheus, P. E., Sher, A. V., Pybus, O. G., Gilbert, M. T. P., Barnes, I., Binladen, J., Willerslev, E., Hansen, A. J., Baryshnikov, G. F., Burns, J. A., Davydov,

S., Driver, J. C., Froese, D. G., Harington, C. R., Keddie, G., Kosintsev, P., Kunz, M. L., Martin, L. D., Stephenson, R. O., Storer, J., Tedford, R., Zimov, S., and Cooper, A. (2004). Rise and Fall of the Beringian Steppe Bison. *Science*, 306(5701):1561–1565.

Soubrier, J., Gower, G., Chen, K., Richards, S. M., Llamas, B., Mitchell, K. J., Ho, S. Y. W., Kosintsev, P., Lee, M. S. Y., Baryshnikov, G., Bollongino, R., Bover, P., Burger, J., Chivall, D., Crégut-Bonnoure, E., Decker, J. E., Doronichev, V. B., Douka, K., Fordham, D. A., Fontana, F., Fritz, C., Glimmerveen, J., Golovanova, L. V., Groves, C., Guerreschi, A., Haak, W., Higham, T., Hofman-Kamińska, E., Immel, A., Julien, M.-A., Krause, J., Krotova, O., Langbein, F., Larson, G., Rohrlach, A., Scheu, A., Schnabel, R. D., Taylor, J. F., Tokarska, M., Tosello, G., van der Plicht, J., van Loenen, A., Vigne, J.-D., Wooley, O., Orlando, L., Kowalczyk, R., Shapiro, B., and Cooper, A. (2016). Early cave art and ancient DNA record the origin of European bison. *Nature Communications*, 7:13158.

Steeves, T. E., Maloney, R. F., Hale, M. L., Tylianakis, J. M., and Gemmell, N. J. (2010). Genetic analyses reveal hybridization but no hybrid swarm in one of the world's rarest birds. *Molecular Ecology*, 19(23):5090–5100.

Tokarska, M., Kawałko, A., Wojcik, J. M., and Pertoldi, C. (2009a). Genetic variability in the European bison (Bison bonasus) population from Białowieża forest over 50 years. *Biological Journal of the Linnean Society*, 97(4):801–809.

Tokarska, M., Marshall, T., Kowalczyk, R., Wojcik, J., Pertoldi, C., Kristensen, T., Loeschcke, V., Gregersen, V., and Bendixen, C. (2009b). Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. *Heredity*, 103(4):326.

Tokarska, M., Pertoldi, C., Kowalczyk, R., and Perzanowski, K. (2011). Genetic status of the European bison Bison bonasus after extinction in the wild and subsequent recovery. *Mammal Review*, 41(2):151–162.

van de Vlasakker, J. (2014). *Rewilding Europe Bison Rewilding Plan, 2014-2024*. Rewilding Europe, Nijmegen, The Netherlands.

Vercoutère, C. and Guérin, C. (2010). Les Bovidae (Mammalia, Artiodactyla) du Pléistocène moyen final de l'Aven de Romain-la-Roche (Doubs, France). *Rev Paléobiol*, 29:655–96.

Verkaar, E. L., Nijman, I. J., Beeke, M., Hanekamp, E., and Lenstra, J. A. (2004). Maternal and paternal lineages in cross-breeding bovine species. Has wisent a hybrid origin? *Molecular Biology and Evolution*, 21(7):1165–1170.

Węcek, K., Hartmann, S., Paijmans, J. L. A., Taron, U., Xenikoudakis, G., Cahill, J. A., Heintzman, P. D., Shapiro, B., Baryshnikov, G., Bunevich, A. N., Crees, J. J., Dobosz, R., Manaserian, N., Okarma, H., Tokarska, M., Turvey, S. T., Wójcik, J. M., Żyła, W., Szymura, J. M., Hofreiter, M., and Barlow, A. (2016). Complex admixture preceded and followed the extinction of wisent in the wild. *Molecular Biology and Evolution*, 34(3):598–612.

Wojciechowska, M., Nowak, Z., Gurgul, A., Olech, W., Drobik, W., and Szmatoła, T. (2017). Panel of informative SNP markers for two genetic lines of European bison: Lowland and Lowland–Caucasian. *Animal Biodiversity and Conservation*, 40(1):17–25.

Wójcik, J., Kawałko, A., Tokarska, M., Jaarola, M., Vallenback, P., and Pertoldi, C. (2009). Post-bottleneck mtDNA diversity in a free-living population of European bison: implications for conservation. *Journal of Zoology*, 277(1):81–87.

Chapter 5

Untangling the evolutionary history of European bison (*Bison bonasus*)

Manuscript prepared for submission

Statement of Authorship

Title of Paper	Untangling the evolutionary history of European Bison (*Bison bonasus*)
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☒ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Manuscript prepared in publication format

Principal Author

Name of Principal Author (Candidate)	Ayla L. van Loenen
Contribution to the Paper	Designed experiments, performed laboratory experiments, compiled data and performed bioinformatics analyses. Analysed and interpreted results. Wrote the paper with comments from co-authors.
Overall percentage (%)	70
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 3/7/18

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 to 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, interpreted results, made comments on the manuscript.		
Signature		Date	3/7/18

Name of Co-Author	Bastien Llamas		
Contribution to the Paper	Designed experiments, interpreted results, made comments on the manuscript.		
Signature		Date	3/7/18

Name of Co-Author	Kieren Mitchell		
Contribution to the Paper	Interpreted results, made comments on the manuscript		
Signature		Date	3/7/18

Name of Co-Author	Rafal Kowalczyk		
Contribution to the Paper	Provided specimens, radiocarbon dates and isotope information, made comments on the manuscript		
Signature		Date	5/7/18

Untangling the Evolutionary History of European Bison (Bison bonasus)

Ayla L. van Loenen,*,[1] Kieren J. Mitchell,[1] Bastien Llamas,[1] Rafał Kowalczyk,[2] and Alan Cooper,[1]

[1]Australian Centre for Ancient DNA (ACAD), University of Adelaide, Australia

[2] Mammal Research Institute Polish Academy of Sciences, Stoczek 1, 17-230, Białowieża, Poland

***Corresponding author:** E-mail: aylalore22@gmail.com

Associate Editor:

Abstract

European bison (*Bison bonasus*) are the largest endemic vertebrates in Europe, and one of the few megafaunal species to have survived the mass megafaunal extinction during the Pleistocene/Holocene transition (12-9 thousand years ago). Untangling their evolutionary history would provide valuable information about the response of European megafauna to periods of rapid environmental change. However, a severe and recent population bottleneck obscures much of the population history that could be inferred from the genomes of modern individuals. While several studies have attempted to analyse ancient European bison populations directly using ancient DNA, their datasets were limited in temporal and geographic range. In this study we present the most comprehensive dataset of ancient European bison mitochondrial genomes to date, with 131 ancient bison samples from across the Eurasian continent covering over 50 thousand years. We reveal patterns of bison distribution and concurrent environmental changes across a broad geographical and temporal range. In particular, population expansions following periods of extensive forest reduction combined with a decrease in anthropogenic pressures suggest that European bison remain preferentially adapted to an open steppe environment through to the present day.

Key words: ancient DNA, European bison, *Bison bonasus*, wisent, evolutionary genetics

Article

Introduction

European bison (also known as wisent or *Bison bonasus*) are one of the few megafaunal species to have survived the Late Pleistocene/early Holocene megafaunal extinctions in Europe (Bocherens *et al.*, 2015; Cooper *et al.*, 2015). While formerly distributed across the continent during the Late Pleistocene and Holocene alongside the now extinct steppe bison (*Bison priscus*), European bison were restricted to two populations representing two subspecies by the end of the 19th century: one in the Białowieża forest in Eastern Europe (*Bison bonasus bonasus*) and the other in the Caucasus Mountains (*Bison bonasus caucasus*) (Pucek, 2004; Krasińska and Krasiński, 2013; Węcek *et al.*, 2016). By the 1920s, both populations were extinct in the wild, with only a handful of individuals - comprising the recent descendants of only 12 individuals - remaining in captivity (Pucek, 2004; Benecke, 2005; Krasińska and Krasiński, 2013). As a result of this severe bottleneck, modern European bison have significantly reduced genetic diversity (Wójcik *et al.*, 2009; Tokarska *et al.*, 2009, 2011; Massilani *et al.*, 2016; Soubrier *et al.*, 2016;

Węcek *et al.*, 2016). Such dramatic bottleneck events typically obscure long-term patterns of evolutionary history (Chang and Shapiro, 2016), and the evolution and palaeobiogeography of European bison consequently remain poorly understood.

Several recent studies have investigated the evolutionary history of European bison using ancient DNA (aDNA) (Soubrier *et al.*, 2016; Massilani *et al.*, 2016; Onar *et al.*, 2017; Marsolier-Kergoat *et al.*, 2015; Grange *et al.*, 2018). The data presented in these studies comprised predominantly mitochondrial D-loop data plus a handful of mitochondrial genomes (where most specimens were from either the Urals Mountains or France). These geographically and temporally limited data supported the division of ancient European bison into two distinct mitochondrial clades (wisent and CladeX in Soubrier *et al.* (2016); or Bb2 and Bb1 in Massilani *et al.* (2016) and Grange *et al.* (2018)). Hereafter, we will refer to these clades according to the Bb1/Bb2 nomenclature from Massilani *et al.* (2016) and Grange *et al.* (2018), where "Bb1" refers to the extinct Late Pleistocene population with a distinct mitochondrial clade and "Bb2" refers to the living European bison (and also ancient specimens that are more closely related to living European bison than they are to Bb1). It has additionally been suggested that Bb2, Bb1 and steppe bison potentially reflect distinct taxa adapted to specific ecological niches in the two regions covered by these studies (Urals mountains, Russia, and France) (Soubrier *et al.*, 2016; Massilani *et al.*, 2016). However, this hypothesis has not been tested at a continental scale, and the exact geographical and temporal distribution of individuals belonging to the two mitochondrial clades remains to be characterised.

In this paper, we present 81 new ancient mitochondrial genome sequences from bison specimens from across Eurasia spanning over 50 thousand years, substantially increasing both the spatial and temporal coverage of the previously available data. By adding these new data to previously published sequences we assembled a total dataset comprising 131 bison mitochondrial genomes. In addition, each new specimen is radiocarbon dated and associated with dietary isotopic information. Using the increased resolution provided by this dataset, we compared our refined reconstruction of the evolutionary history of European bison to concurrent environmental and climatic changes.

Results

Sequencing results

Seventy-nine mitochondrial genomes were successfully sequenced with more than 70% coverage and an average depth $\geq$3. Two additional partial mitochondrial genomes were sequenced with coverage between 40-70% and average depth greater than 0.4, which was sufficient to identify the taxonomic affiliation of both specimens (A17406 and A17482). Damage patterns were

consistent with ancient DNA (Supplementary Table 1), with additional details of sequencing and mapping summary statistics output from PALEOMIX given in Supplementary Table 2. Specimens were assigned to clades of European bison for subsequent analyses according to their position in an initial ML tree (Supplementary Figure 1). Out of the 131 Eurasian bison included in this study, 102 specimens were directly dated and ranged in age from greater than 50,000 calBP (calibrated radiocarbon years before present) to modern day (Supplementary Table 3). The remaining 29 specimens were undated, but 22 came from well-characterised archaeological layers that provided an associated date range for the specimens.

Phylogenetic analysis

The median-joining haplotype network (Figure 2) clearly shows three distinct groups within the mitochondrial genome diversity: two clades of European bison (referred to as Bb1 and Bb2 as per Massilani *et al.* (2016) and Grange *et al.* (2018)), and steppe bison. There was no obvious genetic structure within the European bison clades (i.e. no obvious geographic clusters, and no distinctions between *B. b. bonasus* and *B. b. caucasus*). All three groups (steppe bison, Bb1 and Bb2) were present in Europe until the Pleistocene/Holocene transition (∼9-12 kya), at which point there was a clear reduction in bison diversity in Europe with only Bb2 surviving until the present (Figure 2, Supplementary File 1). The GMRF (Gaussian Markov Random Field) Bayesian skyride plot of Bb1 specimens did not show any evidence of population change over time prior to their extinction, while the GMRF Bayesian skyride plot of Bb2 specimens suggested a population decline starting ∼10 kya (Figure 3). The two distinct European bison clades are significantly less differentiated from each other than either is from the steppe bison (81 base changes between Bb1 and Bb2, compared to 286 base changes between steppe bison and Bb2). However, the available mitochondrial data is not sufficient to either support or exclude the hypothesis that the two European bison clades represent distinct species. For the purposes of subsequent analyses and discussion, Bb1 and Bb2 are considered the same species.

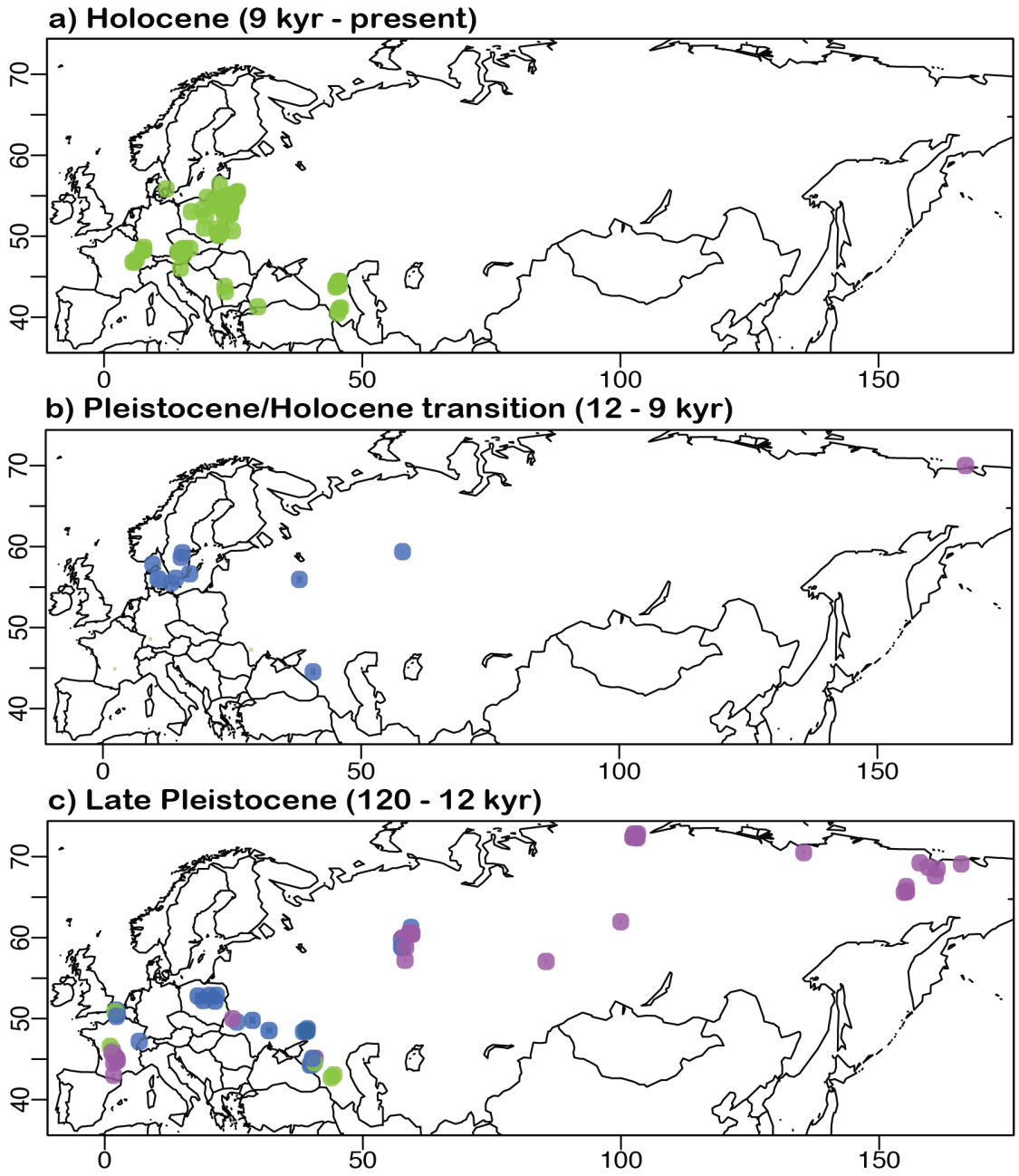


Figure 1. Temporal distribution maps of bison in Eurasia (Bb1 in blue, Bb2 in green, and steppe bison in purple). Sample dates are calibrated radiocarbon dates.

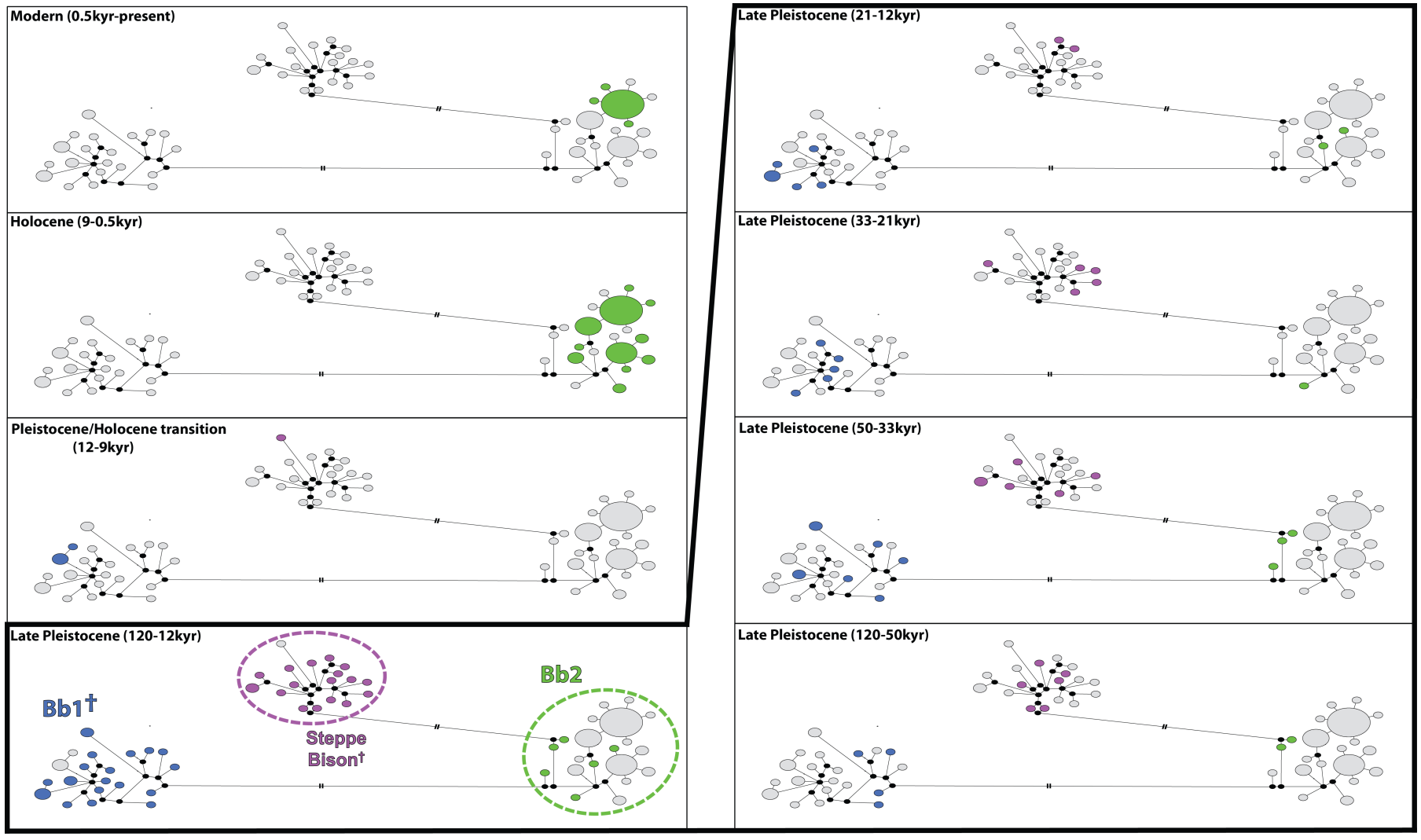


Figure 2. Median Joining haplotype network based on the alignment of 131 Eurasian bison (steppe bison, Bb1 and Bb2 mitochondrial genome sequences. Each circle represents a sampled haplotype, and the size of the circle is proportional to the frequency of that haplotype within the data set. The length of the lines between haplotypes is proportional to the number of base changes between haplotypes (double diagonal lines indicate line interruption due to space constraints). There were 286 base changes between steppe bison and Bb2, and 81 base changes between Bb1 and Bb2. The colours of the circles correspond to the species or clade of the specimen (Bb1 in blue, Bb2 in green, and steppe bison in purple), while black dots represent unsampled haplotypes. The network has been sliced into time periods, greyed out circles represent a sampled haplotype that was not found during the particular time period.

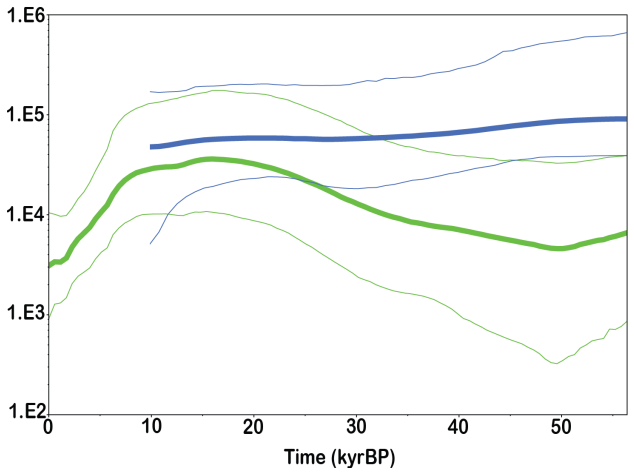


Figure 3. GMRF Bayesian skyride plots of estimated median female effective population size in European bison (thick lines, Bb1 in blue and Bb2 in green) with 95% HPD bounds (thin lines).

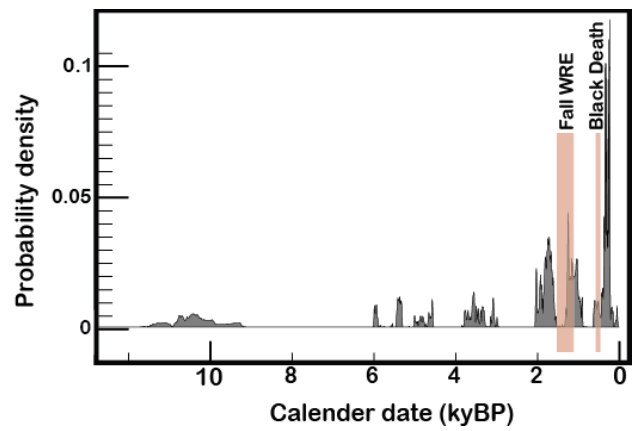


Figure 4. OxCal probability density plot of Holocene European bison radiocarbon dates. Highlighted are two major historical events that are known to have had a significant impact on the human population of Europe (Fall of the Western Roman Empire, and the Black Death).

The MCC (Maximum Clade Credibility) tree generated in BEAST supports the division of European bison into Bb1 and Bb2, and additionally shows that the two clades diverged ∼96.8kya (95% Highest Posterior Density HPD: 113.2-78.8kya). This contrasts with the older divergence dates previously estimated by Massilani *et al.* (2016) (∼246kya: 95% HPD: 283-212kya) and Grange *et al.* (2018) (∼395 kya: 95% HPD: 445-343 kya), but falls within the estimates of Soubrier *et al.* (2016) (∼120kya: 95% HPD: 95-152kya) (Figure 5). The two samples A17406 and A17482 with low sequencing coverage fell within Bb1 diversity in both the initial ML tree generated using PhyML and the constrained tree from RaxML (Supplementary Figures 1 and 2).

Global distribution and ecology

Overlapping geographical distributions of steppe bison, Bb1, and Bb2 can clearly be observed across the continent throughout the Late Pleistocene (Figure 1, Supplementary File 1). After the Pleistocene/Holocene transition, steppe bison are no longer observed in Europe, but survive in Siberia until at least ∼7,000 calBP (Boeskorov *et al.*, 2014; Kirillova *et al.*, 2015). During the early Holocene, Bb1 appear to shift northward with the last identified Bb1 specimen at ∼9,776 calBP in Sweden. No Bb2 specimens were observed during the Late Pleistocene/early Holocene transition or during the early Holocene. However, Bb2 specimens were observed again by the mid-Holocene, initially in Poland ∼6,238 calBP (A17987), before quickly re-appearing across the rest of the European continent (Supplementary File 1). By the late Holocene, Bb2 appeared restricted predominantly to the Caucasus Mountains and the Białowieża forest (Poland), before their extinction in the wild by the 1920s (Supplementary File 1). Both Lowland and Caucasus bison (*B.b. bonasus* and *B.b. caucasus* respectively) fall together in the

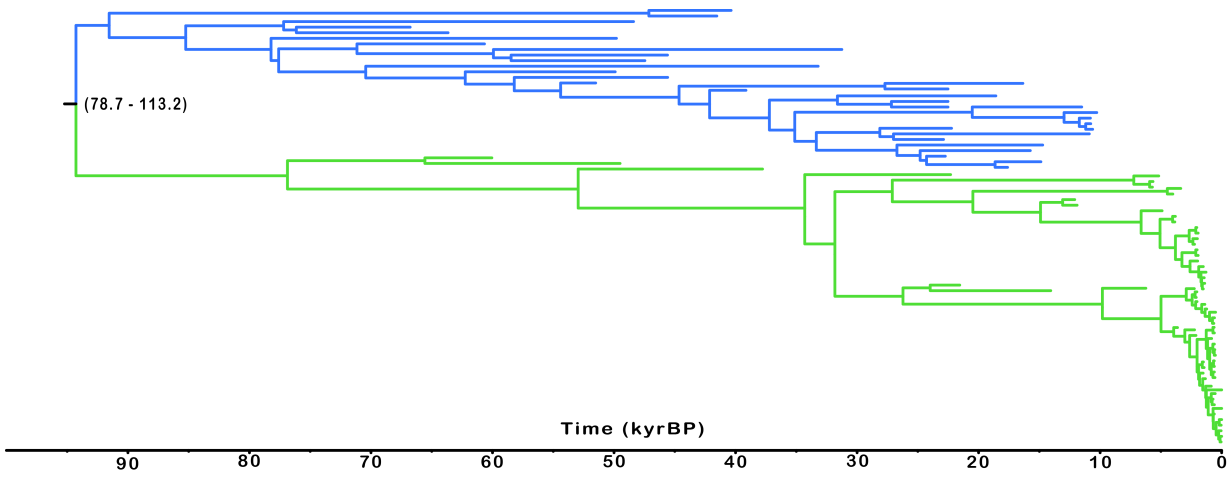


Figure 5. BEAST maximum clade credibility tree of European bison (*Bison bonasus*) mitochondrial genome sequences, calibrated using radiocarbon dates of the specimens. Height of the nodes represents median posterior age. Colours correspond to the clade of the specimens (Bb1 in blue and Bb2 in green).

same mitochondrial clade, and can therefore be collectively classified as Bb2 on the basis of their mitochondrial DNA.

Isotopic data from bison in Europe has been used to analyse their distribution within forested or non-forested landscape over time (Figure 6). Bb1 and steppe bison appear to have had a predominantly graze heavy diet corresponding to an open/steppe habitat (higher proportions of δ13C and δ15N), while Bb2 seems to have had a predominantly 'mixed-feed' diet corresponding to browsing and grazing across a dense/mosaic forest habitat (lower proportions of δ13C and δ15N).

Discussion

The broad geographic and temporal range covered by our expanded dataset provides the first opportunity to delineate clear patterns of bison presence across Europe during the Late Pleistocene and early Holocene. The results of our phylogenetic analyses are consistent with recorded history and illustrate a dramatic loss in bison genetic diversity across the Eurasian continent over time (both within and between clades) (Figures 2 and 5). For example, thirteen mitochondrial haplotypes of European bison were found during the early Holocene, which decreased to four haplotypes by the late Holocene, with only one haplotype remaining in present day populations (Figure 2). Our expanded dataset shows that Bb2 with a broad genetic diversity were regularly present across Europe prior to the Holocene, as opposed to the previously reported isolated occurrences observed in the Caucasus region (Soubrier *et al.*, 2016). The overlapping highest posterior densities for population size through time for Bb1 and Bb2 populations estimated from our Bayesian skyride analysis suggests that both populations were present in comparable numbers during the Late Pleistocene (Figure 3). While we sampled more Bb1 individuals than Bb2 during the Late Pleistocene, this may reflect a sampling bias due to

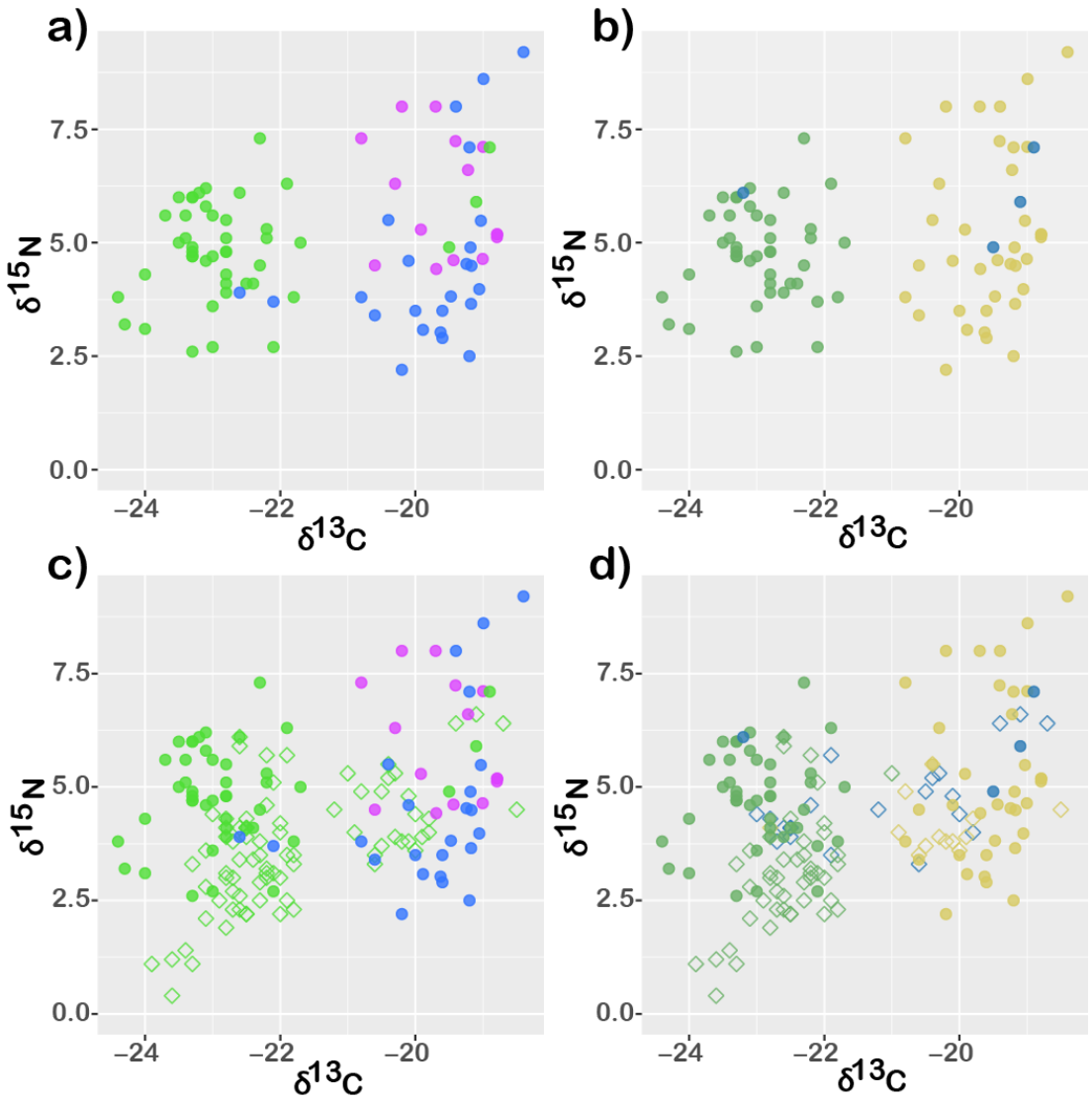


Figure 6. Plots of available isotope values for ancient specimens, and previously published modern Bb2 data from Hofman-Kamińska *et al.* (2018)). Panels A and B contain ancient bison specimens only (represented by coloured circles), while panels C and D also include modern Bb2 specimens (represented by open coloured diamonds) from Hofman-Kamińska *et al.* (2018) which have a known percentage forest cover per population. Colours in panels A and C indicate the species or clade of bison (Bb1 in blue, Bb2 in green, steppe bison in purple). Colours in panels B and D indicate three categories of forest cover (forest in green, non-forest in yellow, and mixed in dark blue.

geographical distribution or sample preservation biases as opposed to different population sizes.

The clear mitochondrial differentiation between Bb2 and Bb1 is well supported by both the haplotype network (Figure 2) and PhyML ML/BEAST MCC trees (Figure 5, Supplementary Figure 1), and the expanded dataset provides a refined estimated divergence time for the two clades of ∼96.8kya (95% HPD 113.2-78.8kya). This is comparable to divergence times between steppe bison mitochondrial lineages (Froese *et al.*, 2017), which suggests that Bb1 and Bb2 are two clades of the same species, at least based on the available mitochondrial data. However, we note that the data cannot unequivocally support that the mitochondrial clades are not different species. Additional mitochondrial and genomic data from ancient Bb1 and Bb2 specimens is required to

further address this question. For example, we may be missing extinct mitochondrial lineages that would reduce the differentiation between the two clades, or genomic data may suggest that the autosomal genetic makeup is not as clearly differentiated between the two clades as the mitochondrial haplotypes.

Bb1 individuals appear to have been restricted to the northern regions of continental Europe during the Late Pleistocene while Bb2 individuals were distributed in southern regions, although there was a broad zone (from France across to the Caucasus Mountains) in which individuals belonging to each mitochondrial clade overlapped (Supplementary File 1). Dietary isotope data suggested that the more southerly Bb2 lived in dense forest in a typically warm/humid environment, alternating between browsing and grazing (with the exception of three Late Pleistocene Bb2). In contrast, the Bb1 and steppe bison individuals, which had more northerly distributions where the environment was cold/dry, showed isotopic signatures characteristic of open habitat grazing such as mammoth steppe or open/mosaic forest (with the exception of two Holocene Bb1 individuals). In concordance with the isotopic data, Late Pleistocene steppe bison and Bb1 were typically distributed in or bordering on non-forested regions, while early Holocene Bb1 individuals (restricted to Scandinavia) were associated with a dense forest environment. Similarly, Late Pleistocene Bb2 specimens were found across a mixed forest/steppe environment while Holocene Bb2 occupied predominantly forested regions.

The stark contrast between dietary isotope signatures from Late Pleistocene and early Holocene bison individuals is consistent with the predicted forest cover across their respective geographical ranges. During the Late Pleistocene, repeated glaciation events restricted cold-sensitive plant species (such as deciduous trees) to glacial refuges throughout Europe, resulting in a 'mosaic' forest-steppe environment (Binney *et al.*, 2017; Leuschner and Ellenberg, 2017; Kaplan *et al.*, 2016). Following deglaciation of the European Ice Sheet Complex (EISC) at the end of the Younger Dryas/Weichselian Glacial (beginning of the Holocene ∼11.6 kya), continuing cold temperatures and seasonal dryness restricted forest expansion for the first two millennia of the Holocene, after which deciduous forest species rapidly expanded to cover the continent from ∼10 kya onwards until it reached maximum extent ∼6.7 kya (Marquer *et al.*, 2014; Theuerkauf *et al.*, 2014; Birks and Tinner, 2016; Patton *et al.*, 2017; Leuschner and Ellenberg, 2017; Binney *et al.*, 2017). In addition, as the EISC retreated there were two migrations of early human populations into Scandinavia, with one group moving northwards through Germany and Denmark, and the second moving southwards down the ice-free Norwegian Atlantic coast (Günther *et al.*, 2018). As a result, the early

Holocene Scandinavian Bb1 populations would have been restricted to a region situated between a rapidly expanding forest, arriving human populations, and the remaining icesheets.

It may be that the different distribution of Bb1 and Bb2 individuals during the early Holocene, while forests were expanding, drove their differential survival. Previous behavioral studies have suggested that Late Pleistocene European bison populations were likely comprised of fairly sedentary herds, with limited ranges and little to no inter-seasonal movement, subsisting on a sub-optimal diet (e.g. lichens) during winter or short periods of forest expansion if necessary (Kowalczyk *et al.*, 2011; Julien *et al.*, 2012; Velivetskaya *et al.*, 2016; van Asperen and Kahlke, 2017). Similarly, modern Bb2 demonstrate dietary plasticity between populations, and are able to survive on a sub-optimal high-browse diet (such as forest environment in winter) (Kerley *et al.*, 2012; Hofman-Kamińska *et al.*, 2018). However, European bison remain poorly adapted to a dense forest environment (Bocherens *et al.*, 2015; Hofman-Kamińska *et al.*, 2018). Collectively, this suggests that while European bison were able to survive in a sub-optimal forested environment during the Late Pleistocene and early Holocene, they were still dependent on access to un-forested 'graze'. This may explain both the apparent absence of Bb2 across mainland Europe during the heights of forest expansion in the early Holocene, and the extinction of Bb1 (the last recorded Bb1 specimen was found in Sweden ∼9,776 calBP, A17482). The restricted range of Bb1 during the Late Pleistocene/early Holocene means they were likely subsisting on a sub-optimal high-browse diet, as reflected by their dietary isotope signatures, for a long period. Prolonged occupancy of a densely forested environment when coupled with increased anthropomorphic pressures may have driven the extinction of Bb1, while Bb2 populations happened to be south of the expanding forest (possibly due to habitat preference), and were potentially able to retreat into nearby mountain regions unhampered by large ice sheets.

Following the early Holocene forest expansion, bison specimens were not detected in Europe until Bb2 re-appeared in Poland around 6,238 calBP (A17987), after which additional bison individuals appeared more widely across the continent. This occurred in parallel with the start of the mid-Holocene elm decline ∼6,347 years ago, which was likely the result of an early form of Dutch elm disease (6,347-5,281 years ago) (Perry and Moore, 1987; Parker, A. G. *et al.*, 2002; Batchelor *et al.*, 2014), as well as the Neolithic expansion and onset of forest clearance by humans that continues to the present day. However, it was not until the Iron Age (2,650-2,450 years ago) and the development of the iron ploughshare that widespread forest clearing started (Leuschner and Ellenberg, 2017). A brief respite from clearing during the fall of the Western Roman Empire between 1,600-1,350 years

ago allowed temporary forest re-expansion, but was followed by the beginning of systematic forest clearance through the Middle Ages (1,350-450 years ago) which continued through to present day, interrupted only by brief periods of regeneration during the Black Death pandemic and periods of civil unrest/wars (Leuschner and Ellenberg, 2017). These periods of extensive forest reduction can be correlated with the temporal distribution of Bb2 specimens identified across Europe (Figure 4); immediately prior to and during the collapse of the Western Roman Empire there are an increased number of Bb2 specimens in our dataset. A similar increase in the number of Bb2 specimens occurs during and immediately after the Black Death pandemic (604-596 years ago), during which time as much as 60% of the human population in Europe was killed (Gottfried, 1983; Benedictow, 1992; Olea and Christakos, 2005).

Our observations suggest that the reduction in dense forest cover during the mid-late Holocene increased the available range for Bb2, and when combined with a reduction in anthropomorphic pressure (e.g. through civil unrest, wars and/or pandemics) Bb2 populations were able to temporarily increase across the landscape. Interestingly, the temporal resolution appears to be sufficient to record an increase in Bb2 specimen numbers immediately prior to and during the collapse of the Western Roman Empire, versus an increase during and immediately after the Black Death armies returning home following civil unrest/war and immediately increasing anthropogenic forces on the surrounding environment, whereas human populations require several generations to recover equivalent sizes following a pandemic. The recovery of Bb2 populations following even short term reductions in forest cover or anthropogenic forces, supports models that megafaunal species (and bison in particular) were able to cope with either the rapid environmental changes or sporadic anthropogenic forces associated with the Late Pleistocene, but struggled under the combination of both stressors during the Holocene.

Conclusion

The survival of European bison through the megafaunal extinctions characteristic of the Late Pleistocene/early Holocene transition is likely in part due to their adaptive capacity for sub-optimal diet (Gautier *et al.*, 2016). This dietary flexibility is a relatively unique trait among other large mammals, but is shared with the few other megafauna that survived well into the Holocene in Europe, including aurochs (*Bos primigenius*), reindeer (*Rangifer tarandus*) and moose (*Alces alces*) (Bocherens *et al.*, 2015; Hofman-Kamińska *et al.*, 2018). Despite the fact that bison populations had an apparent high tolerance for sub-optimal diet, they appear to have been unable to sustain that diet for the length of time during which extremely dense forests covered the majority

of the continent, and slowly became extinct (leading to the extinction of Bb1 following the late Pleistocene/Holocene transition). Following the mid-Holocene elm decline and neolithisation of Europe, Bb2 appear to have recolonised the continent. Population expansions following periods of extensive forest reduction combined with a decrease in anthropogenic pressures suggest that Bb2 remain preferentially adapted to an open steppe environment through to the present day, which is supported by numerous morphological, behavioural and isotopic studies of modern bison (Mendoza and Palmqvist, 2008; Kerley *et al.*, 2012; Kowalczyk *et al.*, 2013; Bocherens *et al.*, 2015; Hofman-Kamińska *et al.*, 2018).

Materials and Methods

Sample collection and processing

Samples of bison specimens were collected from museums and private collections across Europe (Supplementary Table 3) to complement an existing collection at the Australian Centre for Ancient DNA (ACAD), University of Adelaide. The final sample set includes specimens from over 50 thousand years ago (kya) to present across regions from Scandinavia and Siberia (Russia) to Italy and Turkey (Figure 1). All ancient DNA work was performed in the purpose-built ancient DNA facilities at ACAD following previously published guidelines (Cooper and Poinar, 2000; Hofreiter and Shapiro, 2012). Ancient DNA was extracted from bone fragments following the silica based extraction protocol from Brotherton *et al.* (2013) with modifications as described in Soubrier *et al.* (2016). Double-stranded Illumina sequencing libraries were built from 25μl of DNA extract following the partial uracil-DNA-glycosylase (UDG) treatment protocol from Rohland et al. (2015) modified to include dual 7-mer internal barcode sequences as described in Soubrier *et al.* (2016). We used RNA baits targeting bison mitochondrial genome sequences and in-solution hybridisation capture followed by high-throughput sequencing as described in Soubrier *et al.* (2016) to generate complete or near-complete mitochondrial genomes from 81 European bison samples.

Sequencing reads were mapped to published mitochondrial genomes of European bison (chrM from NC014044) or steppe bison (chrM from NC012346) according to the morphological assignment of bones (unidentified bones were mapped to steppe bison). We used the programs AdapterRemoval v2.2.1 (Lindgreen, 2012), BWA v0.7.15 (Li *et al.*, 2009), Picard Tools (http://broadinstitute.github.io/picard), Genome Analysis ToolKit (McKenna *et al.*, 2010), SAMtools v1.3.1 (Li *et al.*, 2009; Li, 2011), and mapDamage v2.0.6 (Jónsson *et al.*, 2013), as implemented in the pipeline PALEOMIX v1.2.9 (Schubert *et al.*, 2014) and as described in Soubrier *et al.* (2016). The summary sequencing statistics for each specimen were compiled from the PALEOMIX pipeline output (Supplementary Table 2). A mitochondrial genome consensus

sequence was generated in Geneious v9.0.5 (Biomatters; http://www.geneious.com) with a minimum read depth of 3 and a consensus call threshold of 75%. Any ambiguities were visually inspected and manually realigned if necessary, and the resulting 81 consensus sequences were aligned to previously published bovid mitochondrial sequences using MAFFT7 (Katoh *et al.*, 2002) to create a multiple sequence alignment (MSA). A Maximum-Likelihood tree (ML tree) was generated in PhyML v3.0 (Guindon *et al.*, 2010) from the MSA (substitution model: HKY+G6, data type: nucleotide, tree improvement: BEST) to classify specimens as steppe bison, Bb1, or Bb2. In cases where the ML tree species assignment did not match the initial morphological assignment, sequencing reads from that specimen were re-mapped to the appropriate reference genome and variants called as described above.

Sequence data from the 81 new specimens were combined with 52 published European bison mitochondrial genomes (i.e. all available specimens from Eurasia identified as either Bb1, Bb2, *B. bonasus* or *B. priscus*) available on NCBI and ENA. Two genome sequences were not included, one was chimeric (KY055664, Węcek *et al.* (2016)), and the other being sourced from a hyena coprolite (GAO1/KU886087, Palacio *et al.* (2017)) and with a contentious taxonomic origin (Grange *et al.*, 2018). The resulting final dataset included 131 bison mitochondrial genomes from Eurasia.

Phylogenetic analyses

Median-joining haplotype networks were generated in PopART (Leigh and Bryant, 2015; Bandelt *et al.*, 1999) from the alignment of all mitochondrial genomes, with sites missing more than 5% of the data masked from the analysis. Bayesian phylogenetic analyses were performed from the MSA of all complete mitochondrial genomes of European bison included in this study using BEAST v1.8.4 (Suchard and Rambaut, 2009; Drummond and Rambaut, 2007) through the CIPRES Science Gateway v3.1 (HKY+G6 nucleotide substitution model, GMRF skyride, uncorrelated relaxed clock).The ages of specimens with infinite radiocarbon error margins were treated as parameters to be estimated in the model. Steppe bison mitochondrial genomes were not included from Bayesian phylogenetic analyses, as the GMRF skyride coalescent tree prior implemented in BEAST requires all samples to be drawn from a single species. Three Markov Chain Monte Carlo (MCMC) chains were run for 100 million generations sampling every 10,000 generations. The first 10% of generations were discarded as burn-in, with the remaining 90% combined using LogCombiner v1.8.3, and all parameters showed convergence when inspected in Tracer v1.5 (Rambaut and Drummond, 2009). The resulting maximum clade credibility tree (MCC tree) was annotated in TreeAnnotator v1.8.4 and visualised in FigTree v1.4.2 (Drummond and Rambaut, 2007). Due to

the relatively distinct geographical distributions of the Bb1 and Bb2 populations, it is likely that they were not randomly mating, which violates the assumptions of Bayesian skyride (Ho and Shapiro, 2011). As a result, we ran two separate skyride analyses, one using the alignment of all Bb1 specimens, and one using the alignment of all Bb2 specimens. Both skyride plots were generated under the GMRF 'skyride' analysis option in Tracer v1.5 (Rambaut and Drummond, 2009; Minin *et al.*, 2008). For two samples with low coverage (A17406 and A17482), the BEAST maximum credibility tree was used as a backbone constraint in RAxML v8 (Stamatakis, 2014) and the position of both samples estimated via ML.

Spatial and ecological patterns

Time slice maps of bison distribution in Europe at 1-2 thousand year (ky) intervals (2 ky slices between 72-22 kya, 1 ky slices between 22 kya-present) were generated in QGIS (v2.14 Essen, QGIS Development Team, 2016), using a 1:10m Shaded Relief raster map layered with cultural vectors (countries) and a time series of palaeobiome predictions in the form of forest/non-forest/sea classes. Raster and vector map data files were sourced from Natural Earth (naturalearthdata.com). The palaeobiome predictions were generated by (Hoogaker *et al.*, 2016) from HadCM3 model data using the BIOME4 vegetation model to infer the dominant biome for each grid cell from functional plant types, and using this data to classify 'forest' or 'non-forest' grid cells for a series of palaeoenviromental snapshots covering the periods from 120 kya to present (Kaplan *et al.*, 2003, 2016; Binney *et al.*, 2017). Ancient specimens were recorded as present across time slices according to the 95% confidence interval range of their associated calibrated radiocarbon dates or archaeological layers. The exported series of maps were stitched together using ffmpeg (available from https://github.com/FFmpeg/FFmpeg) to generate a videographic of the data (Supplementary File 1).

Each ancient specimen was also defined for a single overall category of forest cover across the 95% confidence interval of the calibrated radiocarbon date (forest, non-forest and mixed) (Figure 6). Specifically, 'forest' defined specimens which were found exclusively in forest grid cells across their date range, 'non-forest' defined specimens which were found exclusively in non-forest grid cells across their date range, and 'mixed' defined specimens which were found in both forest and non-forest grid cells across their date range. Modern Bb2 specimens from Hofman-Kamińska *et al.* (2018) which had a known percentage of forest cover per population were also included in our analyses, and were defined into the same categories of forest cover where 'forest' defined a sample with more than 70% forest cover, 'non-forest' defined a specimen with less than 50%

forest cover, and 'mixed' defined a specimen with between 50-70% forest cover.

Stable isotope data

Samples were radiocarbon dated at the Laboratory of Ion Beam Physics, Eidgenössische Technische Hochschule Zürich, Switzerland (Swiss Federal Institute of Technology Zürich) (ETH), and the dates were calibrated with OxCal v4.3 using the IntCal13 curve (Supplementary Table 3) (Reimer *et al.*, 2013; Ramsey, 2009). A probability density plot was generated in OxCal v4.3 from the Holocene dated samples (Ramsey, 2009). Stable isotope data (δ13C and δ15N) were also generated for the samples at the Department of Geosciences at the University of Tübingen (Germany) using an elemental analyzer NC 2500 connected to a Thermo Quest Delta+XL mass spectrometer as described in Hofman-Kamińska *et al.* (2018). Previously published isotope information for European bison for which the mitochondrial clade (Bb1 or Bb2) was known were also collated and included in subsequent analyses (Bocherens *et al.*, 2015; Soubrier *et al.*, 2016; Hofman-Kamińska *et al.*, 2018). Stable isotope values were plotted in combination with estimates of forest cover (described above) using the ggplot2 package in R (R Development Core Team, 2017).

Supplementary Material

Supplementary Figure 1. Maximum-Likelihood tree generated in PhyML to classify specimens as steppe bison, Bb1, or Bb2 (with Water buffalo, Yak, American bison and Cow mitochondrial genomes for reference). Tips are labelled with sample ID, clade ID and location.

Supplementary Figure 2. RaXML tree constrained using the BEAST MCC tree, adding two low coverage specimens (highlighted in red: A17406 and A17482, both falling within the diversity of Bb2). Tips are labelled with sample ID, clade ID, locations and age (in mean calibrated years BP)

Supplementary Table 1. MapDamage values

Supplementary Table 2. PALEOMIX summary values

Supplementary Table 3. Sample information

Supplementary File 1. Videographic of bison distribution across Eurasia through time with a forest/non-forest paleoenvironmental background.

Acknowledgments

This research was supported by the Australian Research Council, the Polish National Science Centre (2013/11/B/NZ8/00914 and N N304 301940) and the European Commission's project No. PIRSES-GA-2009-247652, BIOGEAST within the Seventh Framework Programme. We thank the following for providing access to specimens for sampling: L. Chaix, C. Cupillard, Dr. M. Nussbaumer, E. Keczyńska-Moroz, Dr. L. Wickström, Prof. K. Rauscher, Dr. A. A. Krotova, A. T. Halamski, J. Kobylińska, Prof. D. Anatolie, V. Rusu, Prof. M. Arakelyan, Dr. habil. E. Szuma, A. Stępień, H. Długoszewska-Nadratowska, Dr. M. Krajcarz, B. Stachowiak, Prof. B. Studencka,

I. Lorek, Dr. N. Spassov, A. Archacka, M. Szymkiewicz, Dr. J. Jastrzębski, J. Deptuła, A. D. Ya, B. Antoniuk, A. Sepioł, Dr. M. Križnar, Dr. N. Czeremnyh, Prof. P Kjellander, A. Hultman, V. Gedminas, Dr. K. T. Victorovna, Dr. U. B. Göhlich, Dr. habil. F. E. Zachos, K. Berggren, and K. M. Gregersen. ALvL was supported by an Australian Government Research Training Program (RTP) Scholarship. K.J.M, B.L., and A.C. are supported by the Australian Research Council. BAM files and consensus sequences for each specimen are available at Zenodo (doi:10.5281/zenodo.xxxxxxx).

References

Bandelt, H. J., Forster, P., and Röhl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16(1): 37–48.

Batchelor, C. R., Branch, N. P., Allison, E. A., Austin, P. A., Bishop, B., Brown, A. D., Elias, S. A., Green, C. P., and Young, D. S. 2014. The timing and causes of the Neolithic elm decline: New evidence from the Lower Thames Valley (London, UK). *Environmental Archaeology*, 19(3): 263–290.

Benecke, N. 2005. The holocene distribution of european bison-the archaeozoological record. *Munibe Antropologia-Arkeologia*, (57): 421–428.

Benedictow, O. J. 1992. *Plague in the late medieval Nordic countries: epidemiological studies*. Middelalderforlaget.

Binney, H., Edwards, M., Macias-Fauria, M., Lozhkin, A., Anderson, P., Kaplan, J. O., Andreev, A., Bezrukova, E., Blyakharchuk, T., Jankovska, V., Khazina, I., Krivonogov, S., Kremenetski, K., Nield, J., Novenko, E., Ryabogina, N., Solovieva, N., Willis, K., and Zernitskaya, V. 2017. Vegetation of Eurasia from the last glacial maximum to present: Key biogeographic patterns. *Quaternary Science Reviews*, 157(Supplement C): 80–97.

Birks, H. J. B. and Tinner, W. 2016. *Past forests of Europe*. Publication Office of the European Union.

Bocherens, H., Hofman-Kamińska, E., Drucker, D. G., Schmölcke, U., and Kowalczyk, R. 2015. European Bison as a Refugee Species? Evidence from Isotopic Data on Early Holocene Bison and Other Large Herbivores in Northern Europe. *PLOS ONE*, 10(2): e0115090.

Boeskorov, G. G., Potapova, O. R., Mashchenko, E. N., Protopopov, A. V., Kuznetsova, T. V., Agenbroad, L., and Tikhonov, A. N. 2014. Preliminary analyses of the frozen mummies of mammoth (Mammuthus primigenius), bison (Bison priscus) and horse (Equus sp.) from the Yana-Indigirka Lowland, Yakutia, Russia. *Integrative Zoology*, 9(4): 471–480.

Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Jane Adler, C., Richards, S. M., Sarkissian, C. D., Ganslmeier, R., Friederich, S., Dresely, V., van Oven, M., Kenyon, R., Van der Hoek, M. B., Korlach, J., Luong, K., Ho, S. Y. W., Quintana-Murci, L., Behar, D. M., Meller, H., Alt, K. W., Cooper, A., and Consortium, T. G. 2013. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nature Communications*, 4: 1764.

Chang, D. and Shapiro, B. 2016. Using ancient DNA and coalescent-based methods to infer extinction. *Biology Letters*, 12(2): 20150822.

Cooper, A. and Poinar, H. N. 2000. Ancient DNA: do it right or not at all. *Science*, 289(5482): 1139–1139.

Cooper, A., Turney, C., Hughen, K. A., Brook, B. W., McDonald, H. G., and Bradshaw, C. J. A. 2015. Abrupt warming events drove Late Pleistocene Holarctic megafaunal turnover. *Science*, 349(6248): 602–606.

Drummond, A. J. and Rambaut, A. 2007. Beast: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, 7(1): 214.

Froese, D., Stiller, M., Heintzman, P. D., Reyes, A. V., Zazula, G. D., Soares, A. E., Meyer, M., Hall, E.,

Jensen, B. J., Arnold, L. J., *et al.* 2017. Fossil and genomic evidence constrains the timing of bison arrival in North America. *Proceedings of the National Academy of Sciences*, 114(13): 3457–3462.

Gautier, M., Moazami-Goudarzi, K., Levéziel, H., Parinello, H., Grohs, C., Rialle, S., Kowalczyk, R., and Flori, L. 2016. Deciphering the wisent demographic and adaptive histories from individual whole-genome sequences. *Molecular biology and evolution*, 33(11): 2801–2814.

Gottfried, R. S. 1983. *The black death : natural and human disaster in medieval Europe*. New York Free Press Collier Macmillan.

Grange, T., Brugal, J.-P., Flori, L., Gautier, M., Uzunidis, A., and Geigl, E.-M. 2018. The evolution and population diversity of bison in pleistocene and holocene eurasia: Sex matters. *Diversity*, 10(3): 65.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, 59(3): 307–321.

Günther, T., Malmström, H., Svensson, E. M., Omrak, A., Sánchez-Quinto, F., Kılınç, G. M., Krzewińska, M., Eriksson, G., Fraser, M., Edlund, H., Munters, A. R., Coutinho, A., Simões, L. G., Vicente, M., Sjölander, A., Sellevold, B. J., Jørgensen, R., Claes, P., Shriver, M. D., Valdiosera, C., Netea, M. G., Apel, J., Lidén, K., Skar, B., Storå, J., Götherström, A., and Jakobsson, M. 2018. Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLOS Biology*, 16(1): e2003703.

Ho, S. Y. and Shapiro, B. 2011. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular Ecology Resources*, 11(3): 423–434.

Hofman-Kamińska, E., Bocherens, H., Borowik, T., Drucker, D. G., and Kowalczyk, R. 2018. Stable isotope signatures of large herbivore foraging habitats across Europe. *PLOS ONE*, 13(1): e0190723.

Hofreiter, M. and Shapiro, B. 2012. *Ancient DNA: Methods and Protocols*. Methods in Molecular Biology. Humana Press Incorporated.

Hoogaker, B., Smith, R. S., Singarayer, J. S., Marchant, R., Prentice, I., Allen, J., Anderson, S., Bhagwat, S., Behling, H., Borisova, O., *et al.* 2016. Terrestrial biosphere changes over the last 120 kyr. *Climate of the Past*, 12(1): 51–73.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L., and Orlando, L. 2013. mapdamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, 29(13): 1682–1684.

Julien, M.-A., Bocherens, H., Burke, A., Drucker, D. G., Patou-Mathis, M., Krotova, O., and Péan, S. 2012. Were European steppe bison migratory? ^{18}O, ^{13}C and Sr intra-tooth isotopic variations applied to a palaeoethological reconstruction. *Quaternary International*, 271(Supplement C): 106–119.

Kaplan, J. O., Bigelow, N. H., Prentice, I. C., Harrison, S. P., Bartlein, P. J., Christensen, T. R., Cramer, W., Matveyeva, N. V., McGuire, A. D., Murray, D. F., Razzhivin, V. Y., Smith, B., Walker, D. A., Anderson, P. M., Andreev, A. A., Brubaker, L. B., Edwards, M. E., and Lozhkin, A. V. 2003. Climate change and Arctic ecosystems: 2. Modeling, paleodata-model comparisons, and future projections. *Journal of Geophysical Research: Atmospheres*, 108(D19).

Kaplan, J. O., Pfeiffer, M., Kolen, J. C. A., and Davis, B. A. S. 2016. Large Scale Anthropogenic Reduction of Forest Cover in Last Glacial Maximum Europe. *PLOS ONE*, 11(11): e0166726.

Katoh, K., Misawa, K., Kuma, K.-i., and Miyata, T. 2002. Mafft: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14): 3059–3066.

Kerley, G. I. H., Kowalczyk, R., and Cromsigt, J. P. G. M. 2012. Conservation implications of the refugee species concept and the European bison: king of the forest or

refugee in a marginal habitat? *Ecography*, 35(6): 519–529.

Kirillova, I. V., Zanina, O. G., Chernova, O. F., Lapteva, E. G., Trofimova, S. S., Lebedev, V. S., Tiunov, A. V., Soares, A. E., Shidlovskiy, F. K., and Shapiro, B. 2015. An ancient bison from the mouth of the Rauchua River (Chukotka, Russia). *Quaternary Research*, 84(2): 232–245.

Kowalczyk, R., Taberlet, P., Coissac, E., Valentini, A., Miquel, C., Kamiński, T., and Wójcik, J. M. 2011. Influence of management practices on large herbivore diet—Case of European bison in Białowieża Primeval Forest (Poland). *Forest Ecology and Management*, 261(4): 821–828.

Kowalczyk, R., Krasińska, M., Kamiński, T., Górny, M., Struś, P., Hofman-Kamińska, E., and Krasiński, Z. A. 2013. Movements of european bison (bison bonasus) beyond the białowieża forest (ne poland): range expansion or partial migrations? *Acta theriologica*, 58(4): 391–401.

Krasińska, M. and Krasiński, Z. 2013. *European Bison: The Nature Monograph*. Springer Berlin Heidelberg.

Leigh, J. W. and Bryant, D. 2015. popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6(9): 1110–1116.

Leuschner, C. and Ellenberg, H. 2017. *Ecology of Central European Forests: Vegetation Ecology of Central Europe*. Springer.

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.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16): 2078–2079.

Lindgreen, S. 2012. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Research Notes*, 5(1): 337.

Marquer, L., Gaillard, M.-J., Sugita, S., Trondman, A.-K., Mazier, F., Nielsen, A. B., Fyfe, R. M., Odgaard, B. V., Alenius, T., Birks, H. J. B., *et al.* 2014. Holocene changes in vegetation composition in northern Europe: why quantitative pollen-based vegetation reconstructions matter. *Quaternary Science Reviews*, 90: 199–216.

Marsolier-Kergoat, M.-C., Palacio, P., Berthonaud, V., Maksud, F., Stafford, T., Bégouën, R., and Elalouf, J.-M. 2015. Hunting the Extinct Steppe Bison (Bison priscus) Mitochondrial Genome in the Trois-Frères Paleolithic Painted Cave. *PLOS ONE*, 10(6).

Massilani, D., Guimaraes, S., Brugal, J.-P., Bennett, E. A., Tokarska, M., Arbogast, R.-M., Baryshnikov, G., Boeskorov, G., Castel, J.-C., Davydov, S., Madelaine, S., Putelat, O., Spasskaya, N. N., Uerpmann, H.-P., Grange, T., and Geigl, E.-M. 2016. Past climate changes, population dynamics and the origin of Bison in Europe. *BMC Biology*, 14(1): 93.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., *et al.* 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20: 1297–1303.

Mendoza, M. and Palmqvist, P. 2008. Hypsodonty in ungulates: an adaptation for grass consumption or for foraging in open habitat? *Journal of Zoology*, 274(2): 134–142.

Minin, V. N., Bloomquist, E. W., and Suchard, M. A. 2008. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution*, 25(7): 1459–1471.

Olea, R. A. and Christakos, G. 2005. Duration of Urban Mortality for the 14th-Century Black Death Epidemic. *Human Biology*, 77(3): 291–303.

Onar, V., Soubrier, J., Toker, N. Y., van Loenen, A., Llamas, B., Siddiq, A. B., Pasicka, E., and Tokarska, M. 2017. Did the historical range of the European

bison (Bison bonasus L.) extend further south?—a new finding from the Yenikapı Metro and Marmaray excavation, Turkey. *Mammal Research*, 62(1): 103–109.

Palacio, P., Berthonaud, V., Guérin, C., Lambourdière, J., Maksud, F., Philippe, M., Plaire, D., Stafford, T., Marsolier-Kergoat, M.-C., and Elalouf, J.-M. 2017. Genome data on the extinct Bison schoetensacki establish it as a sister species of the extant European bison (Bison bonasus). *BMC Evolutionary Biology*, 17(1): 48.

Parker, A. G., Goudie, A. S., Anderson, D. E., Robinson, M. A., and Bonsall, C. 2002. A review of the mid-Holocene elm decline in the British Isles. *Progress in Physical Geography*, 26(1): 1–45.

Patton, H., Hubbard, A., Andreassen, K., Auriac, A., Whitehouse, P. L., Stroeven, A. P., Shackleton, C., Winsborrow, M., Heyman, J., and Hall, A. M. 2017. Deglaciation of the Eurasian ice sheet complex. *Quaternary Science Reviews*, 169(Supplement C): 148–172.

Perry, I. and Moore, P. D. 1987. Dutch elm disease as an analogue of Neolithic elm decline. *Nature*, 326(6108): 72.

Pucek, Z. 2004. *European Bison (Bison Bonasus): Current State of the Species and Strategy for Its Conservation*. Council of Europe Publishing.

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

Rambaut, A. and Drummond, A. J. 2009. Tracer version v1.5.0. *http://tree.bio.ed.ac.uk/software/tracer/*.

Ramsey, C. B. 2009. Bayesian analysis of radiocarbon dates. *Radiocarbon*, 51(1): 337–360.

Reimer, P. J., Bard, E., Bayliss, A., Beck, J. W., Blackwell, P. G., Ramsey, C. B., Buck, C. E., Cheng, H., Edwards, R. L., Friedrich, M., Grootes, P. M., Guilderson, T. P., Haflidason, H., Hajdas, I., Hatté, C., Heaton, T. J., Hoffmann, D. L., Hogg, A. G., Hughen, K. A., Kaiser, K. F., Kromer, B., Manning, S. W., Niu, M., Reimer, R. W., Richards, D. A., Scott, E. M., Southon, J. R., Staff, R. A., Turney, C. S. M., and Plicht, J. v. d. 2013. IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. *Radiocarbon*, 55(4): 1869–1887.

Schubert, M., Ermini, L., Sarkissian, C. D., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M. D., Fernández, R., Kircher, M., McCue, M., Willerslev, E., and Orlando, L. 2014. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature Protocols*, 9(5): 1056–1082.

Soubrier, J., Gower, G., Chen, K., Richards, S. M., Llamas, B., Mitchell, K. J., Ho, S. Y. W., Kosintsev, P., Lee, M. S. Y., Baryshnikov, G., Bollongino, R., Bover, P., Burger, J., Chivall, D., Crégut-Bonnoure, E., Decker, J. E., Doronichev, V. B., Douka, K., Fordham, D. A., Fontana, F., Fritz, C., Glimmerveen, J., Golovanova, L. V., Groves, C., Guerreschi, A., Haak, W., Higham, T., Hofman-Kamińska, E., Immel, A., Julien, M.-A., Krause, J., Krotova, O., Langbein, F., Larson, G., Rohrlach, A., Scheu, A., Schnabel, R. D., Taylor, J. F., Tokarska, M., Tosello, G., van der Plicht, J., van Loenen, A., Vigne, J.-D., Wooley, O., Orlando, L., Kowalczyk, R., Shapiro, B., and Cooper, A. 2016. Early cave art and ancient DNA record the origin of European bison. *Nature Communications*, 7: 13158.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9): 1312–1313.

Suchard, M. A. and Rambaut, A. 2009. Many-core algorithms for statistical phylogenetics. *Bioinformatics*, 25(11): 1370–1376.

Theuerkauf, M., Bos, J. A. A., Jahns, S., Janke, W., Kuparinen, A., Stebich, M., and Joosten, H. 2014. Corylus expansion and persistent openness in the early Holocene vegetation of northern central Europe. *Quaternary Science Reviews*, 90(Supplement C): 183–198.

Tokarska, M., Marshall, T., Kowalczyk, R., Wojcik, J., Pertoldi, C., Kristensen, T., Loeschcke, V., Gregersen, V., and Bendixen, C. 2009. Effectiveness of microsatellite and snp markers for parentage and identity analysis in species with low genetic diversity: the case of european bison. *Heredity*, 103(4): 326.

Tokarska, M., Pertoldi, C., Kowalczyk, R., and Perzanowski, K. 2011. Genetic status of the european bison bison bonasus after extinction in the wild and subsequent recovery. *Mammal Review*, 41(2): 151–162.

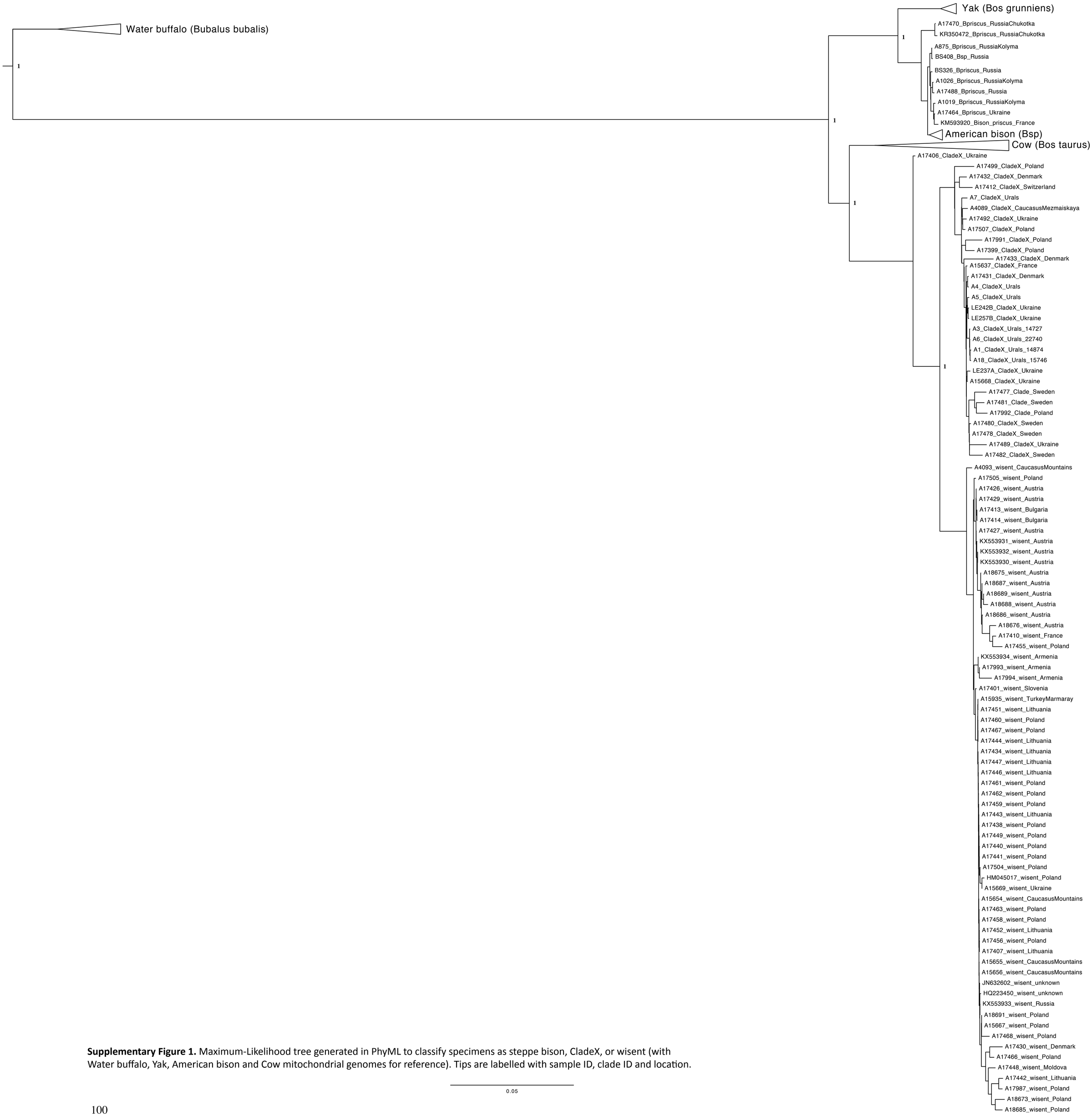
van Asperen, E. N. and Kahlke, R.-D. 2017. Dietary traits of the late Early Pleistocene Bison menneri (Bovidae, Mammalia) from its type site Untermassfeld (Central Germany) and the problem of Pleistocene 'wood bison'. *Quaternary Science Reviews*, 177(Supplement C): 299–313.

Velivetskaya, T. A., Smirnov, N. G., Kiyashko, S. I., Ignatiev, A. V., and Ulitko, A. I. 2016. Resolution-enhanced stable isotope profiles within the complete tooth rows of Late Pleistocene bisons (Middle Urals, Russia) as a record of their individual development and environmental changes. *Quaternary International*, 400(Supplement C): 212–226.

Węcek, K., Hartmann, S., Paijmans, J. L., Taron, U., Xenikoudakis, G., Cahill, J. A., Heintzman, P. D., Shapiro, B., Baryshnikov, G., Bunevich, A. N., *et al.* 2016. Complex admixture preceded and followed the extinction of wisent in the wild. *Molecular Biology and Evolution*, 34(3): 598–612.

Wójcik, J., Kawałko, A., Tokarska, M., Jaarola, M., Vallenback, P., and Pertoldi, C. 2009. Post-bottleneck mtdna diversity in a free-living population of european bison: implications for conservation. *Journal of Zoology*, 277(1): 81–87.

Supplementary Information



Supplementary Figure 1. Maximum-Likelihood tree generated in PhyML to classify specimens as steppe bison, CladeX, or wisent (with Water buffalo, Yak, American bison and Cow mitochondrial genomes for reference). Tips are labelled with sample ID, clade ID and location.

Supplementary Figure 2. RaXML tree constrained using the BEAST MCC tree, adding two low coverage specimens (highlighted in red: A17406 and A17482, both falling within the diversity of CladeX). Tips are labelled with sample ID, clade ID, locations and age (in mean calibrated years BP)

Supplementary Table 1. MapDamage values

Sample ID	Clade ID	Source	3' G>A frequency	5' C>T frequency
A001	CladeX	Soubrier et al., 2016	0.0251	0.0223
A003	CladeX	Soubrier et al., 2016	0.0825	0.0886
A004	CladeX	Soubrier et al., 2016	0.0316	0.0380
A005	CladeX	Soubrier et al., 2016	0.1149	0.0670
A006	CladeX	Soubrier et al., 2016	0.1680	0.1361
A007	CladeX	Soubrier et al., 2016	0.0908	0.1515
A018	CladeX	Soubrier et al., 2016	0.0732	0.0744
A1019	steppe bison	This study	0.1014	0.0738
A1026	steppe bison	This study	0.0400	0.0535
A15637	CladeX	Soubrier et al., 2016	0.3303	0.3890
A15654	wisent	Soubrier et al., 2016	0.0204	0.0493
A15655	wisent	This study	0.0207	0.0233
A15656	wisent	This study	0.0225	0.0291
A15667	wisent	This study	0.0591	0.0466
A15668	CladeX	Soubrier et al., 2016	0.0813	0.0877
A15669	wisent	This study	0.0500	0.0663
A15935	wisent	Onar et al., 2017	0.1117	0.1726
A17399	CladeX	This study	0.0720	0.0833
A17401	wisent	This study	0.1025	0.1025
A17406	CladeX	This study	0.0674	0.0617
A17407	wisent	This study	0.0371	0.0483
A17410	wisent	This study	0.0918	0.0497
A17412	CladeX	This study	0.0236	0.0331
A17413	wisent	This study	0.0842	0.0557
A17414	wisent	This study	0.0674	0.0522
A17426	wisent	This study	0.0618	0.0375
A17427	wisent	This study	0.0486	0.0584
A17429	wisent	This study	0.0574	0.0518
A17430	wisent	This study	0.0438	0.0439
A17431	CladeX	This study	0.1050	0.1133
A17432	CladeX	This study	0.0925	0.0969
A17433	CladeX	This study	0.0344	0.0308
A17434	wisent	This study	0.0263	0.0280
A17438	wisent	This study	0.0609	0.0746
A17440	wisent	This study	0.1118	0.0559
A17441	wisent	This study	0.0644	0.0943
A17442	wisent	This study	0.0253	0.0337
A17443	wisent	This study	0.0365	0.0473
A17444	wisent	This study	0.0293	0.0437
A17446	wisent	This study	0.0629	0.0529
A17447	wisent	This study	0.0308	0.0251
A17448	wisent	This study	0.0345	0.0359
A17449	wisent	This study	0.0210	0.0332
A17451	wisent	This study	0.0294	0.0251
A17452	wisent	This study	0.0250	0.0292
A17455	wisent	This study	0.0735	0.0842
A17456	wisent	This study	0.0323	0.0216
A17458	wisent	This study	0.0334	0.0285
A17459	wisent	This study	0.0306	0.0247
A17460	wisent	This study	0.0463	0.0345
A17461	wisent	This study	0.0366	0.0256
A17462	wisent	This study	0.0750	0.0596
A17463	wisent	This study	0.0322	0.0339

Sample ID	Clade ID	Source	3' G>A frequency	5' C>T frequency
A17464	steppe bison	This study	0.0306	0.0339
A17466	wisent	This study	0.0414	0.0456
A17467	wisent	This study	0.0427	0.0371
A17468	wisent	This study	0.0194	0.0271
A17470	steppe bison	This study	0.0594	0.0261
A17477	CladeX	This study	0.0441	0.0533
A17478	CladeX	This study	0.1032	0.0916
A17480	CladeX	This study	0.0459	0.0103
A17481	CladeX	This study	0.0938	0.0719
A17482	CladeX	This study	0.0620	0.0259
A17488	steppe bison	This study	0.1082	0.1078
A17489	CladeX	This study	0.0753	0.0493
A17492	CladeX	This study	0.1355	0.1431
A17499	CladeX	This study	0.0524	0.0685
A17504	wisent	This study	0.0451	0.0307
A17505	wisent	This study	0.0726	0.0580
A17507	CladeX	This study	0.1331	0.1388
A17987	wisent	This study	0.1089	0.0659
A17991	CladeX	This study	0.0680	0.0707
A17992	CladeX	This study	0.2230	0.1612
A17993	wisent	This study	0.0715	0.0693
A17994	wisent	This study	0.0265	0.0342
A18673	wisent	This study	0.0284	0.0571
A18675	wisent	This study	0.0266	0.0513
A18676	wisent	This study	0.0393	0.0781
A18685	wisent	This study	0.0525	0.0278
A18686	wisent	This study	0.0596	0.0619
A18687	wisent	This study	0.0931	0.1995
A18688	wisent	This study	0.1965	0.1998
A18689	wisent	This study	0.1412	0.1631
A18691	wisent	This study	0.1301	0.1900
A4089	CladeX	Soubrier et al., 2016	0.0431	0.0429
A4093	wisent	Soubrier et al., 2016	0.1326	0.1312
A875	steppe bison	Soubrier et al., 2016	0.0243	0.0084
BS326	steppe bison	This study	0.1257	0.1238
BS408	steppe bison	This study	0.0889	0.1082
LE237A	CladeX	Soubrier et al., 2016	0.1016	0.1006
LE242B	CladeX	Soubrier et al., 2016	0.2702	0.3006
LE247B	CladeX	Soubrier et al., 2016	0.2189	0.2329

Supplementary Table 2. PALEOMIX summary values

Sample ID	RetainedReads	HitsRaw	HitsUniq	Endogenous	Clonal	Raw_Mit	Uniq_Mit	Coverage_Mit	Depth_Mit
A001	162452645	34668211	12096026	0.213	0.651	1932100	121935	1	566.86
A003	2692018	936137	429501	0.348	0.541	377635	66216	1	341.28
A004	5417876	3272115	586606	0.604	0.821	2407791	227167	1	1183.52
A005	2257928	1200675	682145	0.532	0.432	267807	41484	0.99	206.42
A006	14994558	705493	562246	0.047	0.203	123164	45719	0.99	212.07
A007	4014652	1281518	538910	0.319	0.579	182412	46936	0.99	204.4
A018	4965020	1432823	165655	0.289	0.884	1073040	35189	1	176.11
A1019	2238590	304709	299596	0.136	0.017	2075	1984	0.99	6.93
A1026	2502417	1070882	1048721	0.428	0.021	6098	2905	1	10.31
A15637	493371	8555	6073	0.017	0.290	4466	2665	0.92	8.23
A15654	31209944	10862828	9642610	0.348	0.112	36474	35035	1	196.3
A15655	5853676	3174092	2260332	0.542	0.288	1801414	1086686	1	5260.62
A15656	6884164	3193783	2040603	0.464	0.361	1785267	943908	1	4502.3
A15667	471709	5642	5500	0.012	0.025	854	801	0.81	2.87
A15668	245387	5182	4157	0.021	0.198	3641	2876	0.94	10.5
A15669	158197	3453	3380	0.022	0.021	469	427	0.77	2.05
A15935	12712406	4042548	3643802	0.318	0.099	70073	25747	1	124.61
A17399	782666	4044	2963	0.005	0.267	1328	506	0.81	2.05
A17401	7886605	58845	56363	0.007	0.042	872	575	0.85	2.62
A17406	5927661	12222	10691	0.002	0.125	299	127	0.38	0.5
A17407	8120492	120929	112597	0.015	0.069	7307	4591	1	22.05
A17410	11727138	415696	331253	0.035	0.203	27029	10476	1	36.06
A17412	6317299	16351	12217	0.003	0.253	1246	407	0.51	1.18
A17413	2458475	63137	48103	0.026	0.238	8093	1491	0.96	7.71
A17414	2180160	1318566	1009428	0.605	0.234	349367	164619	1	845.7
A17426	819134	5049	4238	0.006	0.161	2359	1712	0.99	7.56
A17427	659565	56766	53998	0.086	0.049	9256	7438	1	36.46
A17429	1190170	316432	261280	0.266	0.174	73347	31154	1	149.5
A17430	10654201	20247	13235	0.002	0.346	4249	1270	0.88	4.31
A17431	7109023	201572	132838	0.028	0.341	59358	18567	0.99	61.07
A17432	461687	12968	12618	0.028	0.027	554	490	0.69	1.5
A17433	15111515	25390	17279	0.002	0.319	1548	426	0.49	1.29
A17434	1276770	203931	122256	0.160	0.401	120062	54062	1	251.98
A17438	950088	9270	8311	0.010	0.103	3813	3057	1	14.05
A17440	3318483	197000	119569	0.059	0.393	55421	8532	1	34.6
A17441	2845968	71948	61874	0.025	0.140	16535	11940	0.99	64.96
A17442	763057	3806	3724	0.005	0.022	264	233	0.56	0.84
A17443	1085000	596145	554066	0.549	0.071	90828	62341	1	269.91

Sample ID	RetainedReads	HitsRaw	HitsUniq	Endogenous	Clonal	Raw_Mit	Uniq_Mit	Coverage_Mit	Depth_Mit
A17444	1097963	740217	723979	0.674	0.022	69261	62788	1	267.38
A17446	943487	47676	45228	0.051	0.051	17084	15396	1	64.14
A17447	1027139	93939	81555	0.091	0.132	24348	15015	1	63.75
A17448	47555835	209214	162370	0.004	0.224	17100	5271	1	22.49
A17449	8708114	98005	59218	0.011	0.396	35980	10783	1	40.74
A17451	911409	18135	16603	0.020	0.084	4599	3483	1	15.79
A17452	1160224	24212	17660	0.021	0.271	12198	6927	1	32.92
A17455	6084504	53244	40248	0.009	0.244	8754	3798	0.99	13.23
A17456	1218246	43448	29968	0.036	0.310	19621	9098	1	43.26
A17458	1011428	21363	14540	0.021	0.319	11134	5548	1	26.09
A17459	1140958	148805	99380	0.130	0.332	69517	29701	1	132.21
A17460	1582190	402428	245095	0.254	0.391	164753	43545	1	178.65
A17461	959685	25701	18231	0.027	0.291	11399	5404	1	24.52
A17462	1036919	299897	279083	0.289	0.069	86944	72290	1	342.04
A17463	893251	9608	8570	0.011	0.108	4806	3990	1	18.11
A17464	1799171	11302	8303	0.006	0.265	1799	545	0.79	1.84
A17466	1655742	21555	12007	0.013	0.443	11837	4202	1	16.81
A17467	1333244	61702	34231	0.046	0.445	39144	17232	1	71.28
A17468	4504596	10510	9102	0.002	0.134	1227	920	0.83	3.51
A17470	2945191	10202	5803	0.003	0.431	3117	238	0.61	1.06
A17477	8986285	66809	50394	0.007	0.246	8985	3550	0.96	11.53
A17478	1011520	30796	20470	0.030	0.335	19929	11502	1	51.23
A17480	768681	6893	5906	0.009	0.143	1878	1123	0.97	5.07
A17481	16527231	44234	35676	0.003	0.193	7609	4405	0.96	13.63
A17482	864775	4617	4322	0.005	0.064	589	412	0.74	1.51
A17488	1157285	57383	50832	0.050	0.114	1058	686	0.79	2.9
A17489	823825	4154	2995	0.005	0.279	1102	168	0.52	0.72
A17492	1017472	22222	17409	0.022	0.217	10132	6310	0.99	26.88
A17499	927352	11060	6542	0.012	0.408	3794	456	0.74	1.53
A17504	5454508	5462	5029	0.001	0.079	1254	989	0.93	4.78
A17505	5691377	235885	218141	0.041	0.075	9844	4034	1	18.03
A17507	4142696	217360	199017	0.052	0.084	12252	3323	0.99	14.24
A17987	6929982	37480	22262	0.005	0.406	7140	2392	0.95	8.38
A17991	2541840	5973	4061	0.002	0.320	1262	373	0.55	1.16
A17992	4074579	35151	12895	0.009	0.633	18221	2600	0.99	10.65
A17993	4476673	1469843	697733	0.328	0.525	377636	21657	1	99.47
A17994	4377430	218252	85925	0.050	0.606	72306	6266	1	23.65
A18673	3862925	17436	12566	0.005	0.279	4224	2228	0.95	7.65

Sample ID	RetainedReads	HitsRaw	HitsUniq	Endogenous	Clonal	Raw_Mit	Uniq_Mit	Coverage_Mit	Depth_Mit
A18675	18369464	9478532	6596625	0.516	0.304	1666426	136342	1	510.55
A18676	2653262	1808207	1784129	0.682	0.013	2049	1982	0.99	7.67
A18685	5168477	133177	25380	0.026	0.809	73774	1966	0.99	6.91
A18686	8274809	4915868	4004601	0.594	0.185	753909	341185	1	1388.68
A18687	17512607	9961905	8564826	0.569	0.140	753101	226524	1	966.87
A18688	18164953	8369158	6967373	0.461	0.167	644354	217954	1	874.99
A18689	19338238	10277151	7692214	0.531	0.252	1795230	380243	1	1466.94
A18691	16686110	8711407	7587787	0.522	0.129	584355	173166	1	701.42
A4089	8618722	6447432	104557	0.748	0.984	5436517	44869	1	238.68
A4093	9612598	3146908	1085197	0.327	0.655	111495	19959	1	85.9
A875	988026064	555627150	480724760	0.562	0.135	245137	192223	1	652.77
BS326	671162	162667	159642	0.242	0.019	2002	1953	1	6.82
BS408	509414	95993	94567	0.188	0.015	228	222	0.54	0.83
LE237A	507023	11861	9389	0.023	0.208	4272	2678	0.98	9.67
LE242B	6912671	134742	112132	0.019	0.168	48788	35415	0.99	119.49
LE247B	4156307	258923	57104	0.062	0.779	184219	28771	0.99	93.63

Supplementary Table 3. Sample Information

Sample ID	Clade ID	Source	Museum	Museum ID	Location	$\delta^{13}C$	$\delta^{15}N$	Raw date (C^{14}/ archaeological)
A001	CladeX	Soubrier et al., 2016	IPAE	888/117	Rasik 1, Urals, Russia	-19.887	3.08	12565±55
A003	CladeX	Soubrier et al., 2016	IPAE	1871/01	Voronovka, Urals, Russia	-18.993	8.608	12505±55
A004	CladeX	Soubrier et al., 2016	IPAE	888/1705	Rasik 1, Urals, Russia	-19.055	3.976	19010±80
A005	CladeX	Soubrier et al., 2016	IPAE	929/1	Ladeinyi Kamen, Urals, Russia	-19.171	4.491	15310±70
A006	CladeX	Soubrier et al., 2016	IPAE	994/714	Sur'ya 5, Urals, Russia	-19.244	4.532	18880±90
A007	CladeX	Soubrier et al., 2016	IPAE	884/19	Sur'ya 3, Urals, Russia	-19.474	3.816	58300±2900
A018	CladeX	Soubrier et al., 2016	IPAE	994/315	Sur'ya 5, Urals, Russia	-19.627	3.025	13120±60
A1019	steppe bison	This study	Ian Barnes	CRS-SY-11	Stanchikovsky Yar, Kolyma, Siberia, Russia	-	-	-
A1026	steppe bison	This study	Ian Barnes	CRS-IC-18	Bolshoi-Homoos-Uriakh River, Kolyma, Siberia, Russia	-20.2	8	>56800
A15637	CladeX	Soubrier et al., 2016	Musee de Prehistoire d'Orgnac-l'Aven	-	Aven de l'Arquet, Gars, France	-19.75	-	>48300
A15654	wisent	Soubrier et al., 2016	Zoological Museum of the Zoological Institute RAS in St. Petersburg	8853	Kuban, Caucasus, Russia	-	-	1911AD
A15655	wisent	This study	Zoological Museum of the Zoological Institute RAS in St. Petersburg	7985	Kubanskaja oblast, Kuban, Northern Caucasus, Russia	-	-	1906AD
A15656	wisent	This study	Zoological Museum of the Zoological Institute RAS in St. Petersburg	7985	Kubanskaja oblast, Kuban, Northern Caucasus, Russia	-	-	1906AD
A15667	wisent	This study	Private Museum of Etnography and Archeology (W. Litwińczuk)	NA	Narew river, Suraz, Poland	-	-	3386±29
A15668	CladeX	Soubrier et al., 2016	Museum in Vinnytsia	Gp-673 (3)	Zmerynka, Ukraine	-	-	13573±36
A15669	wisent	This study	State Museum of Natural History Lviv (old Museum Dzieduszyckich) dr Natalia Cheremnyh, Klimishin DO, Boretskiy V.	D-1492	Sokalskiy, Ukraine	-	-	1634±26
A15935	wisent	Onar et al., 2017	Faculty of Veterinary Medicine Istambul University	-	Marmaray, Istanbul, Turkey	-	-	1283±26

Sample ID	Clade ID	Source	Museum	Museum ID	Location	δ^{13}C	δ^{15}N	Raw date (C^{14}/ archaeological)
A17399	CladeX	This study	Museum of the Turek City them. Joseph Mehoffer	MRzTT/P/21 old no 14	Poland	-	-	29046±280
A17401	wisent	This study	Slovenian Museum of Natural History	PMSL OV. 90	Volcija Jama, Slovenia	-	-	17792±78
A17406	CladeX	This study	State Museum of Natural History Lviv (old Museum Dzieduszyckich) dr Natalia Cheremnyh	D-919	Ukraine	-19.4	8.0	42623±816
A17407	wisent	This study	Lietuvos Nacionalinis Muziejus, dr. Giedrė Piličiauskienė	VLC 46	Vilinius Castle, Lithuania	-23.3	6.0	600BP
A17410	wisent	This study	L. Chaix and C. Cupillard	-	Gouffre de la Cabosse, Gex, Ain, France	-22.8	5.5	4325±35
A17412	CladeX	This study	Bern Naturhistorisches Museum M. Nussbaumer	-	Lajoux , Switzerland	-20.8	3.8	27394±161
A17413	wisent	This study	National Museum of Natural History, Sofia. Dr Nikolai Spassov	-	Ponor, Bulgaria	-21.9	6.3	3535±31
A17414	wisent	This study	National Museum of Natural History, Sofia. Dr Nikolai Spassov	-	Ponor, Bulgaria	-22.8	4.8	3561±30
A17426	wisent	This study	Naturhistorisches Museum Wien M. Pacher	-	Styria, Austria	-24.3	3.2	2038±29
A17427	wisent	This study	Naturhistorisches Museum Wien M. Pacher	H77-49-1	Zwergweiden, Austria	-23.3	4.7	2073±29
A17429	wisent	This study	Naturhistorisches Museum Wien M. Pacher	H65-1-1	Stainsenkogel, Austria	-22.1	2.7	2231±26
A17430	wisent	This study	Natural History Museum, Copenhagen	ZMK 68/1945 P183/2015KMG	Sjaelland, Denmark	-22.6	6.1	1480±50
A17431	CladeX	This study	Natural History Museum, Copenhagen	ZMK 12/1921 P187/2015KMG	Akkerup, Fyn, Denmark	-19.6	2.9	9540±85
A17432	CladeX	This study	Natural History Museum, Copenhagen	ZMK 26/1944 P186/2015KMG	Harndrup, Fyn, Denmark	-19.6	3.5	12000
A17433	CladeX	This study	Natural History Museum, Copenhagen	ZMK 68/1944 P184/2015KMG	Tranum, Denmark	-20.0	3.5	10000±80
A17434	wisent	This study	Lietuvos Nacionalinis Muziejus, dr. Giedrė Piličiauskienė	20	Vilinius Castle, Lithuania	-23.1	5.8	537BP
A17438	wisent	This study	Prof.Daniel Makowiecki	nr. inw.14/11	Lidzbark Warminski, Poland	-23.7	5.6	639±28
A17440	wisent	This study	Prof.Daniel Makowiecki	nr inw.N 11/1	Nienawiszcz, Poland	-23.4	5.1	2128±28
A17441	wisent	This study	Prof.Daniel Makowiecki	nr inw. 63/93	Napole, Poland	-22.8	4.3	894±27

Sample ID	Clade ID	Source	Museum	Museum ID	Location	δ^{13}C	δ^{15}N	Raw date (C^{14}/ archaeological)
A17442	wisent	This study	Lietuvos Nacionalinis Muziejus, dr. Giedrė Piličiauskienė	AL.'11; P1,J2	Alytus, Lithuania	-22.3	4.5	667±26
A17443	wisent	This study	Lietuvos Nacionalinis Muziejus, dr. Giedrė Piličiauskienė	VR'03, VLC2	Vilinius Castle, Lithuania	-23.5	6.0	469±26
A17444	wisent	This study	Lietuvos Nacionalinis Muziejus, dr. Giedrė Piličiauskienė	VLC7	Vilinius Castle, Lithuania	-24.0	4.3	500BP
A17446	wisent	This study	Lietuvos Nacionalinis Muziejus, dr. Giedrė Piličiauskienė	VLC 4	Vilinius Castle, Lithuania	-23.3	4.8	600BP
A17447	wisent	This study	Lietuvos Nacionalinis Muziejus, dr. Giedrė Piličiauskienė	VLC 15	Vilinius Castle, Lithuania	-23.0	3.6	600BP
A17448	wisent	This study	Institute of Zoology of the Academy of Sciences of Moldova, Viorelia Rusu, prof. David Anatolie	1S	Sipoteni, Moldova	-18.9	7.1	12383±42
A17449	wisent	This study	Nature Museum in Drozdowo	B5 KW631	Barzykowo, Poland	-22.2	5.1	1536±27
A17451	wisent	This study	Tadas Ivanauskas Zoological Museum in Kaunas (Vaclovas Gedminas)	M-4644 S5-st218	Telsiai, Lithuania	-22.8	4.8	694±28
A17452	wisent	This study	Tadas Ivanauskas Zoological Museum in Kaunas (Vaclovas Gedminas)	M-4652 S5-st218	Marijampole, Lithuania	-23.1	4.6	614±28
A17455	wisent	This study	Nature Museum in Olsztyn (Marian Szymkiewicz)	-	Nadl. Gorowo Ilaweckie, Poland	-22.4	4.1	3145±30
A17456	wisent	This study	Białowieża National Park (E. Keczyńska-Moroz)	-	Zlobin, Poland	-22.5	4.1	502±28
A17458	wisent	This study	Private Museum in Dobrzyniewo Duże Bogusław Antoniuk	-	Gora, Poland	-22.2	5.3	788±25
A17459	wisent	This study	Museum of North Mazowsze, Łomża	-	Lomza, Poland	-21.8	3.8	543±25
A17460	wisent	This study	Museum of North Mazowsze, Łomża	-	Lomza, Poland	-22.8	4.1	575±25
A17461	wisent	This study	Prof.Daniel Makowiecki	5	Elk Castle, Poland	-24.4	3.8	371±23
A17462	wisent	This study	Prof.Daniel Makowiecki	nr inw. 61 (8)	Elk Castle, Poland	-22.8	5.1	600BP
A17463	wisent	This study	Prof.Daniel Makowiecki	12	Staswiny, Poland	-23.3	4.9	1442±27
A17464	steppe bison	This study	Paleontological Museum I. Franko in Lviv	MP 867 (5)	Ukraine	-20.3	6.3	26448±142

Sample ID	Clade ID	Source	Museum	Museum ID	Location	δ^{13}C	δ^{15}N	Raw date (C^{14}/ archaeological)
A17466	wisent	This study	State Museum of Natural History Lviv (old Museum Dzieduszyckich) Natalia Cheremnyh found prof. Friedberg	D-222	Wisla, Poland	-21.7	5.0	2253±26
A17467	wisent	This study	State Museum of Natural History Lviv (old Museum Dzieduszyckich) Natalia Cheremnyh found prof. Friedberg	D-220	Rzeszow, Poland	-23.4	5.6	2068±26
A17468	wisent	This study	Museum of Gentry of Mazowsze in Ciechanów (ING PAN in Warsaw dr Magdalena Krajcarz)	GRU/PRB/ 94/10	Grudusk, Poland	-23.3	6.0	938±26
A17470	steppe bison	This study	The National Museum of Natural History of NAS Dep. of vertebrate paleozoology V.A.Topachevsky Paleontological Museum in Kiev	-	Chukotka, Siberia, Russia	-19.7	8.0	22441±91
A17477	CladeX	This study	Geological Survey of Sweden (SGU) Uppsala dr. Linda Wickström	SGU:Typnr 7721	Hagebyhoga, Sweden	-20.2	2.2	9370±110
A17478	CladeX	This study	Geological Survey of Sweden (SGU) Uppsala dr. Linda Wickström	SGU:Typnr 6224	Oland, Sweden	-20.6	3.4	9480±130
A17480	CladeX	This study	Stjerngranat Museum at Stjärneborg (Ann-Marie Hultman)	SMS 1537	Tranas, Sweden	-19.2	2.5	9460±130
A17481	CladeX	This study	Zoological Museum in Lund, Collection of Zoology and Entomology, Lund University	Lzz/3276	Trelleborg, Sweden	-22.1	3.7	9100±110
A17482	CladeX	This study	Zoological Museum in Lund, Collection of Zoology and Entomology, Lund University	Lzz/3305	Skane, Sweden	-22.6	3.9	8700±130
A17488	steppe bison	This study	The National Museum of Natural History of NAS Dep. of vertebrate paleozoology V.A.Topachevsky Paleontological Museum in Kiev	2	Chukotka, Siberia, Russia	-20.6	4.5	48967±3233
A17489	CladeX	This study	Institute of Archeology Ukrainian Academy of Sciences, Kiev (Aleksandra Aleksandrovna Krotova)	A-89	Amvrosievka, Donetsk Oblast, Ukraine	-19.2	7.1	18361±74
A17492	CladeX	This study	Nature Museum Lviv (old Museum Dzieduszyckich) Natalia Czeremnyh	D-927	Ukraine	-18.4	9.2	45147±999

Sample ID	Clade ID	Source	Museum	Museum ID	Location	$\delta^{13}C$	$\delta^{15}N$	Raw date (C^{14}/ archaeological)
A17499	CladeX	This study	County Museum in Konin Izabela Lorek	MOK/P/583	Goslawice, Poland	-20.1	4.6	43993±971
A17504	wisent	This study	Museum of the Earth in Warsaw PAS, dr Barbara Studencka	VIII Vm 755	Zajeczniki, Poland	-	-	1620±24
A17505	wisent	This study	Museum of the Earth in Warsaw PAS, dr Barbara Studencka	VIII Vm 762	Pyskowice, Poland	-23.5	5	3677±27
A17507	CladeX	This study	Museum of the Earth in Warsaw PAS, dr Barbara Studencka	VIII Vm 814	Bolimow, Poland	-20.4	5.5	>47618
A17987	wisent	This study	Institute of Paleobiology PAS dr. A. T. Halamski	MZR L.m. 29A (1)	Wislok river, Poland	-22.3	7.3	5432±37
A17991	CladeX	This study	Adam Stępień	-	Vistula river, Warsaw, Poland	-	-	47801±2790
A17992	CladeX	This study	Adam Stępień	-	Vistula river, Warsaw, Poland	-	-	49790±3581
A17993	wisent	This study	Institute of Zoology Marine Arakelyan	1	Sevan lake, Armenia	-19.5	4.9	4919±37
A17994	wisent	This study	Institute of Zoology Marine Arakelyan	2	Sevan lake, Armenia	-19.1	5.9	4537±37
A18673	wisent	This study	Muzeum Regionalne w Jaśle (Alfred Sepioł)	123 P-P	Rzeszow, Poland	-23.1	6.2	2057±36
A18675	wisent	This study	Department of Palaeontology, University of Vienna, Prof. Karl Rauscher (Institut für Paläontologie an der Universität Wien)	-	Aschbach, Austria	-23.3	2.6	1628±36
A18676	wisent	This study	Department of Palaeontology, University of Vienna, Prof. Karl Rauscher (Institut für Paläontologie an der Universität Wien)	-	Aschbach, Austria	-24.0	3.1	1629±36
A18685	wisent	This study	Natural History Museum Vienna, Department of Geology and Palaeontology Collection Dr. Ursula B. Göhlich	372/1960	Krakow, Poland	-23.0	5.6	2114±36
A18686	wisent	This study	Natural History Museum Vienna, Zoology (vertebrates) Collection dr. habil. Frank E. Zachos	H79-59-2	Styria, Austria	-23.0	2.7	1530±70
A18687	wisent	This study	Natural History Museum Vienna, Zoology (vertebrates) Collection dr. habil. Frank E. Zachos	H70-30-2	Durrenstein, Lunz am See, Austria	-23.0	4.7	2010±60

Sample ID	Clade ID	Source	Museum	Museum ID	Location	δ^{13}C	δ^{15}N	Raw date (C^{14}/ archaeological)
A18688	wisent	This study	Natural History Museum Vienna, Zoology (vertebrates) Collection dr. habil. Frank E. Zachos	H77-49-1	Styria, Austria	-23.3	4.7	2038±88
A18689	wisent	This study	Natural History Museum Vienna, Zoology (vertebrates) Collection dr. habil. Frank E. Zachos	H77-49-1	Styria, Austria	-22.8	3.9	1980±45
A18691	wisent	This study	MRI PAS	-	Kurpiewskie, Ostroleka Country, Poland	-23.2	6.1	2265±25
A4089	CladeX	Soubrier et al., 2016	Liubov Golovanova, Vladimir Doronichev	M3M N9	Mezmaiskaya cave (Level 2B4, Horizon 1, Quadrant O18, Plan number 203), Russia	-19.12		>59400
A4093	wisent	Soubrier et al., 2016	Liubov Golovanova, Vladimir Doronichev	-	Mezmaiskaya Cave, Level 2B3, Horizon 2, Quadrant N18, Plan number 64, Russia	-18.91	-	>56300
A875	steppe bison	Soubrier et al., 2016	PIN	3658-131	Alyoshkina Zaimka, Kolyma Lowland, Siberia, Russia	-20.8	7.3	>50000
BS326	steppe bison	This study	(ChNRS) Northern Research Station, Cherskii Russia	CRS-IC-29	Kolyma lowlands, Siberia, Russia	-	-	>57300
BS408	steppe bison	This study	(ZIN) Zoological Institute, St. Petersburg, Russia	StP-1	Yana-Indigirka lowland, Siberia, Russia	-	-	>54100
HM045017.1	wisent	Zeyland et al., 2012	-	-	Wroclaw Zoo, Poland	-	-	Modern blood
HQ223450	wisent	Derr et al., 2010	-	-	-	-	-	Modern
JN632602	wisent	Hassanin et al., 2012	-	-	CYTO, B. Dutrillaux	-	-	Modern tissue
KM593920.1	steppe bison	Marsolier-Kergoat et al., 2015	Museum of the Association Louis Begouen (Montesquieu-Avantes)	-	Trois-Frères Paleolithic cave	-	-	15,880±70BP
KR350472.1	steppe bison	Kirillova et al., 2015	National Alliance of Shidlovskiy Ice Age Museum, Moscow, Russia	-	Rauchua River, Chukotka, Russia	-	-	9497±92
KX269109.1	steppe bison	Froese et al., 2017	-	AE006	Chukotka, Siberia, Russia	-	-	20910±120
KX269110.1	steppe bison	Froese et al., 2017	-	AE007	Krasnoyarsk Territory, Taimyr, Siberia, Russia	-	-	-

Sample ID	Clade ID	Source	Museum	Museum ID	Location	δ^{13}C	δ^{15}N	Raw date (C^{14}/ archaeological)
KX269111.1	steppe bison	Froese et al., 2017	-	AE009	Krasnoyarsk Territory, Taimyr, Siberia, Russia	-	-	-
KX269112.1	steppe bison	Froese et al., 2017	-	AE010	Krasnoyarsk Territory, Taimyr, Siberia, Russia	-	-	44100±2000
KX269113.1	steppe bison	Froese et al., 2017	-	AE012	Krasnoyarsk Territory, Taimyr, Siberia, Russia	-	-	31870±570
KX269114.1	steppe bison	Froese et al., 2017	-	AE014	Krasnoyarsk Territory, Taimyr, Siberia, Russia	-	-	32310±470
KX269115.1	steppe bison	Froese et al., 2017	-	AE015	Krasnoyarsk Territory, Taimyr, Siberia, Russia	-	-	43200±1800
KX269116.1	steppe bison	Froese et al., 2017	-	AE017	Krasnoyarsk Territory, Taimyr, Siberia, Russia	-	-	32610±490
KX553930.1	wisent	Węcek et al., 2016	Vienna Museum of Natural History.	H-1981-28-6	Styria, Austria	-	-	1480±70
KX553931.1	wisent	Węcek et al., 2016	Vienna Museum of Natural History.	H-1977-49-1	Styria, Austria	-	-	1980±45
KX553932.1	wisent	Węcek et al., 2016	Vienna Museum of Natural History.	H-1979-48-1	Upper Austria, Austria	-	-	1370±50
KX553933.1	wisent	Węcek et al., 2016	State Darwin Museum, Moscow, Russia		North Ossetia-Alania, Russia	-	-	-
KX553934.1	wisent	Węcek et al., 2016	-	-	Sevan Lake region, Armenia	-	-	4972±29
KX898005.1	CladeX	Massilani et al., 2016	Musee de Prehistoire d'Orgnac-l'Aven	18	L'Aven de l'Arquet, Ardèche, Gard, France	-	-	46700-34300yr old
KX898006.1	CladeX	Massilani et al., 2016	Musee de Prehistoire d'Orgnac-l'Aven	4455	L'Aven de l'Arquet, Ardèche, Gard, France	-	-	46700-34300yr old
KX898007.1	CladeX	Massilani et al., 2016	Musee de Prehistoire d'Orgnac-l'Aven	78531	L'Aven de l'Arquet, Ardèche, Gard, France	-	-	46700-34300yr old
KX898008.1	wisent	Massilani et al., 2016	Natural History Museum Geneva, Department of Archeozoology, Geneva, Switzerland	GRAL01-P47d4-76	Igue du Gral, Sauliac-sur-Célé, Lot, France	-	-	12100-11700yr old
KX898009.1	wisent	Massilani et al., 2016	Natural History Museum Geneva, Department of Archeozoology, Geneva, Switzerland	P47-dec006-125	Igue du Gral, Sauliac-sur-Célé, Lot, France	-	-	12400-11800yr old
KX898010.1	steppe bison	Massilani et al., 2016	Natural History Museum Geneva, Department of Archeozoology, Geneva, Switzerland	Q47-dec013-232	Igue du Gral, Sauliac-sur-Célé, Lot, France	-	-	20200-14700yr old

Sample ID	Clade ID	Source	Museum	Museum ID	Location	δ^{13}C	δ^{15}N	Raw date (C^{14}/ archaeological)
KX898011.1	wisent	Massilani et al., 2016	Zentrum fur Naturwissenschaftliche Archaologie, Tubingen, Germany	-	Kesslerloch Cave, Switzerland	-	-	12180±70BP
KX898012.1	wisent	Massilani et al., 2016	Zoological Museum of the Zoological Institute RAS, St. Petersburg, Russia	2S-133; 1977; layer 2; sq. P-9	Kudaro 3 Cave, Kvaisa City, Southern Ossetia, Southern Caucasus, Georgia	-	-	22500-22100yr old
KX898013.1	wisent	Massilani et al., 2016	Zoological Museum of the Zoological Institute RAS, St. Petersburg, Russia	4S-136; 1984; layer 3ab-2	Kudaro 3 Cave, Kvaisa City, Southern Ossetia, Southern Caucasus, Georgia	-	-	38500-37000yr old
KX898014.1	steppe bison	Massilani et al., 2016	Musee de Prehistoire, Les Eyzies de Tayac-Sireuil, France	LB00-N6a	La Berbie, Castels, Dordogne, France	-	-	39100-36300yr old
KX898015.1	CladeX	Massilani et al., 2016	Zoological Museum of the Zoological Institute RAS, St. Petersburg, Russia	2N-127; 1988; layer 2b; sq.O-21	Mezmaiskaya Cave, Northern Caucasus Mountains, Russia	-	-	46500-44600yr old
KX898016.1	CladeX	Massilani et al., 2016	Zoological Museum of the Zoological Institute RAS, St. Petersburg, Russia	3N-128; 1988; layer 2b; sq. H-22	Mezmaiskaya Cave, Northern Caucasus Mountains, Russia	-	-	47000-44000yr old
KX898017.1	wisent	Massilani et al., 2016	Zoological Museum of the Zoological Institute RAS, St. Petersburg, Russia	5N-130; 1989; layer 2b; horizon 4	Mezmaiskaya Cave, Northern Caucasus Mountains, Russia	-	-	51000-47700yr old
KX898018.1	steppe bison	Massilani et al., 2016	Diamond and Precious Metal Geology Institute of the Siberian Branch of the RAS	115	Lower Kolyma, Yakutia, East Siberia, Russia	-	-	29600-26000yr old
KX898019.1	steppe bison	Massilani et al., 2016	Diamond and Precious Metal Geology Institute of the Siberian Branch of the RAS	118	Lower Kolyma, Yakutia, East Siberia, Russia	-	-	27300-26000yr old

Sample ID	Clade ID	Source	Museum	Museum ID	Location	δ^{13}C	δ^{15}N	Raw date (C^{14}/ archaeological)
KX898020.1	steppe bison	Massilani et al., 2016	Diamond and Precious Metal Geology Institute of the Siberian Branch of the RAS	124	Lower Kolyma, Yakutia, East Siberia, Russia	-	-	44000-29600yr old
LE237A	CladeX	Soubrier et al., 2016	IAKiev	A-89 KB VI B	Amvrosievka, Ukraine	-	-	18630±220
LE242B	CladeX	Soubrier et al., 2016	IAKiev	A-89 KB 1	Amvrosievka, Ukraine	-	-	18630±220
LE247B	CladeX	Soubrier et al., 2016	IAKiev	A93 K4 b/33	Amvrosievka, Ukraine	-	-	18630±220

Supplementary File 1. Videographic of bison distribution across Eurasia through time with a forest/non-forest paleoenvironmental background (available in the electronic version of this thesis).

Chapter 6

Low genomic diversity in European bison (*Bison bonasus*) prior to their near-extinction during the 20th Century

Manuscript prepared for submission

Statement of Authorship

Title of Paper	Low genomic diversity in European bison (Bison bonasus) prior to their near-extinction during the 20th Century
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☒ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Manuscript prepared in publication format

Principal Author

Name of Principal Author (Candidate)	Ayla L. van Loenen		
Contribution to the Paper	Designed experiments, performed laboratory experiments, compiled data and performed bioinformatics analyses. Analysed and interpreted results. Wrote the paper with comments from co-authors.		
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	3/7/18

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, interpreted results, made comments on the manuscript.
Signature	Date 3/7/18

Name of Co-Author	Bastien Llamas
Contribution to the Paper	Designed experiments, interpreted results, made comments on the manuscript.
Signature	Date 3/7/18

Name of Co-Author	Kieren Mitchell
Contribution to the Paper	Interpreted results, made comments on the manuscript.
Signature	Date 3/7/18

RESEARCH

Low genomic diversity in European bison prior to their near-extinction during the 20th Century.

Ayla L. van Loenen*, Kieren J. Mitchell, Bastien Llamas and Alan Cooper

*Correspondence:
aylalore22@gmail.com
Australian Centre for Ancient DNA (ACAD), University of Adelaide, Adelaide, Australia
Full list of author information is available at the end of the article

Abstract

Untangling long-term patterns of evolutionary history is particularly complicated when researching extant populations after extreme bottlenecks. European bison (*Bison bonasus*) were initially found throughout Europe during the Late Pleistocene/early Holocene until suffering an extreme bottleneck as they went extinct in the wild in the 1920s. The extant population was recovered through a captive breeding program of the 12 remaining European bison. This severe bottleneck was thought to have been the cause of the observed low diversity present in extant European bison. Prior research has focused on historic (20th Century) and modern genomes only, and consequently was unable to accurately characterise the loss of diversity post-bottleneck. Here, we present the first ancient European bison genomes (∼2 thousand years old), and describe patterns of genomic diversity (heterozygosity) through time between ancient, historic and modern European bison. With this information, we can identify that the human-mediated near-extinction event in the 1900s was not the likely cause of the low diversity, but more likely the extension of a period of decreasing diversity that began earlier in the Holocene.

Keywords: European bison (*Bison bonasus*; evolutionary genetics; genomics; low coverage genomes

Introduction

The European bison (*Bison bonasus*) is the largest extant terrestrial mammal in Europe and is an iconic European species [1]. Initially found throughout Europe during the Late Pleistocene and Holocene, by the end of the 19th century European bison were found exclusively in two populations, representing two subspecies [2, 3, 4]: one in Bialoweiża forest (*Bison bonasus bonasus*) and the other in the Caucasus Mountains (*Bison bonasus caucasus*). During the 1920s, both populations were hunted to extinction in the wild, with only a handful remaining in captivity [2, 4]. The 54 remaining captive European bison were all first to third generation descendants from only 12 founders [2, 4]. In an attempt to maintain purity of the *B. b. bonasus* subspecies during the captive breeding program after the extreme bottleneck, the descendants of the 12 founders were divided into two genetic lines: the Lowland line, which was comprised of the offspring of seven 'pure' Lowland founders, while the Lowland-Caucasian line consisted of the offspring from all founders (11 Lowland bison, and one Caucasian bull) that did not qualify for the Lowland line [2, 5, 4]. There are now more than 6,500 bison in Europe, over 4,500 of which are in free or semi-free ranging herds [6, 7]. However, modern European bison are known to have a low genetic diversity, thought to be a result of the extreme bottleneck suffered in the 1900s [8, 3, 9].

A series of genomic data for modern and historic European bison has been published [9, 3], including information from members of both Lowland and Lowland-Caucasian lines (MdL1, MdL2, MdLC, BBO3569, BBO3574), historical 'pure' Caucasian bison (Cc1 and Cc2), and one of the 12 founders of the modern lines and her son (F42/PLANTA and M158/PLANTEN). Both [9] and [3] characterised introgression between European bison and cattle lineages prior to recent bottlenecks. [9] also described a potential decline in population size following the Last Glacial Maximum (LGM) around 20 thousand years ago (kya), and identified a number of genes differentially selected between bison and cattle lineages. However, [9] were restricted to two modern bison genomes (BBO3569 and BBO3574), making it difficult to confidently identify directional patterns of selection between lineages as they lacked genomic information from ancestral lineages. [3] used both modern and historic low coverage genomes to identify patterns of recent admixture between the two subspecies (*B. b. bonasus* and *B. b. caucasus*), occurring prior to extinction in the wild and also during the captive breeding program. Genomes from both the founder/son pair and historical Caucasian bison provided key reference genomes to compare to modern genomes and allowed the characterisation of variation in the proportion of Caucasian bison admixture in modern European bison immediately before and after the bottleneck event [3]. However, the founder/son pair of specimens analysed by [3] were still very recent examples of European bison diversity (early 1900s) and it is known from the historical record that the number of European bison was already significantly reduced by that time [4].

Ancient DNA (aDNA) has been used to directly investigate evolutionary processes through time in European bison [Chapter 5, 8, 10, 11, 12]. Demographic inferences based on mitochondrial DNA [Chapter 5] suggest a pattern of loss of diversity beginning ∼10 kya during the Late Pleistocene/early Holocene. This supports the findings of [9], who suggested that there may have been a decline in effective population size of European bison beginning around 20kya. However, previous aDNA studies were restricted to predominantly mitochondrial DNA information (initially D-loop only, more recently mitochondrial genomes), with a subset of nuclear SNPs for a few specimens [10, 8, 11, 12]. Genomes from ancient specimens prior to the recent bottleneck in the 1900s would further our understanding of the genomic consequences of said bottleneck. For example, while we assume that genomic diversity has been severely affected by the near-extinction event, we lack comparative pre-bottleneck specimens with which to quantify the effect. [3] attempted to address this by incorporating low coverage genomes from a pair of founder/son specimens. However, we know from the historical record that many of the founders were kept isolated in zoos and parks prior to their inclusion in the captive breeding program [4]. In addition, the wild population was already significantly reduced as a result of hunting, and the remaining reduced populations were under the protection of the Tsars prior to their extinction in the wild [4]. As a result, the founders may not accurately represent the true diversity present in wild European bison prior to the population bottleneck. Here we present the first truly ancient (two thousand year old) low coverage genomes for European bison, and characterise the changes in genomic diversity between ancient, historic, and modern European bison. In addition, we also analysed the diversity of two American bison (*Bison bison*) and

two steppe bison (*Bison priscus*) in order to benchmark our analyses, as the demographic history of bison in America is better understood than that of European bison.

Methods

Sample collection and radiocarbon dating

The four ancient specimens used in this study (A18676, A18688, A16121, and A16171) were from the existing collection at the Australian Centre for Ancient DNA (ACAD), University of Adelaide. Samples were radiocarbon dated, and the dates were calibrated with OxCal v4.3 using the IntCal13 curve (Table 1) [13]. Out of the sixteen bison specimens included in this study, fourteen specimens were directly dated and range in age from over 40,000calBP (calibrated years before present) to modern day (Table 1). The remaining two specimens (Cc1 and Cc2) were undated [3]. For clarity, hereafter the specimens are referred to by descriptive names (e.g. AncientEuropean1, ModernEuropean1) instead of by sample IDs (see Table 1) for a complete list of matching names and sample IDs.

DNA analyses

Modern samples

The ModernEuropean1, ModernWoods, and ModernPlains DNA extracts were processed into sequencing libraries using the Illumina TruSeq Nano 350bp following the manufacturer's protocol (Illumina). Libraries were shotgun sequenced on an Illumina NextSeq 500 and HiSeq 2500 platforms (paired-end chemistry) to generate nuclear genomes.

Ancient samples

All ancient DNA work was performed in the purpose-built ancient DNA facilities at ACAD following previously published guidelines [14, 15]. Ancient DNA was extracted from bone fragments following the silica based extraction protocol from [16] (as described in [8]). Double-stranded Illumina sequencing libraries were built from 25μl of DNA extract following the partial uracil-DNA-glycosylase (UDG) treatment protocol from [17] modified to include dual 7-mer internal barcode sequences (as described in [8]). High-throughput sequencing on an Illumina HiSeq X Ten was used to generate low coverage nuclear genomes. Sequence data from the two new specimens were combined with previously published European bison genomic sequence data available on NCBI/SRA, and modern bison genomic sequencing data described above.

Sequencing reads from both the newly generated data and previously published data from Wecek et al., (2016) and Gautier et al., (2016) were mapped to a published genome of the cow (UMD3.1). We used the programs AdapterRemoval v2.2.1 [18], BWA v0.7.15 [19], Picard Tools (http://broadinstitute.github.io/picard), Genome Analysis ToolKit [20], SAMtools v1.3.1 [21, 22], and mapDamage v2.0.6 [23], as implemented in the pipeline PALEOMIX v1.2.9 [24] and as described in [8]. The summary sequencing statistics for each specimen were compiled from the PALEOMIX pipeline output (Supplementary Table 1). Damage patterns for the ancient specimens were consistent with ancient DNA (Supplementary Table 2). Pseudo-haploid

sequences were generated by random allele selection, with filtering to exclude indels and transitions, using scripts available from (https://github.com/grahamgower/eig-utils). Transitions and indels were excluded because low coverage genomes do not provide enough data to distinguish true SNPs from potential post-mortem DNA damage (commonly observed as C to T transitions). A Maximum-Likelihood tree (ML tree) incorporating all the genomes used in this study was generated in RaxML [25] using pseudo-haploid sequences (algorithim: a, bootstrap replicates: 100, model: ASC_GTRGAMMA).

Estimates of heterozygosity (θ)

Heterozygosity was estimated using the estimateTheta tool within ATLAS [26, 27]. Unlike other tools, ATLAS does not require a priori estimates of population genetics parameters (such as mutation rates or population size estimates), instead calculating them on a sliding window basis, and as such it can be used for species (like European bison) for which insufficient high coverage genomic data are available to infer accurate estimates of these parameters. We used estimateTheta to calculate estimates of heterozygosity for non-overlapping 1Mb blocks across the genome. To ensure our results were not biased by window size we also re-ran the analysis using 10Mb blocks for genomes with depth of coverage less than 10 (Table 1), as it was been demonstrated that estimates across larger windows are more accurate for low coverage genomes[26]. The results were plotted as a violin plot using the `ggplot2` package in `R` [28].

Controlling for batch effects

Bioinformatics analyses involving low coverage genomes can be sensitive to potential 'batch effects' where data generated following different methods can appear more similar or different than they truly are [29, 30]. While the data used in this study have been generated using different DNA extraction and library preparation techniques (as a result of being produced by different research groups), all sequencing data included here was generated using Illumina sequencing platforms [3, 9]. To reduce the potential for batch effects, the raw sequencing data from all specimens was collectively processed through the same bioinformatics pipeline described above.

Results

Our maximum likelihood phylogenetic analysis produced a tree with very high bootstrap support (all nodes 100%) The results of our maximum likelihood phylogenetic analysis support the reciprocal monophyly of American bison and European bison, and within European bison there are several distinct clades (Figure 1). Our two ancient European bison (AncientEuropean1 and AncientEuropean2) form a clade that is sister to a clade comprising the modern/historic individuals, within which the historical Caucasian bison are reciprocally monophyletic with respect to the modern diversity. Interestingly, the modern Lowland line bison form a monophyletic clade to the exclusion of the sole modern descendent of the Lowland-Caucasus line included in this study (ModernEuropean4) and the founder/son pair (HistoricalFounder1 and HistoricalFounder2).

The founder/son specimens (HistoricalFounder1 and HistoricalFounder2) show high levels of heterozygosity (Figure 2), with the lower heterozygosity values of the remaining bison individuals forming two broad 'clusters'. Within these two clusters, cluster 1 includes AncientSteppe2, HistoricalCaucaus1, HistoricalCaucasus2, ModernEuropean3, ModernEuropean2, ModernEuropean4, and AncientEuropean2, while cluster 2 includes ModernEuropean6, AncientEuropean1, ModernEuropean5, ModernWoods, ModernEuropean1, AncientSteppe1, and ModernPlains (Figure 2). There was no obvious clustering that corresponded to different library preparation protocols or levels of ancient DNA damage (that could have indicated batch effects). Estimates of heterozygosity did not vary substantially based on the size of the window used (Figure 2).

Discussion

The American and steppe bison specimens (AncientSteppe1, AncientSteppe2, ModernWoods, and ModernPlains) exhibited a pattern consistent with what is know about their evolutionary history. The older ancient specimen (AncientSteppe2) fell within the higher group of diversity, exhibiting the second highest observed values other than the European founder individuals. A high value for the older steppe bison is fully consistent with the results of demographic analyses based on mitochondrial data, which suggest that steppe bison populations peaked in size during this period [31]. The younger ancient steppe bison specimen grouped within the lower cluster of diversity, likewise consistent with previous demographic analyses, which suggested that steppe bison populations declined markedly beginning ∼30 kya [31]. For the modern plains and woods bison, this is likely reflective of their history of severe population bottlenecks [32]. Both plains and woods bison are known to have low genetic diversity, with plains bison in the United States in particular all descended from a single population in Yellowstone National Park, which was reduced to 25 individuals in the 1900s [32]. These results suggest that estimations of theta from low-coverage bison genomes do recapitulate effective population size.

Surprisingly, our estimates of heterozygosity (as a proxy for genomic diversity and effective population size) suggest that a major loss of diversity may have occurred in the nuclear genome of European bison prior to the drastic bottleneck they experienced in the 1920s. The ancient European bison samples presented in this study do not exhibit substantially higher diversity than historical and modern European bison, which would be expected if a major loss of diversity among European bison occurred in the 20^{th} Century. This suggests that the levels of diversity observed in historic and modern European bison are comparable to the levels of diversity present in the ancient Austrian bison around two thousand years ago.

The lack of diversity in European bison over the last two thousand years supports prior theories of subsistence of European bison populations in sub-optimal environments/habitat fragmentation during the Late Holocene [Chapter 5, 3, 9]. Following the Neolithic expansion and forestation of Europe during the mid Holocene, the geographic and temporal distribution patterns of European bison remains suggest they were likely surviving in small isolated population groups in fragmented habitat [3, 9, 33, 34, 35]. If the ancient European bison were already subsisting in small habitat fragments, this would result in smaller isolated populations with limited opportunities for gene flow. As a result, the effective population size was likely much

smaller (supported by [9]), which would in turn explain why patterns of diversity may be similar between modern, historic, and two thousand year old European bison. As such, the potential loss of diversity immediately following the near extinction of European bison in the 1920s may not be as significant as previously considered. Instead, the reduced diversity observed in modern European bison may be the extension of a loss of diversity that began much earlier than two thousand years ago (the age of both our ancient European bison specimens). Mitochondrial genome data from a large number of European bison specimens has previously suggested that the loss of diversity (in the mitochondrial genome) began as early as ∼10kya [Chapter 5]. [9] also identified a trend towards decreasing population size following the LGM from ∼20kya to present based on modern genomic data (specifically ModernEuropean5 and ModernEuropean6).

Interestingly, the historical founder/son pair (HistoricalFounder1, HistoricalFounder2) [3] have unusually high diversity. Until the late 19th century European bison and ancestral species of domestic cow (aurochs or *Bos primigenius*, Bojanus, 1827) were considered to be the same species, often collectively referred to as 'auerochs' in historical records [36]. Consequently, it is possible that the recent ancestors of the founder individuals were housed together and subsequently admixed with a *Bos* species, potentially explaining the unusually high proportions of heterozygosity identified in the founder specimens (HistoricalFounder1, HistoricalFounder2). However, this would likely also result in ancestry 'blocks' across the genome, which were not identified in the pairwise distance estimates performed by [3]. Alternatively, the high heterozygosity of the founders may be because they are recent descendants of admixture between Caucasian wisent and Lowland wisent populations [3]. Based on pedigrees, HistoricalFounder1 is known to be a recent descendent of a Caucasian bull (M100/KAUKASUS), and based on pairwise distance estimates from [3] is thought to have passed on significant blocks of Caucasian ancestry to her son (HistoricalFounder2). As recent descendants of an admixture event, this could explain the temporary increase in diversity observed in these specimens that is lost again in their modern descendants due to genetic drift (ModernEuropean2, ModernEuropean3, ModernEuropean4). This pattern of increased diversity is also observed in the historical Caucasus bison, which are similarly highly structured and consequently may also represent recent mixing between distantly related populations [3]. These hypotheses are supported by our phylogenetic results, which show that the historic Caucasian bison represent a distinct lineage (Figure 1), and by pairwise distance results from [3]. Lowland wisent and Caucasian wisent were geographically isolated for some time prior to their extinction in the wild and may therefore have had sufficiently different allele frequencies to cause such an increase in diversity immediately following an introgression event.

Conclusion

We have demonstrated that there does not appear to have been a substantial loss of diversity between our ancient specimens and modern European bison. This suggests that European bison may have subsisted in small isolated habitat fragments for extended periods during the Holocene. This is supported by data on distribution and diet presented in Chapter 5, and population size estimates from [9]. Alternatively,

this pattern may reflect a recovery of diversity following the bottleneck as a result of introgression between two geographically distinct populations (*B. b. caucasus* and *B. b. bonasus*) identified in the founder/son pair [3]. While we took all practical steps to reduce potential batch effects, and did not observe any obvious shared patterns between genomes sequenced in the same studies, there is a potential that these results may be biased due to differences in laboratory methods between studies. A potential approach to control for this involves applying a new 'recal' tool in ATLAS (https://bitbucket.org/phaentu/atlas/wiki/Home) to recalibrate the base quality scores using a region of the genome known to be monomorphic (such as ultraconserved regions). This approach is being developed to account for an potential biases introduced in ancient DNA sequencing due to variation in laboratory methods or sequencing technologies. However this tool is still under development, and requires additional verification. The inclusion of genomic data from additional (or older) European bison specimens would provide additional reference time points to more specifically characterise potential losses or gains of diversity over time, and validate the current results.

Competing interests
The authors declare that they have no competing interests.

Author's contributions
ALvL, BL and AC designed experiments, ALvL, KJM, BL and AC interpreted results, ALvL wrote the manuscript with comments from all co-authors.

Acknowledgements
ALvL was supported by an Australian Government Research Training Program (RTP) Scholarship. K.J.M, B.L., and A.C. are supported by the Australian Research Council.

References

1. 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**(2), 0115090 (2015)
2. Pucek, Z.: European Bison (Bison Bonasus): Current State of the Species and Strategy for Its Conservation, Council of Europe Publishing (2004)
3. Węcek, K., Hartmann, S., Paijmans, J.L.A., Taron, U., Xenikoudakis, G., Cahill, J.A., Heintzman, P.D., Shapiro, B., Baryshnikov, G., Bunevich, A.N., Crees, J.J., Dobosz, R., Manaserian, N., Okarma, H., Tokarska, M., Turvey, S.T., Wójcik, J.M., Żyła, W., Szymura, J.M., Hofreiter, M., Barlow, A.: Complex admixture preceded and followed the extinction of wisent in the wild. Molecular Biology and Evolution **34**(3), 598–612 (2016)
4. Krasińska, M., Krasiński, Z.: European Bison: The Nature Monograph, Springer Berlin Heidelberg (2013)
5. Olech, W.: The influence of inbreeding on European bison sex ratio. Animals, Zoo and Conservation. E. Zgrabczynska, P. Cwiertnia, J. Ziomek eds. Published by The Zoological Garden in Poznañ, 29–33 (2006)
6. van de Vlasakker, J.: Rewilding Europe Bison Rewilding Plan, 2014-2024, Rewilding Europe, Nijmegen, The Netherlands (2014)
7. EBPB: European Bison Pedigree Book, Białowieża National Park (2016)
8. Soubrier, J., Gower, G., Chen, K., Richards, S.M., Llamas, B., Mitchell, K.J., Ho, S.Y.W., Kosintsev, P., Lee, M.S.Y., Baryshnikov, G., Bollongino, R., Bover, P., Burger, J., Chivall, D., Crégut-Bonnoure, E., Decker, J.E., Doronichev, V.B., Douka, K., Fordham, D.A., Fontana, F., Fritz, C., Glimmerveen, J., Golovanova, L.V., Groves, C., Guerreschi, A., Haak, W., Higham, T., Hofman-Kamińska, E., Immel, A., Julien, M.-A., Krause, J., Krotova, O., Langbein, F., Larson, G., Rohrlach, A., Scheu, A., Schnabel, R.D., Taylor, J.F., Tokarska, M., Tosello, G., van der Plicht, J., van Loenen, A., Vigne, J.-D., Wooley, O., Orlando, L., Kowalczyk, R., Shapiro, B., Cooper, A.: Early cave art and ancient DNA record the origin of European bison. Nature Communications **7**, 13158 (2016)
9. Gautier, M., Moazami-Goudarzi, K., Levéziel, H., Parinello, H., Grohs, C., Rialle, S., Kowalczyk, R., Flori, L.: Deciphering the Wisent Demographic and Adaptive Histories from Individual Whole-Genome Sequences. Molecular Biology and Evolution **33**(11), 2801–2814 (2016)
10. Marsolier-Kergoat, M.-C., Palacio, P., Berthonaud, V., Maksud, F., Stafford, T., Bégouën, R., Elalouf, J.-M.: Hunting the Extinct Steppe Bison (Bison priscus) Mitochondrial Genome in the Trois-Frères Paleolithic Painted Cave. PLoS ONE **10**(6) (2015)
11. Massilani, D., Guimaraes, S., Brugal, J.-P., Bennett, E.A., Tokarska, M., Arbogast, R.-M., Baryshnikov, G., Boeskorov, G., Castel, J.-C., Davydov, S., Madelaine, S., Putelat, O., Spasskaya, N.N., Uerpmann, H.-P., Grange, T., Geigl, E.-M.: Past climate changes, population dynamics and the origin of Bison in Europe. BMC Biology **14**(1), 93 (2016)

12. Onar, V., Soubrier, J., Toker, N.Y., van Loenen, A., Llamas, B., Siddiq, A.B., Pasicka, E., Tokarska, M.: Did the historical range of the European bison (Bison bonasus L.) extend further south?—a new finding from the Yenikapı Metro and Marmaray excavation, Turkey. Mammal Research **62**(1), 103–109 (2017)
13. Reimer, P.J., Bard, E., Bayliss, A., Beck, J.W., Blackwell, P.G., Ramsey, C.B., Buck, C.E., Cheng, H., Edwards, R.L., Friedrich, M., Grootes, P.M., Guilderson, T.P., Haflidason, H., Hajdas, I., Hatté, C., Heaton, T.J., Hoffmann, D.L., Hogg, A.G., Hughen, K.A., Kaiser, K.F., Kromer, B., Manning, S.W., Niu, M., Reimer, R.W., Richards, D.A., Scott, E.M., Southon, J.R., Staff, R.A., Turney, C.S.M., Plicht, J.v.d.: IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. Radiocarbon **55**(4), 1869–1887 (2013)
14. Cooper, A., Poinar, H.N.: Ancient DNA: do it right or not at all. Science **289**(5482), 1139–1139 (2000)
15. Hofreiter, M., Shapiro, B. (eds.): Ancient DNA: Methods and Protocols. Methods in Molecular Biology, Humana Press Incorporated, New York (2012)
16. Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Jane Adler, C., Richards, S.M., Sarkissian, C.D., Ganslmeier, R., Friederich, S., Dresely, V., van Oven, M., Kenyon, R., Van der Hoek, M.B., Korlach, J., Luong, K., Ho, S.Y.W., Quintana-Murci, L., Behar, D.M., Meller, H., Alt, K.W., Cooper, A., Consortium, T.G.: Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. Nature Communications **4**, 1764 (2013)
17. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., Reich, D.: Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. Phil. Trans. R. Soc. B **370**(1660), 20130624 (2015)
18. Lindgreen, S.: AdapterRemoval: easy cleaning of next-generation sequencing reads. BMC Research Notes **5**(1), 337 (2012)
19. Li, H., Durbin, R.: Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics **25**(14), 1754–1760 (2009)
20. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., *et al.*: The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research **20**, 1297–1303 (2010)
21. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.: The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**(16), 2078–2079 (2009)
22. Li, H.: A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics **27**(21), 2987–2993 (2011)
23. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L., Orlando, L.: mapdamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics **29**(13), 1682–1684 (2013)
24. Schubert, M., Ermini, L., Sarkissian, C.D., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M.D., Fernández, R., Kircher, M., McCue, M., Willerslev, E., Orlando, L.: Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nature Protocols **9**(5), 1056–1082 (2014)
25. Stamatakis, A.: RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**(9), 1312–1313 (2014)
26. Kousathanas, A., Leuenberger, C., Link, V., Sell, C., Burger, J., Wegmann, D.: Inferring Heterozygosity from Ancient and Low Coverage Genomes. Genetics **205**(1), 317–332 (2017)
27. Link, V., Kousathanas, A., Veeramah, K., Sell, C., Scheu, A., Wegmann, D.: ATLAS: Analysis Tools for Low-depth and Ancient Samples. bioRxiv, 105346 (2017)
28. R Development Core Team: R: A Language and Environment for Statistical Computing. Vienna, Austria (2017). http://www.R-project.org
29. Lambert, C.G., Black, L.J.: Learning from our GWAS mistakes: from experimental design to scientific method. Biostatistics **13**(2), 195–203 (2012)
30. Leigh, D., Lischer, H., Grossen, C., Keller, L.: Batch effects in a multiyear sequencing study: False biological trends due to changes in read lengths. Molecular Ecology Resources **18**(4), 778–788 (2018)
31. Shapiro, B., Drummond, A.J., Rambaut, A., Wilson, M.C., Matheus, P.E., Sher, A.V., Pybus, O.G., Gilbert, M.T.P., Barnes, I., Binladen, J., Willerslev, E., Hansen, A.J., Baryshnikov, G.F., Burns, J.A., Davydov, S., Driver, J.C., Froese, D.G., Harington, C.R., Keddie, G., Kosintsev, P., Kunz, M.L., Martin, L.D., Stephenson, R.O., Storer, J., Tedford, R., Zimov, S., Cooper, A.: Rise and Fall of the Beringian Steppe Bison. Science **306**(5701), 1561–1565 (2004)
32. Hedrick, P.W.: Conservation genetics and north american bison (bison bison). Journal of Heredity **100**(4), 411–420 (2009)
33. Hoogaker, B., Smith, R.S., Singarayer, J.S., Marchant, R., Prentice, I., Allen, J., Anderson, S., Bhagwat, S., Behling, H., Borisova, O., *et al.*: Terrestrial biosphere changes over the last 120 kyr. Climate of the Past. **12**(1), 51–73 (2016)
34. Kaplan, J.O., Pfeiffer, M., Kolen, J.C.A., Davis, B.A.S.: Large Scale Anthropogenic Reduction of Forest Cover in Last Glacial Maximum Europe. PLOS ONE **11**(11), 0166726 (2016)
35. Leuschner, C., Ellenberg, H.: Ecology of Central European Forests: Vegetation Ecology of Central Europe, Springer (2017)
36. van Vuure, C.: Retracing the Aurochs: History, Morphology and Ecology of an Extinct Wild Ox, Pensoft Pub (2005)

Figure and Tables

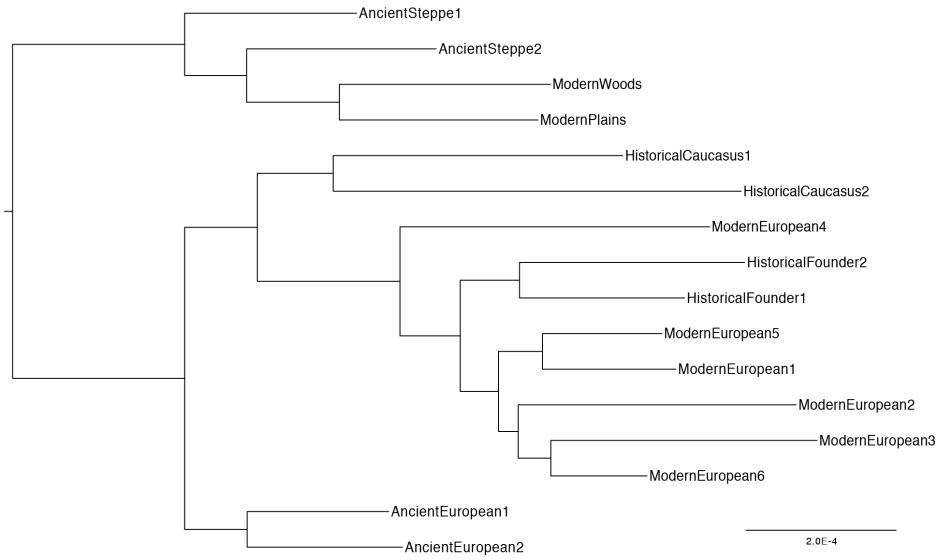


Figure 1 RAxML ML tree including all genomes used in this study (using pseudo-haploid genome sequences).

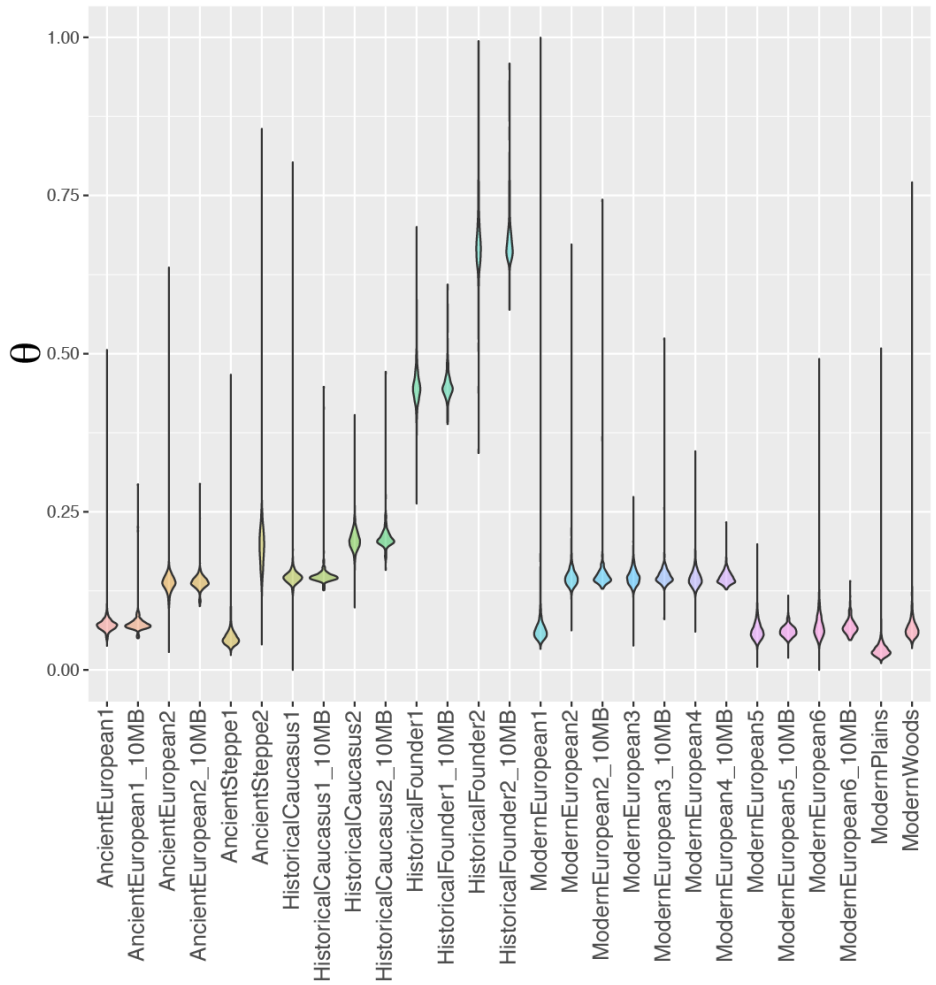


Figure 2 Violin plot of estimates of heterozygosity (θ) for all samples (window size 1MB), with replicated estimates of heterozygosity (θ) for all specimens with an average depth of coverage less than 10 (replicated with an increased window size 10MB). The violin plot illustrates kernel probability density, i.e. the width of the shaded area represents the proportion of the data located there.

Table 1 Sample information

Name	Sample ID	Source	Species ID	Museum	Museum ID	Location	Age
AncientEuropean1	A18676	This study	*B. bonasus*	Institut für Paläontologie an der Universität Wien	-	Aschbach, Austria	1629±36
AncientEuropean2	A18688	This study	*B. bonasus*	Natural History Museum Vienna (VMNH)	H77-49-1	Styria, Austria	2038±88
AncientSteppe1	A16121	This study	*B. priscus*	Yukon palaeontology program	-	Eastern Beringia, Canada	12530±55
AncientSteppe2	A16171	This study	*B. priscus*	Yukon palaeontology program	408.53	Hawk mine, Eastern Beringia, Canada	36450±750
HistoricalCaucasus1	Cc1	Węcek et al., 2016	*B. b. caucasus*	Zoological Institute RAS, Sankt Petersburg, Russia		Kuban Oblast, Russia	-
HistoricalCaucasus2	Cc2	Węcek et al., 2016	*B. b. caucasus*	State Darwin Museum, Moscow, Russia		North Ossetia-Alania, Russia	-
ModernEuropean1	EuropeanBison	This study	*B. bonasus*	-	-	Białowieża Forest, Poland	modern
ModernEuropean2	MdL1	Węcek et al., 2016	*B. bonasus*	-	-	Białowieża Forest, Poland	2006
ModernEuropean3	MdL2	Węcek et al., 2016	*B. bonasus*	-	-	Białowieża Forest, Belarus	2009
ModernEuropean4	MdLC	Węcek et al., 2016	*B. bonasus*	-	-	Białowieża Forest, Poland	2012
ModernEuropean5	BBO3569	Gautier et al., 2016	*B. bonasus*	-	-	Białowieża Forest, Poland	modern
ModernEuropean6	BBO3574	Gautier et al., 2016	*B. bonasus*	-	-	Białowieża Forest, Poland	modern
HistoricalFounder1	F42 (PLANTA)	Węcek et al., 2016	*B. b. bonasus*	Upper Silesian Museum (Poland)		Pszczyna, Poland	1931
HistoricalFounder2	M158 (PLANTEN)	Węcek et al., 2016	*B. b. bonasus*	Upper Silesian Museum (Poland)		Pszczyna, Poland	1933
ModernPlains	PlainsBison	This study	*B. priscus*	-	-	USA	modern
ModernWoods	WoodBison	This study	*B. priscus*	-	-	USA	modern

Supplementary Information

Supplementary Table1. PALEOMIX supplementary values

Name	Sample ID	RetainedReads	HitsRaw	HitsUniq	Endogenous	Clonal	Coverage_Mit	Depth_Mit	Coverage_Nuc	Depth_Nuc
AncientEuropean1	A18676	535661644	226437194	186596539	0.42	0.18	1	288.83	0.7	3.01
AncientEuropean2	A18688	440268808	176701555	151555324	0.4	0.14	1	201.24	0.72	2.7
AncientSteppe1	A16121	1357061970	826386884	666250846	0.61	0.19	1	2023.96	0.9	15.61
AncientSteppe2	A16171	834951509	477255135	413846592	0.57	0.13	1	918.84	0.9	10.85
HistoricalCaucasicus1	Cc1	370322404	93133254	89112782	0.25	0.04	1	76.78	0.75	2.05
HistoricalCaucasicus2	Cc2	377301034	86186049	84053178	0.23	0.02	1	294.24	0.74	1.97
ModernEuropean1	EuropeanBison	558730449	506868788	489450420	0.91	0.03	1	42379.95	0.95	27.11
ModernEuropean2	MdL1	180174985	143074293	140012316	0.79	0.02	1	1754.39	0.91	3.87
ModernEuropean3	MdL2	176805402	144670378	141532507	0.82	0.02	1	1831.31	0.91	3.91
ModernEuropean4	MdLC	161306362	132005051	128975939	0.82	0.02	1	4644.13	0.9	3.55
ModernEuropean5	BBO3569	232315015	228743534	228488727	0.98	0	1	405.32	0.93	8.46
ModernEuropean6	BBO3574	265833657	261419653	260902458	0.98	0	1	310.29	0.93	9.81
HistoricalFounder1	F42 (PLANTA)	475502981	237681548	231184388	0.5	0.03	1	1244.17	0.87	5.66
HistoricalFounder2	M158 (PLANTEN)	444188730	161957426	158373491	0.36	0.02	1	2512.69	0.81	3.6
ModernPlains	PlainsBison	797793746	695734967	691295535	0.87	0.01	1	791.49	0.95	24.77
ModernWoods	WoodBison	490945347	426076994	413094513	0.87	0.03	1	3027.26	0.95	23.1

Supplementary Table 2. MapDamage values

Name	Sample ID	3' G>A frequency	5' C>T frequency
AncientEuropean1	A18676	0.0130	0.0217
AncientEuropean2	A18688	0.0139	0.0249
AncientSteppe1	A16121	0.0899	0.0879
AncientSteppe2	A16171	0.0536	0.0538
HistoricalCaucasicus1	Cc1	0.0278	0.0177
HistoricalCaucasicus2	Cc2	0.0158	0.0518
ModernEuropean1	EuropeanBison	0.0088	0.0080
ModernEuropean2	MdL1	0.0080	0.0052
ModernEuropean3	MdL2	0.0081	0.0064
ModernEuropean4	MdLC	0.0083	0.0055
ModernEuropean5	BBO3569	0.0076	0.0074
ModernEuropean6	BBO3574	0.0075	0.0070
HistoricalFounder1	F42 (PLANTA)	0.0064	0.0349
HistoricalFounder2	M158 (PLANTEN)	0.0041	0.0286
ModernPlains	PlainsBison	0.0042	0.0080
ModernWoods	WoodBison	0.0094	0.0074

Chapter 7

Discussion and Conclusion

Chapter 7: Discussion and Conclusion

Thesis summary and significance

Understanding the complex and dynamic patterns and processes driving the evolutionary history of species is key to furthering our understanding of how species evolve. The aim of my thesis was to develop and identify approaches to combine multiple genetic and non-genetic datasets in order to improve our understanding of the evolutionary history of species, contextualised by environmental change. By improving our understanding of how species have evolved over time and responded to change, we can use this knowledge to forecast likely responses to change in the future and consequently develop appropriate conservation management plans to address such changes. I have specifically focused on the incorporation of key non-genetic datasets that can be combined with genetic data in order to generate a more accurate understanding of the evolutionary history of species. In addition, I have provided novel strategies to improve the recovery and processing of genetic data from ancient specimens and developed a series of approaches to integrate DNA data with large associated non-genetic datasets such as environmental predictions, geographic and temporal information, historical records, and ecological information. I applied these approaches to European bison in order to delineate patterns of evolutionary change and identify potential environmental drivers. Below, I summarise my findings and discuss their implications and significance for our understanding of the evolutionary history of European bison, and for research involving ancient DNA and evolutionary genetics generally.

Summary

In Chapters 2 and 3, I investigated the implications of various improvements on molecular techniques and sequencing technologies on current aDNA research. Future projects, where improving the spatial and temporal scale of samples is of key importance to our understanding of the evolutionary history of species, will

benefit from improved aDNA recovery. More specifically, in Chapter 2 I developed a novel bioinformatics method to characterise the distribution of post-mortem DNA damage (PMD) across mitochondrial genomes using HTS data, and showed that PMD occurs randomly across the genome. This method can also be used to identify appropriate consensus calling thresholds for downstream analyses, which is particularly important for analyses incorporating ancient DNA. In Chapter 3 I demonstrated that archived DNA extracts (stored for over 10 years) can be successfully re-sequenced using modern techniques in order to generate whole mitochondrial genomes and nuclear data (including specimens which were unsuccessful with Sanger sequencing approaches). I identified potential links between treatment of the archived extracts and improved recovery of DNA, and outlined approaches to efficiently generate additional DNA data. The findings of Chapter 3 are especially relevant for specimens that cannot be re-extracted. Collectively Chapters 2 and 3 validate methods for successfully obtaining ancient DNA data from larger numbers of low quality sub-fossil material. I apply these methods in subsequent chapters to generate a large number of ancient and historic mitochondrial genomes, and two ancient genomes of European bison. These methods will allow future projects to expand the spatial and temporal scale covered by specimens by improving the recovery of ancient DNA data. Inclusion of a large number of samples with a variety of spatial and temporal information is key for improving our understanding of the evolutionary history of species.

In Chapter 4, I critically reviewed the current literature on the evolutionary history of European bison, and discussed the differing points of view currently proposed in the literature. In particular, I addressed the validity of recent claims surrounding the continued survival of *Bison schoetensaki* into the Late Pleistocene, and suggested expanding the geographic and temporal coverage of the sampling, and including more detailed supplementary non-genetic datasets to help improve our understanding. These datasets include geographic and temporal specimen metadata, historical records, and recently developed palaeoenviromental models. This chapter identified key gaps in our

understanding of the evolutionary history of European bison, a number of which I subsequently addressed in Chapters 5 and 6.

In Chapter 5 I used the approaches described in Chapters 2 and 3 to generate a large dataset of ancient, historical, and modern, European bison mitochondrial genome sequences. This is the largest genetic dataset of European bison to date, spanning the period from ~70 thousand years ago (kya) until modern times, and includes a broad spatial coverage, with specimens from Scandinavia to Italy and from the North Sea to Siberia. I demonstrated the increased resolution provided by a large well-annotated dataset by refining previous divergence date estimates, and revealing geographic and temporal patterns of distribution cryptic to previous smaller scale studies. The inclusion of complementary historical, ecological and environmental data supported the characterisation of evolutionary history and further contributed towards the identification of potential environmental drivers. For example, with the dataset presented in Chapter 5 I was able to observe patterns of response to major events occurring during the Holocene, including the fall of the Western Roman Empire and the Black Death pandemic. These historical events had a significant impact on the human populations and environmental conditions, which in turn impacted the distribution and number of European bison over time.

Finally, in Chapter 6, I generated the two oldest European bison genomes to date (~2 thousand years old), and combined these genomes with previously published historical and modern European bison genomes for subsequent analysis. For the first time, I analysed genomic diversity (heterozygosity) through time between ancient, historic, and modern European bison. With this additional information I can now identify that the human-mediated near-extinction event in the 1920s was not the likely cause of low diversity suffered by European bison populations today, but rather the extension of a period of low diversity already in progress since the early Holocene. This provided additional support to theories proposed in Chapter 5 from mitochondrial analyses. That is, that European bison were already suffering from a reduction in diversity from ~10kya, which may

have been a consequence of extended periods of subsistence in habitat fragments between the Late Pleistocene and the modern day.

Collectively the methods and approaches demonstrated in the previous chapters can be used to compile and analyse large 'total-evidence' datasets to provide key information to untangle the evolutionary history of individual species. In an extension of the single-species approach demonstrated in this thesis, individual datasets from multiple species (e.g. humans or other megafauna such as aurochs/mammoth) can be combined to identify common patterns and consequently potential drivers or causes for shared patterns of evolutionary history (e.g. population replacements, mass extinctions, and haplotype frequency changes). Those datasets can be combined with the expanded European bison datasets generated in Chapters 5 and 6, and other complementary non-genetic datasets (e.g. palaeoenvironmental models, historical records, or other species with equivalent datasets such as humans) to identify potentially similar patterns of change over the same time period. Analyses of a combination of multiple datasets, including data from other species (such as humans or aurochs), can also help untangle the relationship between environmental and climatic changes, human presence, and the interactions between other species in an ecosystem. This approach requires an extensive series of cross-discipline collaborations and expertise, but would ultimately lead to a better understanding of the interactions between species and their environment, as well as improving our interpretations of evolutionary history and application of those interpretations in conservation management plans.

Significance

A detailed understanding evolutionary history is key for developing appropriate conservation management strategies for species currently under threat or predicted to come under threat in future. DNA from ancient specimens can be used to directly observe genetic changes through the evolutionary history of a species as well as reconstruct the demographic history of both extinct and extant species and populations (Leonard, 2008; Orlando and Cooper, 2014).

However, the small number of specimens typically included in ancient DNA studies often limits their interpretations to fairly general or location specific conclusions. Further refinement of techniques for sequencing poor quality samples and the inclusion of supplementary non-genetic data, such as those approaches described in my thesis, will improve the resolution of future research on the evolutionary history of species. Comprehensive continental or global scale studies of the evolutionary history of a species across its former temporal and geographic range will allow for the delineation of specific patterns of genetic change through the evolutionary history of that species. In combination with concurrent advances in related palaeoenviromental disciplines (such as palaeoecology and palaeoclimate), patterns of genetic change can be interpreted in the context of environmental changes in order to identify potential drivers. As demonstrated in this thesis, including a large number of specimens across a broad geographic and temporal range is key for successfully delineating continental patterns of genetic change and identifying concurrent environmental changes that may be acting as potential drivers (Chapters 5 and 6).

Ultimately the research presented in my thesis contributes towards the 'next-generation' of ancient DNA research by highlighting a series of approaches for future research into the evolutionary history of species. These approaches can be used to generate equivalent datasets of multiple species, which could ultimately be combined to identify shared patterns of evolution across multiple species (e.g. the shared distribution of humans and European bison, or characterisation of mass extinction phases such as the Late Pleistocene/early Holocene megafaunal extinctions). The remainder of this chapter summarises some of the limitations encountered during the course of my research, and discusses potential approaches that could be implemented in future to address them, as well as potential further research directions that can build upon the work presented in this thesis.

Limitations and Future Directions

Although the combination of genetic and non-genetic datasets included in this thesis gave me the opportunity to investigate in further detail a number of

questions about the evolutionary history of bison in Europe, a number of questions remain to be answered. Several limitations that became apparent during my research are inherent to the specific datasets generated in this thesis, while others are linked to the analyses used, and some are general issues inherent to ancient DNA research and other palaeodisciplines (e.g. palaeoecology). Below I identify and discuss these limitations and potential opportunities to expand on my research in the future.

Fossil discovery/limitations in sample size

As I showed in Chapter 5 of this thesis, expanding the temporal and geographic range covered by specimens included in subsequent analyses substantially improves the accuracy of our interpretations of evolutionary history. However, there are still a few key gaps in the temporal coverage of European bison specimens. For example, though we have now multiple specimens with 'wisent' mitochondrial lineages prior to the Holocene (as opposed to just one), we still lack sufficient numbers of specimens to characterise the extent of their late Pleistocene distribution. Nor are we able to identify if there were distinct ecological preferences differentiating wisent and specimens with 'CladeX' mitochondrial lineages during that time. In addition, there is a temporal gap in dated European bison specimens between a 9.8kya specimen in Scandinavia, and a 6.8kya specimen in Poland. Given the extensive sampling effort undertaken as part of my project, this may be an accurate reflection of European bison distribution. During that period was the height of the forest extent across Europe, and bison may have retreated to a refugium until ~7kya when the forest cover began to thin again. In order for wisent to re-appear in Poland 6.8kya, they must have been living in a refugium somewhere on the Eurasian continent between 9kya and 7kya. However, the precise location of this refugium (or refugia) remains unknown.

Ancient specimens are often difficult to source, irreplaceable, and for many species there are only limited specimens available in collections around the world. Limited sample sizes are common across all palaeontological studies, as

the number of specimens available is not only dependent on the sampling effort, but also on a number of preservation (taphonomic) biases (Behrensmeyer, 1984; Kidwell and Flessa, 1995; Behrensmeyer et al., 2000; Benton et al., 2011). Multiple taphonomic processes can impact specimen preservation, including: climate at time of death (from hot and humid to permafrost), habitat preferences (e.g. caves, forests, steppe-grassland), and probability of immediate burial compared to long-term exposure of the specimen, which is largely dictated by the depositional environment (e.g. river valley or hill slope) (Behrensmeyer, 1984; Kidwell and Flessa, 1995; Behrensmeyer et al., 2000; Benton et al., 2011).

A recent paper by Block et al. (2016) proposes a new method for fossil discovery by generating a species distribution model (SDM) to estimate prior distribution, which can be combined with data on the likelihood of fossil discovery in the field and estimates of geological suitability for fossil preservation. Given the appropriate environmental and geographical data to construct a SDM, this method could be used to identify potential locations for the refugium European bison retreated to between 9kya and 7kya, or alternatively where additional late Pleistocene specimens may be found. Additional specimens from either of these key time periods would further our understanding of the evolutionary history of European bison. Further improvements in palaeoenvironmental modelling are required to address this issue.

Limitations of palaeoenvironmental data

In Chapter 5 of this thesis, I used stable isotope data and palaeoecological predictions of forested/non-forested grid zones covering the last 70 thousand years to infer the likely environment for each specimen in my dataset. With these combined data I was able to characterise potential variation in ecological patterns between CladeX and wisent during the Late Pleistocene, and identify an apparent transition in European bison dietary patterns from a predominantly grazing diet during the Late Pleistocene to a more mixed-feeding diet during the Holocene. However, the categories I used (forest vs non-forest) were coarse, and likely did not encapsulate the variety of biomes and forest cover that existed in reality. In

addition, the resolution of the grid cells I used was one degree latitude/longitude, with 'snapshots' taken every 1,000 years. This geographic and temporal range could include a variety of ecological biomes across the entire grid cell. For example, there may be small local regions of open habitat within a grid cell characterised as 'forest' biome, or short periods (less than 500 years) where the grid cell was predominantly 'non-forest' biome. This is due to the limitations of the data generated by the hadcm3 models (Pope et al., 2000; Gordon et al., 2000; Kaplan et al., 2003), and accuracy of the predictions that could be made from that data. However, the effect of spatial resolution of the palaeoenvironmental models is dependent on the species studied. For example, a species with a small home range could theoretically find enough 'forest' habitat within a grid cell defined as 'non-forest', whereas for a wider ranging species such as European bison the model predictions may be more informative. In contrast, researchers studying modern European bison are able to characterise the specific percentage of forest cover in the environments occupied by each bison population (Hofman-Kamińska et al., 2018). As a result, such studies can draw much more accurate inferences about habitat preferences and ecological niches.

Palaeoecological modelling is improving rapidly, with more detailed and finer scale predictions recently made available covering the period from 21kya until present (Fordham et al., 2017). There will always be limitations at local scales for palaeoenvironmental modelling, as it is almost impossible to accurately predict small local environmental changes, but additional refinement beyond one degree grid cells every 1,000 years would improve the characterisation of ecological and environmental change through time. The improvements presented in Fordham et al. (2017) and the extension of similarly detailed predictions prior to 21kya would provide a significant improvement to characterisations of ecological niches of species or populations, and the correlation of ecological and genetic changes. In the case of European bison, such improvements would support the characterisation of potential ecological niches, and provide a better understanding of environmental processes potentially driving observed patterns of evolutionary history.

Ancient DNA preservation

Ancient specimens typically contain varying levels of degraded DNA, which impacts sequencing success and efficiency, and downstream analyses. Consequently, even once a specimen has been found and sampled, there is no guarantee that DNA can be successfully extracted and sequenced from it. Issues surrounding the successful extraction and sequencing of DNA from ancient samples are well documented (Hofreiter et al., 2015). In particular, specimens from cold and dry environments (such as permafrost) are typically much better preserved than those from comparatively warmer and more humid environments such as continental Europe (i.e. Western Europe) (Hofreiter et al., 2015). Recently the petrous bone has been identified as a particularly good source of endogenous DNA compared to other bones (up to 183 times more) (Gamba et al., 2014), and consequently is now a common sampling target for ancient specimens. However, subsequent research has also identified substantial variation within individual petrous bones (Pinhasi et al., 2015), and while overall petrous tends to contain higher proportions of endogenous DNA compared to other samples, this DNA also tends to have a higher C to T damage proportion (Hansen et al., 2017). As a result, while petrous bone can be a good target for specimens with particularly poor preservation (such as particularly arid or humid locations), they are not always better suited for genomic analyses (Pinhasi et al., 2015; Hansen et al., 2017).

For a subset of specimens included in this thesis (predominantly those collected after 2014) we were able to obtain samples of petrous bone. DNA extracted from two of those petrous bones was subsequently sequenced at sufficient depth to generate the whole genome data presented in Chapter 6. However, the vast majority of specimens were sampled prior to 2014 (when petrous bone was not a well known source for high-quality ancient DNA) and for a number of specimens sampled post-2014, petrous bone was not available for sampling. Consequently, as per my findings in Chapters 2 and 3, for the

remaining lower-quality samples used for Chapter 5 I applied a variety of methods, including partial UDG treatment of DNA extracts during library construction, hybridisation capture, and HTS technologies (Rohland et al., 2015; Soubrier et al., 2016). These approaches improved the recovery of whole mitochondrial genomes for a large number of specimens. The mitochondrial genome has a greater chance of recovery from ancient specimens because mtDNA has a larger copy number per cell compared to nuclear DNA (Ho and Gilbert, 2010). However, despite the implementation of these improved methods, I was unable to successfully recover whole mitochondrial genome sequences for ~55% of ancient specimens sequenced as opposed to successful recovery of DNA from 100% of historic or modern specimens. However, this rate of success for ancient specimens is already a significant improvement over results with previous molecular techniques and sequencing technologies.

Further developments in laboratory techniques and sequencing technologies have continued to improve the recovery of ancient DNA fragments from poorly preserved specimens. For example, recent developments in library preparation methods (such as single stranded libraries; Gansauge and Meyer (2013); Wales et al. (2015)) and further improvements in sequencing technologies could be applied to the DNA extracts generated in this thesis and may improve the recovery of mitochondrial genomes and nuclear DNA from specimens we were unable to generate DNA data from. The development of a nuclear genome hybridisation capture panel for European bison (either for the whole genome or for a subset of nuclear SNPs) would also improve the recovery of nuclear DNA data and complement the mitochondrial genomes already recovered. Soubrier et al. (2016) incorporated nuclear data from a SNP panel, but these SNPs were ascertained from a domestic cow SNP panel (i.e. they were designed to characterise the diversity in domestic cattle) and as a result poorly characterise diversity between European bison specimens.

The development of an unbiased SNP panel for European bison would substantially improve the delineation of genomic diversity (and evolutionary

relationships) between European bison specimens. For example, a clear outgroup species with a well-annotated reference genome (such as the goat/*Capra hircus*: ARS1) can be used to randomly select potential SNP sites and consequently avoid ascertainment biases. In addition, the recent generation of multiple genomic sequences for European bison (Chapter 6; Węcek et al., 2016; Gautier et al., 2016) provides additional references that can be used to identify variation to target in European bison. Further, an increased number of specimens, in particular late Pleistocene specimens, would complement our current understanding of the evolutionary history of European bison, and the patterns and processes involved prior to the late Pleistocene/early Holocene transition, as well as further delineate specific patterns during the Holocene. For example, additional late Pleistocene wisent specimens (and their associated stable isotope data) would contribute towards our understanding of the relationship between populations possessing CladeX and wisent mitochondrial lineages, and help define their geographical distribution and potential ecological niche preferences.

Limitations of mitochondrial DNA

For Chapter 5 (WMGs), I used mitochondrial genomes to reconstruct the demographic and phylogeographic history of European bison across Eurasia. Mitochondrial DNA (mtDNA) is especially well suited for ancient DNA studies due to its high copy number (relative to nuclear DNA), and due to its high mutation rate and lack of recombination is especially suited for genetic analyses (Ramakrishnan and Hadly, 2009). However, mtDNA is uniparentally inherited and only represents the demographic history of the maternal line, which can introduce biases in demographic reconstructions (Ballard and Whitlock, 2004). For example, a single locus such as the mitochondrial genome represents only a single realisation of the coalescent process, which can result in large confidence intervals on estimates of divergence and demographic inference (Heled and Drummond, 2010). In order to have the full picture of demographic history it is important to also have recombining bi-parentally inherited markers (i.e. autosomal nuclear DNA), which provide many independent realisations of the

coalescent process (Ballard and Whitlock, 2004; Ho and Shapiro, 2011). Estimates of demographic history are often substantially improved with the incorporation of multiple independent nuclear loci (Ballard and Whitlock, 2004; Heled and Drummond, 2010; Ho and Shapiro, 2011). With the generation of not only nuclear loci, but also whole nuclear genomes, multiple additional methods to infer demographic history can be implemented including; PSMC (pair-wise sequentially Markovian coalescent; Li and Durbin (2011)) and MSMC (multiple sequentially Markovian coalescent; Schiffels and Durbin (2014)). However, these methods require sufficient genome depth to accurately call alleles, which typically needs a minimum average depth of 10-15x (Nadachowska-Brzyska et al., 2016; Li and Durbin, 2011).

Nuclear DNA is also necessary to untangle patterns of hybridisation and introgression from incomplete lineage sorting (ILS) (Ballard and Whitlock, 2004). Mitochondrial DNA is especially susceptible to ILS and introgression due to its small effective population size (haploid and uniparentally inherited) (Ballard and Whitlock, 2004; Mastrantonio et al., 2016). There are multiple examples of complete mitochondrial replacement within a population, without any corresponding observed nuclear introgression (Ballard and Whitlock, 2004; Chan and Levin, 2005; Toews and Brelsford, 2012; Pons et al., 2014; Mastrantonio et al., 2016). For example, extensive mitochondrial introgression from American herring gulls into North American black-backed gulls has been characterised, but there was no corresponding evidence for nuclear introgression (Pons et al., 2014). As a result, mitochondrial DNA (or any single locus dataset) is usually insufficient to differentiate between patterns of ILS and introgression or hybridisation (Ballard and Whitlock, 2004; Mastrantonio et al., 2016).

Limitations of low coverage genomes

The average depth of coverage for the genomes I presented in Chapter 6 (2-3x) was not sufficient to apply PSMC or MSMC methods to estimate demographic history, as both require accurately called diploid genotypes (Li and Durbin, 2011; Nadachowska-Brzyska et al., 2016). A number of methods to infer

population history from low coverage genomes have recently been developed with ancient DNA research applications in mind (Pasaniuc et al., 2012; Keller et al., 2012; Schaefer et al., 2017; Kousathanas et al., 2017; Ros-Freixedes et al., 2017; Link et al., 2017). However, several of these methods require a level of knowledge about the demographic processes taking place that is not commonly known for species other than humans (i.e. non model species). For example, to estimate admixture in low coverage genomes using AD-LIBS (Schaefer et al., 2017), input requirements include reference sequences for the admixing populations, estimated admixed population size, generations since admixture, and the proportion of population ancestry for each individual (or an outgroup reference sequence).

Most non-model species (including European bison) currently lack the high coverage genomic data to infer such values. As a consequence, many of these programs cannot yet be implemented without the generation of reference sequences for each species and additional high coverage genomes to estimate population genetic parameters (e.g. mutation or growth rates). Another suite of tools currently under development called ATLAS (Link et al., 2017) includes a novel method for generating estimates of heterozygosity in low coverage genomes (Kousathanas et al., 2017). I applied this method in Chapter 6 to compare estimates of heterozygosity across all available low coverage genome sequences for European bison. ATLAS is particularly unique in that it does not require many of the population genetics parameter inputs of other programs, but instead calculates them on a sliding window basis as the analyses are performed. As such it can be used for species that do not have sufficient high coverage genomic data or reference genomes to infer accurate estimates of parameters required by other programs. However, multiple tools in this program are still under development, and require further validation and testing.

Pseudo-haploidistion, as used in Chapter 6, is a common approach in studies of low coverage genomes to remove the need for deep enough sequencing to call diploid genotypes (typically around 10-15x; Haak et al., 2015; Song et al., 2016; Schaefer et al., 2017; Martiniano et al., 2017). To generate a pseudohaploid

sequence, genome level HTS data is computationally reduced to the equivalent of a haploid genome (i.e. 1x at every site), usually either by random allele selection or by selecting the major allele at each site (Haak et al., 2015; Song et al., 2016; Schaefer et al., 2017; Martiniano et al., 2017). However, a potential issue with pseudo-haploidising genomes is that a large proportion of information is lost, potentially limiting the power of downstream analyses. In addition, there are various ways to call pseudohaploid genomes (randomly, major allele, etc.), and different methods may result in different biases for downstream analyses. For example, in Martiniano et al. (2017), they demonstrated a bias towards increased shared drift with the reference genome in pseudo-haploidised genomes from ancient human specimens when randomly sampling a single allele for each site. However, this may also be due to reference bias during initial alignment, and depends on the relationships of the specimens to the reference genome. Further investigation of the consequences of reference bias for pseudo-haploidisation is necessary to untangle the cause of these biases. Ideally, deeper sequencing of a few key specimens would help define the backbone of the phylogenetic tree, while low coverage genomes from additional specimens would improve the resolution of the tree (Schaefer et al., 2017; Schraiber, 2018).

Conclusion

My thesis contributes towards our understanding of the evolutionary history of European bison, and develops a series of tools and approaches that can be used in subsequent research to further our understanding of the evolutionary history of species in general. Throughout the course of this research, numerous challenges and limitations were identified, as well as additional research questions beyond the scope of this thesis. However, each new research question or current limitation identified provides exciting opportunities for future research.

Additional sampling of European bison specimens focusing on petrous bones will likely increase the successful recovery of nuclear DNA data, and consequently provide more accurate estimates of demographic history. Similarly, expanding the sample coverage to include specimens in key time periods

currently lacking (e.g. between 9-7kya, or wisent specimens during the Late Pleistocene), would improve our understanding of the evolutionary patterns and processes taking place during either of these time periods. Generating additional low coverage genomes and/or high coverage genomes for key time points would have multiple benefits of a range of different analyses. Additional genomes would provide more reference points for comparison, so that a clearer picture of changes in distribution and diversity can be generated. Genome sequences from CladeX specimens in particular would provide a reference point to characterize the genomic variation between wisent and CladeX populations, and delineate appropriate species designations. Genomes with deeper sequencing from key time points could be used to generate reference points to infer demographic history, estimate population genetics parameters to infer patterns of admixture and introgression in low coverage genomes, and identify patterns of selection and functional variation.

I hope that the findings presented here in this thesis contributes towards the 'next-generation' of ancient DNA research by highlighting a series of methods and approaches for future research into the evolutionary history of species. These approaches, and subsequent research that follows in future can be used to compile equivalent datasets of multiple species, which can then subsequently be combined to identify shared patterns of evolution across multiple species, and further our understanding of how species evolve and respond to change.

References

Ballard, J. W. O. and Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, 13(4):729–744.

Behrensmeyer, A. K. (1984). Taphonomy and the fossil record: the complex processes that preserve organic remains in rocks also leave their own traces, adding another dimension of information to fossil samples. *American Scientist*, 72(6):558–566.

Behrensmeyer, A. K., Kidwell, S. M., and Gastaldo, R. A. (2000). Taphonomy and paleobiology. *Paleobiology*, 26(sp4):103–147.

Benton, M. J., Dunhill, A. M., Lloyd, G. T., and Marx, F. G. (2011). Assessing the quality of the fossil record: insights from vertebrates. *Geological Society, London*, Special Publications, 358(1):63–94.

Block, S., Saltré, F., Rodríguez-Rey, M., Fordham, D. A., Unkel, I., and Bradshaw, C. J. A. (2016). Where to Dig for Fossils: Combining Climate-Envelope, Taphonomy and Discovery Models. *PLOS ONE*, 11(3):e0151090.

Chan, K. M. and Levin, S. A. (2005). Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution*, 59(4):720–729.

Fordham, D. A., Saltré, F., Haythorne, S., Wigley, T. M., Otto-Bliesner, B. L., Chan, K. C., and Brook, B. W. (2017). Paleoview: a tool for generating continuous climate projections spanning the last 21 000 years at regional and global scales. *Ecography*, 40(11):1348–1358.

Gamba, C., Jones, E. R., Teasdale, M. D., McLaughlin, R. L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kovári, I., Pap, I., Anders, A., Whittle, A., Dani, J., Raczky, P., Higham, T. F. G., Hofreiter, M., Bradley, D. G., and Pinhasi, R. (2014). Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications*, 5.

Gansauge, M.-T. and Meyer, M. (2013). Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols*, 8(4):737.

Gautier, M., Moazami-Goudarzi, K., Levéziel, H., Parinello, H., Grohs, C., Rialle, S., Kowalczyk, R., and Flori, L. (2016). Deciphering the Wisent Demographic and Adaptive Histories from Individual Whole-Genome Sequences. *Molecular Biology and Evolution*, 33(11):2801–2814.

Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., Mitchell, J. F. B., and Wood, R. A. (2000). The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, 16(2-3):147–168.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., Fu, Q., Mittnik, A., Bánffy, E., Economou, C., Francken, M., Friederich, S., Pena, R. G., Hallgren, F., Khartanovich, V., Khokhlov, A., Kunst, M., Kuznetsov, P., Meller, H., Mochalov, O., Moiseyev, V., Nicklisch, N., Pichler, S. L., Risch, R., Guerra, M. A. R., Roth, C., Szécsényi-Nagy, A., Wahl, J., Meyer, M., Krause, J., Brown, D., Anthony, D., Cooper, A., Alt, K. W., and Reich, D. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, 522(7555):207–211.

Hansen, H. B., Damgaard, P. B., Margaryan, A., Stenderup, J., Lynnerup, N., Willerslev, E., and Allentoft, M. E. (2017). Comparing Ancient DNA

Preservation in Petrous Bone and Tooth Cementum. *PLOS ONE*, 12(1):e0170940.

Heled, J. and Drummond, A. J. (2010). Bayesian Inference of Species Trees from Multilocus Data. Molecular Biology and Evolution, 27(3):570–580. Ho, S. Y. W. and Gilbert, M. T. P. (2010). Ancient mitogenomics. *Mitochondrion*, 10(1):1–11.

Ho, S. Y. W. and Shapiro, B. (2011). Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular Ecology Resources*, 11(3):423–434.

Hofman-Kamińska, E., Bocherens, H., Borowik, T., Drucker, D. G., and Kowalczyk, R. (2018). Stable isotope signatures of large herbivore foraging habitats across Europe. *PLOS ONE*, 13(1):e0190723.

Hofreiter, M., Paijmans, J. L. A., Goodchild, H., Speller, C. F., Barlow, A., Fortes, G. G., Thomas, J. A., Ludwig, A., and Collins, M. J. (2015). The future of ancient DNA: Technical advances and conceptual shifts. *BioEssays*, 37(3):284–293.

Kaplan, J. O., Bigelow, N. H., Prentice, I. C., Harrison, S. P., Bartlein, P. J., Christensen, T. R., Cramer, W., Matveyeva, N. V., McGuire, A. D., Murray, D. F., Razzhivin, V. Y., Smith, B., Walker, D. A., Anderson, P. M., Andreev, A. A., Brubaker, L. B., Edwards, M. E., and Lozhkin, A. V. (2003). Climate change and Arctic ecosystems: 2. Modeling, paleodata-model comparisons, and future projections. *Journal of Geophysical Research: Atmospheres*, 108(D19).

Keller, A., Graefen, A., Ball, M., Matzas, M., Boisguerin, V., Maixner, F., Leidinger, P., Backes, C., Khairat, R., Forster, M., et al. (2012). New insights into the tyrolean iceman's origin and phenotype as inferred by whole-genome sequencing. *Nature Communications*, 3:698.

Kidwell, S. M. and Flessa, K. W. (1995). The quality of the fossil record: populations, species, and communities. *Annual Review of Ecology and Systematics*, 26(1):269–299.

Kousathanas, A., Leuenberger, C., Link, V., Sell, C., Burger, J., and Wegmann, D. (2017). Inferring Heterozygosity from Ancient and Low Coverage Genomes. *Genetics*, 205(1):317–332.

Leonard, J.A. (2008). Ancient DNA applications for wildlife conservation. *Molecular Ecology*, 17(19):4186–4196.

Li, H. and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357):493.

Link, V., Kousathanas, A., Veeramah, K., Sell, C., Scheu, A., and Wegmann, D. (2017). ATLAS: Analysis Tools for Low-depth and Ancient Samples. *bioRxiv*, page 105346.

Martiniano, R., Cassidy, L. M., Ó'Maoldúin, R., McLaughlin, R., Silva, N. M., Manco, L., Fidalgo, D., Pereira, T., Coelho, M. J., Serra, M., Burger, J., Parreira, R., Moran, E., Valera, A. C., Porfirio, E., Boaventura, R., Silva, A. M., and Bradley, D. G. (2017). The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLOS Genetics*, 13(7):e1006852.

Mastrantonio, V., Porretta, D., Urbanelli, S., Crasta, G., and Nascetti, G. (2016). Dynamics of mtDNA introgression during species range expansion: insights from an experimental longitudinal study. *Scientific Reports*, 6:30355.

Nadachowska-Brzyska, K., Burri, R., Smeds, L., and Ellegren, H. (2016). Psmc analysis of effective population sizes in molecular ecology and its application to black-and-white ficedula flycatchers. *Molecular Ecology*, 25(5):1058–1072.

Orlando, L. and Cooper, A. (2014). Using Ancient DNA to Understand Evolutionary and Ecological Processes. *Annual Review of Ecology*, Evolution, and Systematics, 45(1):573–598.

Pasaniuc, B., Rohland, N., McLaren, P. J., Garimella, K., Zaitlen, N., Li, H., Gupta, N., Neale, B. M., Daly, M. J., Sklar, P., et al. (2012). Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nature Genetics*, 44(6):631.

Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpaslan-Roodenberg, S., Gerritsen, F., Moiseyev, V., Gromov, A., Raczky, P., et al. (2015). Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLOS ONE*, 10(6):e0129102.

Pons, J., Sonsthagen, S., Dove, C., and Crochet, P. (2014). Extensive mitochondrial introgression in North American great black-backed gulls (*Larus marinus*) from the American herring gull (*Larus smithsonianus*) with little nuclear DNA impact. *Heredity*, 112(3):226.

Pope, V. D., Gallani, M. L., Rowntree, P. R., and Stratton, R. A. (2000). The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3. *Climate Dynamics*, 16(2-3):123–146.

Ramakrishnan, U. and Hadly, E. A. (2009). Using phylochronology to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies. *Molecular Ecology*, 18(7):1310–1330.

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Phil. Trans. R. Soc. B*, 370(1660):20130624.

Ros-Freixedes, R., Gonen, S., Gorjanc, G., and Hickey, J. M. (2017). A method for allocating low-coverage sequencing resources by targeting haplotypes rather than individuals. *Genetics Selection Evolution*, 49(1):78.

Schaefer, N. K., Shapiro, B., and Green, R. E. (2017). AD-LIBS: inferring ancestry across hybrid genomes using low-coverage sequence data. *BMC Bioinformatics*, 18:203.

Schiffels, S. and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, 46(8):919.

Schraiber, J. G. (2018). Assessing the Relationship of Ancient and Modern Populations. *Genetics*, 208(1):383–398.

Song, J.-J., Wang, W.-Z., Otecko, N. O., Peng, M.-S., and Zhang, Y.-P. (2016). Reconciling the conflicts between mitochondrial DNA haplogroup trees of Canis lupus. *Forensic Science International: Genetics*, 23:83–85.

Soubrier, J., Gower, G., Chen, K., Richards, S. M., Llamas, B., Mitchell, K. J., Ho, S. Y. W., Kosintsev, P., Lee, M. S. Y., Baryshnikov, G., Bollongino, R., Bover, P., Burger, J., Chivall, D., Crégut-Bonnoure, E., Decker, J. E., Doronichev, V. B., Douka, K., Fordham, D. A., Fontana, F., Fritz, C., Glimmerveen, J., Golovanova, L. V., Groves, C., Guerreschi, A., Haak, W., Higham, T., Hofman-Kamińska, E., Immel, A., Julien, M.-A., Krause, J., Krotova, O., Langbein, F., Larson, G., Rohrlach, A., Scheu, A., Schnabel, R. D., Taylor, J. F., Tokarska, M., Tosello, G., van der Plicht, J., van Loenen, A., Vigne, J.-D., Wooley, O., Orlando, L., Kowalczyk, R., Shapiro, B., and Cooper, A. (2016). Early cave art and ancient DNA record the origin of European bison. *Nature Communications*, 7:13158.

Toews, D. P. and Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, 21(16):3907–3930.

Wales, N., Carøe, C., Sandoval-Velasco, M., Gamba, C., Barnett, R., Samaniego, J. A., Ramos Madrigal, J., Orlando, L., and Gilbert, M. T. P. (2015). New insights on single-stranded versus double-stranded DNA library preparation for ancient DNA. *Biotechniques*, 59(6):368–371.

Węcek, K., Hartmann, S., Paijmans, J. L., Taron, U., Xenikoudakis, G., Cahill, J. A., Heintzman, P. D., Shapiro, B., Baryshnikov, G., Bunevich, A. N., et al. (2016). Complex admixture preceded and followed the extinction of wisent in the wild. *Molecular Biology and Evolution*, 34(3):598–612.

Appendix 1

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

This appendix contains the publication of a preliminary study of the evolutionary history of European bison (*Bison bonasus*). I contributed to this work by performing laboratory genetic analyses of mitochondrial and nuclear data that contributed towards the body of genetic data analysed in this paper, performed the data processing of the aforementioned genetic data, and provided edits and comments on the manuscript. The samples analysed in this study are included in Chapter 5 in the multiple sequence alignment of all available mitochondrial genomes of European bison. This article has been published in Nature Communications.

Soubrier J*, Gower G*, Chen K, Richards SM, Llamas B, Mitchell KJ, Ho SYW, Kosintsev P, Lee MSY, Baryshnikov G, Bollongino R, Bover P, Burger J, Chivall D, Crégut-Bonnoure E, Decker JE, Doronichev VB, Douka K, Fordham DA, Fontana F, Fritz C, Glimmerveen J, Golovanova LV, Groves C, Guerreschi A, Haak W, Higham T, Hofman-Kamińska E, Immel A, Julien MA, Krause J, Krotova O, Langbein F, Larson G, Rohrlach A, Scheu A, Schnabel RD, Taylor JF, Tokarska M, Tosello G, van der Plicht J, **van Loenen A**, Vigne JD, Wooley O, Orlando L, Kowalczyk R, Shapiro B, Cooper A (2016). Early cave art and ancient DNA record the origin of European bison. Nature Communications, 7, 13158-1-13158-7.

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 (~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, ~22–18 kya), while the latter dominates Magdalenian art (~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 ~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).

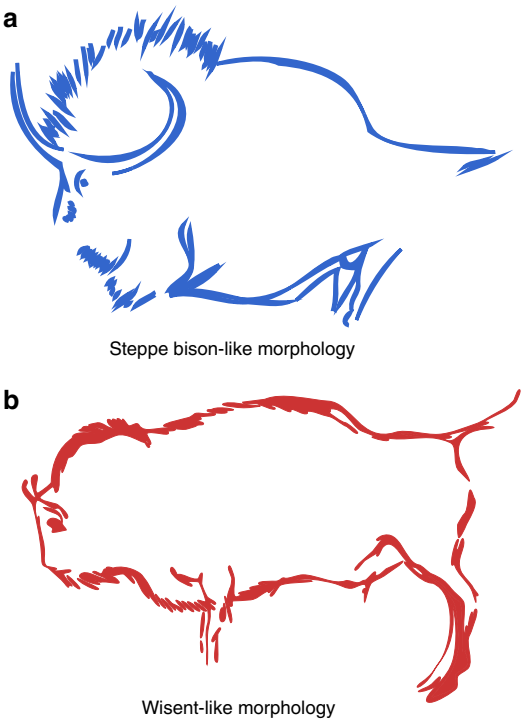


Figure 1 | Cave painting example of steppe bison-like and wisent-like morphs. (**a**) Reproduction from Lascaux cave (France), from the Solutrean or early Magdalenian period (~20,000 kya—picture adapted from ref. 53). (**b**) Reproduction from the Pergouset cave (France), from the Magdalenian period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

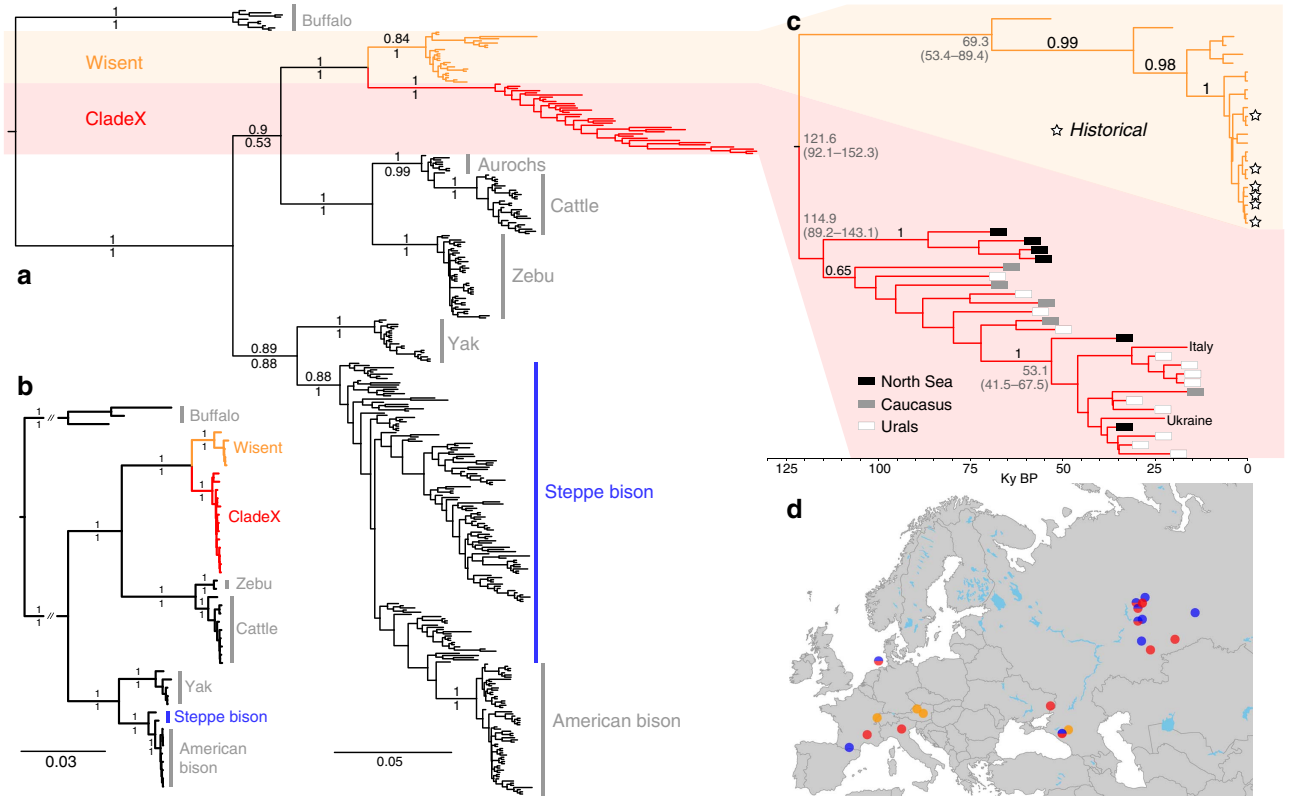


Figure 2 | Identification of CladeX. (**a**) Phylogenetic tree inferred from bovine mitochondrial control region sequences, showing the new clade of bison individuals. The positions of the newly sequenced individuals are marked in red for CladeX. (**b**) Bovine phylogeny estimated from whole-mitochondrial 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 ($\delta^{13}C/\delta^{15}N$; 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

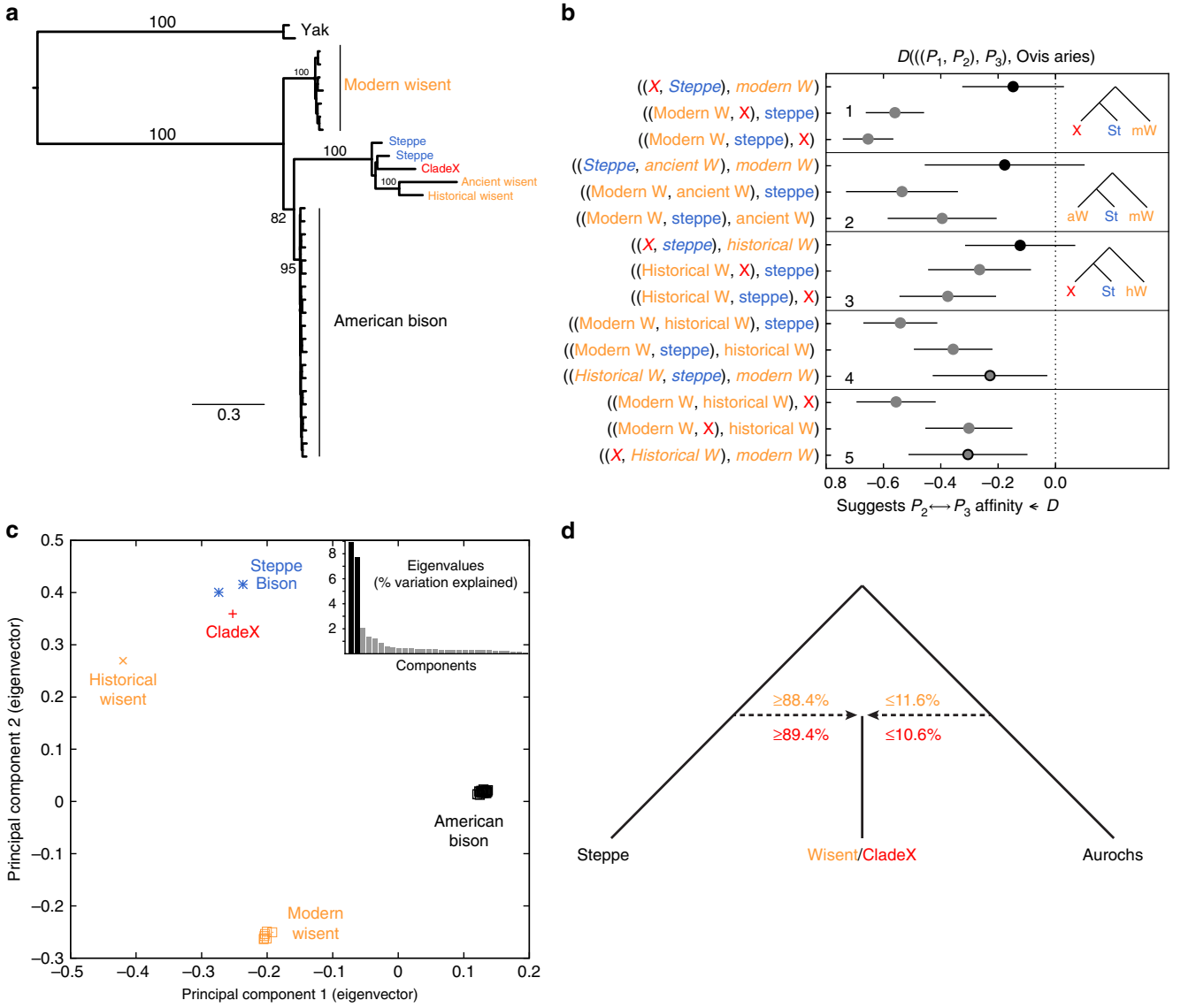


Figure 3 | Genome-wide data comparison of bison. (**a**) Maximum-likelihood phylogeny of modern and ancient bison from ~10,000 genome-wide nuclear sites, showing the close relationship between CladeX and steppe bison. However, a bifurcating phylogeny is not capable of displaying the complex relationships 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).

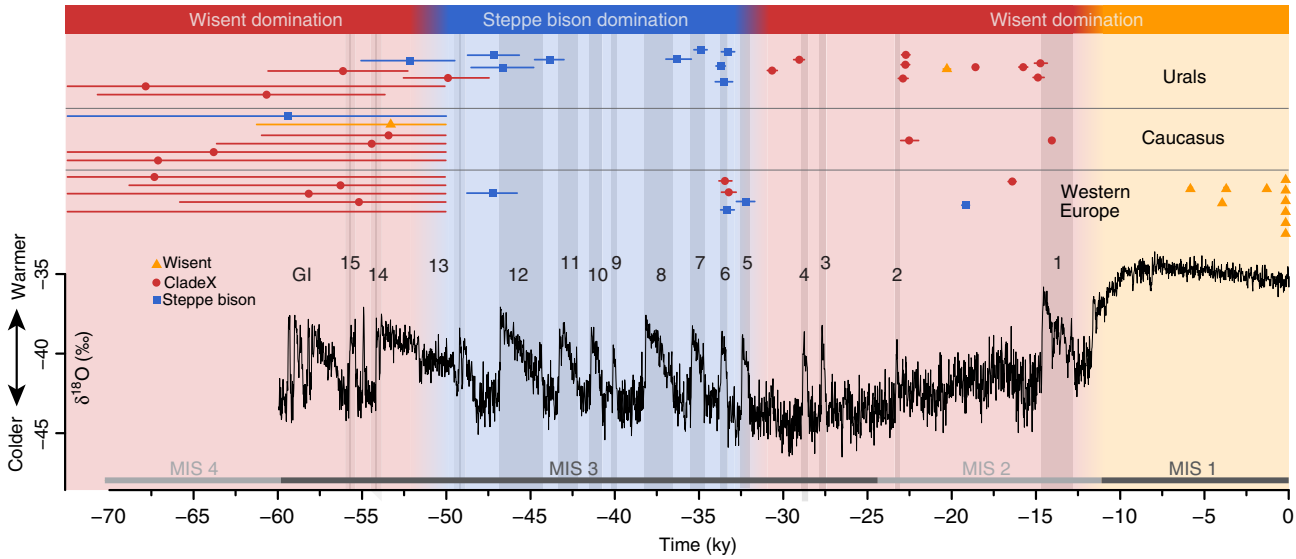


Figure 4 | Temporal and geographical distribution of bison in Europe. Individual calibrated AMS dates from the present study and published data are plotted on top of the NGRIP $\delta^{18}O$ record[55]. Age ranges for infinite AMS dates are from molecular clock estimates (Fig. 2c). Greenland interstadials (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, $n^{X_1} \sim \text{Mult}(N, p^{T,X_1})$, and for tree topology X_2, $n^{X_2} \sim \text{Mult}(N, p^{T,X_2})$. 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

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

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 $1° \times 1°$ 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).

Appendix 2

Did the historical range of the European bison (Bison bonasus L.) extend further south? - a new finding from the Yenikapı Metro and Marmaray excavation, Turkey.

This appendix contains the publication of a mitochondrial genome from Marmaray (Turkey) that expands the known southern extent of European bison distribution. I contributed to this work by performing the laboratory genetic analyses of mitochondrial data for the specimen, sequencing data processing for the specimen, and provided edits and comments on the manuscript. The specimen analysed in this study is included in Chapter 5 in the multiple sequence alignment of all available mitochondrial genomes of European bison. This article has been published in Mammal Research.

Onar, V., Soubrier, J., Toker, N., **van Loenen, A.**, Llamas, B., Siddiq, A. Pasicka, E. and Tokarska, M. (2017). Did the historical range of the European bison (Bison bonasus L.) extend further south? - a new finding from the Yenikapı Metro and Marmaray excavation, Turkey. Mammal Research, 62, 1, 103-10

Mamm Res
DOI 10.1007/s13364-016-0299-4

ORIGINAL PAPER

Did the historical range of the European bison (*Bison bonasus* L.) extend further south?—a new finding from the Yenikapı Metro and Marmaray excavation, Turkey

Vedat Onar[1] · Julien Soubrier[2] · Nezir Yaşar Toker[3] · Ayla van Loenen[2] · Bastien Llamas[2] · Abu Bakar Siddiq[4] · Edyta Pasicka[5] · Małgorzata Tokarska[6]

Received: 23 May 2016 / Accepted: 7 November 2016

Abstract The origin of the European bison (*Bison bonasus*, Linnaeus, 1758) has been widely discussed and investigated in recent years. The species had a wide historic geographic distribution throughout the European continent during the middle and late Holocene, ranging from France in the west to the Caucasus in the east. However, archaeological evidence is needed to resolve the southern extent of the European bison distribution. We discovered one bison skull fragment during archaeological excavations in 2008 in the area of Yenikapı Metro and Marmaray (Turkey). Radiocarbon dating indicated the skull was deposited during the Byzantine period (seventh to eighth century AD). Mitochondrial genome analyses provided clear evidence that the skull was from a European bison. This is the first unambiguous evidence of the presence of this species in southeastern Europe during Byzantine times, which validates the historical written records of a potentially wider range of the European bison in historical times.

Keywords European bison · Skull morphology · mtDNA · Radiocarbon dating · Zooarchaeology · Turkey

Communicated by: Allan McDevitt

✉ Małgorzata Tokarska
tokarska@ibs.bialowieza.pl

[1] Osteoarchaeology Practice and Research Center and Faculty of Veterinary Medicine, Istanbul University, Avcılar, Istanbul, Turkey

[2] Australian Centre for Ancient DNA, School of Biological Sciences, The University of Adelaide, Adelaide, SA 5005, Australia

[3] Department of Biochemistry, Faculty of Veterinary Medicine, Istanbul University, Avcılar, Istanbul, Turkey

[4] Department of Prehistory, Archaeology Division, Faculty of Letters, Istanbul University, Beyazıt, Istanbul, Turkey

[5] Department of Biostructure and Animal Physiology, Faculty of Veterinary Medicine, Wrocław University of Environmental and Life Sciences, Wrocław, Poland

[6] Department of Genetics and Evolution, Mammal Research Institute Polish Academy of Sciences, Waszkiewicza 1, 17-230 Bialowieza, Poland

Introduction

During the early Holocene, European bison (*Bison bonasus*) were widespread throughout the European continent, but did not occupy Scandinavia. *B. bonasus* appeared in Denmark and Sweden only at the end of the last glaciation period (c. 10,000 BP) (Pucek 2004). During the middle and late Holocene (last 6000 years), the population of the European bison was widely distributed from France in the west to Russia's Volga River and the Caucasus in the east (Benecke 2005; Pucek 2004). It was proposed that bison were also present in other regions (modern Armenia, Azerbaijan, Georgia, Northern Iran and Asiatic parts of Russia) although evidence is needed to substantiate this hypothesis (Pucek 2004; Flint et al. 2002). About 100 years ago, *B. bonasus* reached an extinction threshold due to habitat degradation and fragmentation linked to agricultural and logging activities, as well as unlimited hunting and heavy poaching (Pucek 2004). Fortunately, extinction was prevented through several conservation initiatives and controlled breeding in recent years, which resulted in a population increase (Pucek 1986, 2004).

Information on the presence and distribution of the *B. bonasus* in historical times has been mainly based on ancient written sources (Benecke 2005). However, names used to describe European bison have varied wildly, which creates some confusion and weakens the reliability of historical written sources (Benecke 2005). For example, Aristotle used the word "bonasus" to describe European bison in his reports on the

Published online: 23 November 2016

history of animals (Aristotle and Balme 1991). In later years, between 1450 and 1850, "auerochs" was used as the official name of wisent/bison/żubr. There were attempts to eliminate this confusion in subsequent accounts by assigning the names "aerox" for aurochs and "bisont" for wisent (Ahrens 1921). Yet, some twentieth century and medieval descriptions use the name "aurochs" for European bison (Avebury 1913; Von Lengerken 1953).

Natural habitats of *B. bonasus* include large deciduous forests, bushes, shrubs, meadows and marsh areas (Benecke 2005; Pucek 2004). Consequently, it is expected that European bison could have occupied an area from the northern European plains to interior Asia based on the suitability of niches (Filean 2006). In fact, archaeological evidence indicates the widespread presence of European bison in Europe at least during the sixteenth century (Aleksandrowicz 1999). However, interpretation of the archaeological record is problematic: domestic and wild cattle remains could easily be incorrectly identified as European bison because of their similar size and morphological features (Benecke 2005). In particular, domestic male cattle and wild female cattle were possibly described as bison in some written sources based on their basic features (Filean 2006). For example, a bison bone was allegedly reported in faunal remains from Sos Höyük (on the border of Turkey and Iran), although the finding was classified as "doubtful" (Howell-Meurs 2001). Similarly, two bones were described as bison remains in the faunal assemblage of Pınarbaşı, Central Anatolia, but the accuracy of the identification of these remains is also doubtful because of their possibility of being mixed with other bovid species (Carruthers 2005).

A bison skull fragment was found in Yenikapı Metro and Marmaray (Turkey) excavation in 2008. This was one of the most important findings from the Byzantine period found at this site. The 58-km^2 excavation area is located in the Yenikapı area of Istanbul, about 1.5 km from the Marmara Sea (Başaran et al. 2007) (Fig. 1).

Relatively large number of animal species have been found during subsequent excavations (Onar et al. 2013a, b). Amongst these, one identified European bison skull fragment was excavated during the 2008 season. Geographically, the

Fig. 1 Yenikapı Metro and Marmaray excavation area

Fig. 2 The Yenikapı Metro and Marmaray European bison skull fragment. **a** Dorsal view. **b** Basal (ventral) view. **c** Oral (rostral) view. **d** Aboral (nuchal) view

nearest findings of other bison remains have been reported from the southeastern Europe (Bulgaria and Romania) (Benecke 2005). This paper reports a unique finding of the bison remains from the Istanbul site is of utmost importance in relation to the addition of new geographical ancient distribution range.

Material and methods

A single skull fragment identified as European bison, from level −4.00 of trench 1Hc2 from the 2008 excavation season of Istanbul Yenikapi Metro and Marmaray excavation area, was used in this study (Fig. 2). The skull fragment was sent

Table 1 Primer sequences for DNA typing

Amplicon	Primer	Primer sequence (5′-3′)	Length with primers
1	Bovid 16351F BISCR-16457R	CAACCCCCAAAGCTGAAG TGGTTRGGGTACAAAGTCTGTG	133 bp
2	BISCR-16420F BISCR-16642R	CCATAAATGCAAAGAGCCTCAYCAG TGCATGGGGCATATAATTTAATGTA	240 bp
3	BISCR-16507F BISCR-16755R	AATGCATTACCCAAACRGGG ATTAAGCTCGTGATCTARTGG	248 bp
4	BisonCR-16633F[a] BisonCR_16810R[a]	GCCCCATGCATATAAGCAAG GCCTAGCGGGTTGCTGGTTTCACGC	170 bp
5	BisonCR-16765F[a] BISCR-16998R	GAGCTTAAYTACCATGCCG CGAGATGTCTTATTTAAGAGGAAAGAATGG	180 bp
6	BISCR-16960F BisonCR-80R[a]	CATCTGGTTCTTTCTTCAGGGCC CAAGCATCCCCCAAAATAAA	153 bp

[a] Primer sequences from Shapiro et al. (2004)

Table 2 Radiocarbon dating of the Marmaray bison skull fragment

Site	Species	$\delta^{13}C$	OxA-32358 uncalBP[a]	Confidence	calAD	Century
Yenikapı Metro and Marmaray	European bison (*Bison bonasus*)	−20.98	1283 ± 26	95.4%	669–770	7-8th

[a] Before present (BP)–AD 1950

to the Oxford Radiocarbon Accelerator Unit, Oxford University, for radiocarbon dating.

Ancient DNA work Samples were sent to the Australian Centre for Ancient DNA (University of Adelaide, Australia) where sample preparation and DNA extraction were performed in an off-campus, purpose-built, physically isolated, ancient DNA facility equipped with positive air pressure. The laboratory follows strict clean-room conditions to minimise the impact of DNA contamination. All rooms and workbenches are routinely decontaminated with a 3% sodium hypochlorite solution (bleach) and UV irradiated overnight. Consumable packages are discarded where possible or extensively wiped with 3% bleach and UV irradiated before entering the facility. Workers use sterile and disposable hooded coveralls, face masks, dedicated shoes and two pairs of gloves. A second laboratory on campus accommodates all post-DNA-amplification work. A strict one-way movement of materials and personnel is implemented, from pre- to post-amplification laboratory. Methodological precautions include total removal of sample surface, repeated irradiation with UV light, mock extractions, DNA amplification reaction blanks and replication of DNA amplification.

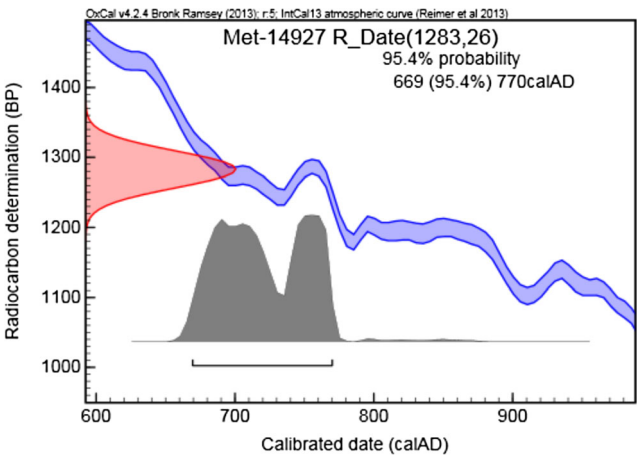


Fig. 3 Calibrated date range (calAD) of the European bison from Yenikapı Metro and Marmaray. This plot shows how the radiocarbon measurement 1283 ± 26 BP was calibrated. The *left-hand axis* shows radiocarbon concentration expressed in years "before present, 1950", and the *bottom axis* shows calendar years (derived from the IntCal13 calibration curve). The *blue curve* shows the IntCal13 calibration curve derived from direct ^{14}C measurements (±one standard deviation), and the *red curve* on the *left* indicates the radiocarbon concentration in the sample. The *grey histogram* shows possible ages for the sample (the higher the histogram, the more likely that age is)

Sample preparation and DNA extraction Potential DNA contamination on the surface of the bison skull fragment was reduced by UV irradiation (260 nm) of all surfaces for 15 min, thorough wiping with 3% bleach and removal of ~1 mm of the sample surface by abrasion using a dremel tool and a disposable carborundum disk. The sample (200 mg) was powdered using a Mikrodismembrator (Sartorius) and immediately decalcified and lysed overnight in 4 mL 0.5 M EDTA pH 8.0, 200 μL SDS and 40 μL proteinase K at 55 °C under constant rotation. After lysis, DNA was purified using an in-solution silica-based method previously described (Brotherton et al. 2013) and resuspended in 200 μL of 10 mM Tris–1 mM EDTA buffer.

Sequencing of the mitochondrial control region Polymerase chain reaction (PCR) was prepared in the ancient DNA laboratory. To test if the sample contained bison DNA, the mitochondrial DNA D-loop was amplified in six overlapping fragments. PCR was performed using 0.4 mM of each primer (see Table 1 for primer sequences), 2.5 mM $MgSO_4$, 0.25 mM dNTPs, 1 mg/mL rabbit serum albumin (Sigma, fraction V) and 1.25 U Platinum Taq Hi-Fi polymerase (Invitrogen). Thermal cycling parameters were 94 °C for 2 min, 50 cycles of 94 °C for 20 s, 58 °C for 30 s and 68 °C for 30 s, then 68 °C for 10 min. Resulting PCR products were visualised on a pre-stained 3.5% agarose gel under UV light. For successful PCR, amplicons were purified using Agencourt Ampure (Beckman Coulter) according to the manufacturer's protocol. Sanger sequencing was performed at the Australian Genome Research Facility Ltd (AGRF). For successfully sequenced amplicons, forward and reverse sequences were aligned to the bison control region, cleaned from primer sequences, and a consensus was made using the software Geneious v7 (Kearse et al. 2012).

Sequencing of the whole mitochondrial genome Double-stranded Illumina libraries were built from 20 μL of DNA extract using partial uracil-DNA glycosylase (UDG) treatment (Rohland et al. 2015) and truncated Illumina adapters with dual 7-mer internal barcodes, following the protocol from (Soubrier et al. 2016). The complete mitochondrial genome of the Marmaray sample was captured using hybridisation with commercial RNA baits (MYcroarray) (Soubrier et al. 2016). The enriched library was sequenced in paired-end reactions on an Illumina MiSeq machine. Sequencing reads were processed using the pipeline Paleomix v1.0.1 (Schubert

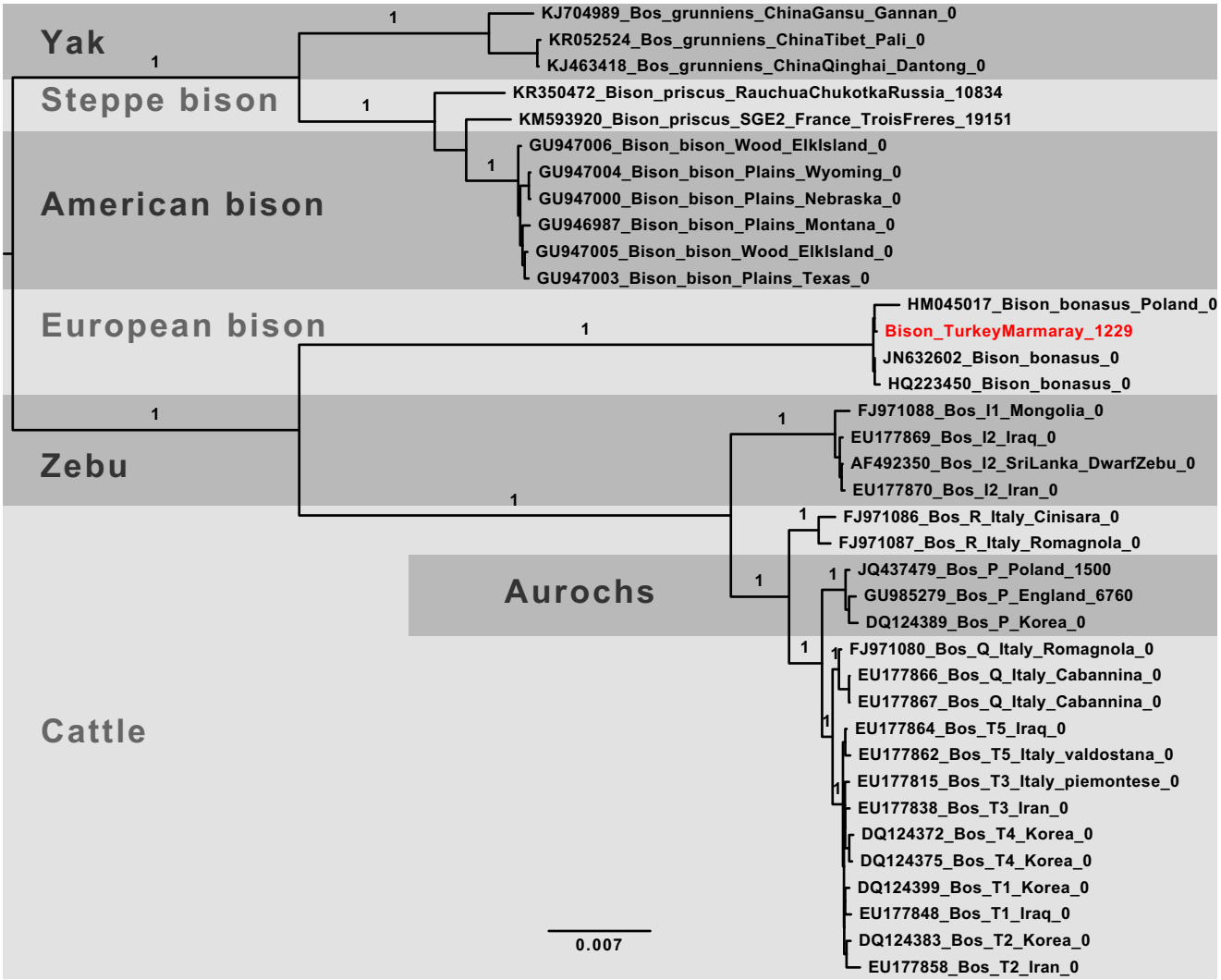


Fig. 4 Maximum-likelihood phylogenetic tree of whole mitochondrial genomes showing the position of the Yenikapı Metro and Marmaray excavation sample within the European bison clade. *Numbers* above branches represent approximate likelihood-ratio test support values, and the *scale bar* represents nucleotide substitutions per site

et al. 2014). AdapterRemoval v2 (Lindgreen 2012) was used to trim adapter sequences, merge the paired reads and discard reads shorter than 25 bp (Li and Durbin 2009). BWA v0.6.2 was then used to map the processed reads to the reference mitochondrial genome of the European bison (GenBank ref. NC_014044). Minimum mapping quality was set at 25, seeding was disabled, the maximum number or fraction of gap opens was set to 2 and the edit distance was relaxed. MapDamage v2 (Jónsson et al. 2013) was used to verify the presence of expected contextual mapping and damage patterns considering the partial UDG treatment of the library (i.e. only the first and last base of sequencing reads should show higher levels of C-to-T and G-to-A transitions, respectively). Finally, the Marmaray consensus sequence was called using samtools/bcftools (Li et al. 2009) with minimum base quality at 30 and minimum depth of coverage at 3, and the Paleomix script "vcf_to_fasta".

Phylogenetic analysis The Marmaray consensus mitochondrial genome sequence was aligned to 36 published bovid sequences by hand using the software SeaView v4.3.5 (Gouy et al. 2010). Published sequences represented the following species: European bison (*Bison bonasus*), American bison (*Bison bison*), steppe bison (*Bison priscus*), yak (*Bos grunniens*), zebu (*Bos indicus*), cattle (*Bos taurus*) and aurochs (*Bos primigenius*).

To place the studied individual within the modern bovid assemblage, a maximum-likelihood phylogenetic tree was constructed using the program PhyML v3 (Guindon et al. 2010), HKY+G6 as the model of nucleotide substitution selected by comparison of Bayesian information criterion (BIC) scores in ModelGenerator v0.85 (Keane et al. 2006), NNI and SPR rearrangements to search for the tree topology and approximate likelihood-ratio tests to establish statistical support for the internal branches. The resulting tree, comparing the

Marmaray skull DNA with DNA of other Bovini tribe representatives, is shown in Fig. 4.

Results

Radiocarbon dating The skull fragment belonged to the seventh–eighth century AD as determined by radiocarbon dating (OxA-32358—Table 2). Calibration of the radiocarbon date was performed using OxCal v4.2.4 and the IntCal13 curve (Reimer et al. 2013), and the calibrated age range curve is given in Fig. 3.

Ancient DNA analysis No amplicons were detected for the PCR amplification negative controls. Fragments of expected sizes were detected for amplicons 1, 4, 5 and 6, a larger fragment than expected was detected for amplicon 2, and amplicon 3 could not be generated. Sanger sequencing results showed that fragments 1, 4, 5 and 6 corresponded to the bison control region, whereas fragment 2 was identified as cattle, therefore considered as a potential contaminant. Based on these results, hybridisation capture and sequencing of the whole mitochondrial genome were attempted. A total of 25,486 sequencing reads mapped uniquely to the European bison mitochondrial reference sequence (NC_014044), covering 100% of the reference at an average depth of 123×. Damage pattern analysis showed 13.8% of C-to-T at 5′ ends and 7.1% of G-to-A at 3′ ends, with less than 0.5% substitution level at all other positions, as expected from an ancient sample and the partial UDG treatment of the DNA library. The phylogenetic analysis showed that the Marmaray specimen is nested within the *B. bonasus* clade, with very high confidence (Fig. 4).

Discussion and conclusions

The *B. bonasus* skull fragment found in Istanbul, at the meeting point of Asia and Europe, is the only remain of this species ever found in this region. Previous reports of European bison bones in Anatolian sites such as Sos Höyük (Howell-Meurs 2001) and Pınarbaşı (Carruthers 2005) are debatable because of the highly likely possibility of confounding some morphological features with other bovid species. We propose that DNA analysis of these other Anatolian remains is required to identify them accurately, such as in the present study. Thus, the nearest unequivocal presence of *B. bonasus* archaeological remains is in Bulgaria and Romania, in southeastern Europe (Benecke 2005).

In Byzantine times, the Theodosius Harbour, being the second largest port of the Mediterranean Sea, was used as the most important centre for grain shipments from Egypt and remained a large harbour until the twelfth century (Kocabaş 2008, 2015). The discovery of seventh–eighth century European bison remains in this harbour area suggests that commercial activities related to this species may have spread through Bulgaria and Romania to interior areas of Thrace. Bison may have been traded from the Balkans/Carpates as food supply for people living in the Constantinople area, although there was no sign of cut marks on the Marmaray bison skull fragment.

This is the 57th animal species found in the Yenikapı Harbour area, which exhibits an extraordinarily diverse faunal assemblage from Byzantine times (Onar et al. 2013a, b). More importantly, the presence of a *B. bonasus* skull fragment in the Yenikapı Metro and Marmaray area strongly suggests that the historic distribution of European bison may have extended in the south as far as Asia Minor. This finding supports historical reports of the presence of European bison in the Constantinople area during the Byzantine period (Ahrens 1921).

Acknowledgments The authors of this study offer their grateful thanks to Prof. Alan Cooper and Dr. Pere Bover Arbos for their substantial contributions of expertise and knowledge to this study as well as Amy Eycott.

This work was supported by the Scientific Research Projects Coordination Unit of Istanbul University (Project number: THZ-2016-20107) and the Australian Research Council.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Data availability The Marmaray European bison whole mitochondrial sequence is available at GenBank under the accession number KX773459.

References

Ahrens TG (1921) The present status of the European bison or wisent. J Mammal 2:58–62

Aleksandrowicz NP (1999) Zubr (*Bison bonasus bonasus* Linnaeus 1758) v okhotnichei dobyche naseleniya srednіovekovych beloruskich gorodov (Bones remains of European bison in the hunt trophies of medieval Belarussian towns). In: Proceedings of conference: Nauchno-prakticheskaya konferentsiya posvyashchenna 60-letiyu so dnia obrazovanya Gosudarstvennogo zapovednika "Belovezhskaya Pushcha", Minsk, 22–24 Dec 1999 [In Russian]

Aristotle, Balme DM (1991) History of animals. Cambridge, MA: Harvard UP. Print

Avebury L (1913) Prehistoric times. As illustrated by ancient remains and the manners and customs of modern savages. Williams and Norgate, London

Başaran S, Kocabas U, Kocabas I, Yılmaz R (2007) Istanbul University Yenikapı shipwrecks project: Documentation, lifting, conservation and reconstruction. In: Istanbul: 8000 years brought to daylight,

Marmaray, Metro, Sultanahmet excavations. Istanbul: Vehbi Koç Foundation. pp.190–195

Benecke N (2005) The Holocene distribution of European bison—the archaeozoological record. MUNIBE (Anthropologia-Arkeologia) 57: 421–428. Brehm-Bücherei 105. Leipzig

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

Carruthers DB (2005) Hunting and herding in central Anatolian prehistory: the sites at Pınarbaşı. Archaezoology of the near east VI. Proceedings of the sixth international symposium on the archaeozoology of southwestern Asia and adjacent areas. ARC-Publicaties 123 Groningen, The Netherlands, pp.85–95

Filean EP (2006) Domestic cattle and economic change in the Roman period lower Rhineland: the Civitas Batavorum. Thesis The University of Iowa, Iowa

Flint VE, Belousova IP, Pererva VI, Kazmin VD, Kiseleva EG, Kudryavtsev IV, Pierozikov EN, Sipko TP (2002) Strategy for conservation the European bison in the Russian Federation. Russian Academy of Sciences, Moscow, pp. 1–45

Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27:221–224

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321

Howell-Meurs S (2001) Archaeozoological evidence for pastoral systems and herd mobility: the remains from Sos Höyük and Büyüktepe Höyük. Int J Osteoarchaeol 11:321–328

Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. BİOİNFORMATİCS 29:1682–1684. doi:10.1093/bioinformatics/btt193

Keane TM, Creevey CJ, Pentony MM, Naughton TJ, Mclnerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol Biol 6:29. doi:10.1186/1471-2148-6-29

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C et al (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649

Kocabaş U (2008) Life at the Theodosian Harbour, wrecks and a rapid siting. In: Kocabaş U (ed) The "old ships" of the "new gate" 1. Ege Publishing, İstanbul, pp. 23–36

Kocabaş U (2015) Geçmişe açılan kapı Yenikapı Batıkları. Ege Yayınları, İstanbul

Krasińska M, Krasiński Z A (2013) European bison. The nature monograph. Second Ed., Springer-Verlag Berlin Heidelberg

Lengerken HV (1953) Der Ur und seine Beziehungen zum Menschen. Neue Brehm-Bücherei/Leipzig 105:1–80

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

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) 1000 genome project data proc the sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079

Lindgreen S (2012) AdapterRemoval: easy cleaning of next generation sequencing reads. BMC RES NOTES 5:337. doi:10.1186/1756-0500-5-337

Onar V, Alpak H, Pazvant G, Armutak A, Gezer İnce N, Kızıltan Z (2013a) A bridge from Byzantium to modern day Istanbul: an overview of animal skeleton remains found during Metro and Marmaray excavations. JOURNAL OF THE FACULTY OF VETERINARY MEDICINE, Istanbul University 39:1–8

Onar V, Pazvan G, Alpak H, Gezer-İnce N, Armutak A, Kızıltan ZS (2013b) Animal skeletal remains of the Theodosius harbor: general overview. Turk J Vet Anim Sci 37:81–85

Pucek Z (1986) Bison bonasus (Linnaeus 1758)—Wisent. In. Niethammer J, Krapp F (eds) Handbuch der Saugetiere Europas. Band 2/II (Paarhufer), AULA-Verlag GmnH, Wiesbaden: 278–315

Pucek Z (ed); Pucek Z, Belousova IP, Krasińska M, Krasiński ZA, Olech W (comps.) (2004) European bison. Status survey and conservation action plan. IUCN/SSB Bison Specialist Group IUCN, Gland, Cambridge

Reimer PJ, Bard E, Bayliss A, Warren Beck J, Blackwell PG, Bronk Ramsey C, Buck et al (2013) IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. RADİOCARBON 55(4):1869–1887. doi:10.2458/azu_js_rc.55.16947

Rohland N, Harney E, Mallick S, Nordenfelt S, Reich D (2015) Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. PHILOS T ROY SOC B 22:939–949

Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, Pybus OG, Thomas M, Gilbert P, Barnes I, Binladen J, Willerslev E, Hansen AJ, Baryshnikov GF, Burns JA, Davydov S, Driver JC, Froese DG, Harington CR, Keddie G, Kosintsev P, Kunz ML, Martin LD, Stephenson RO, Storer J, Tedford R, Zimov S, Cooper A (2004) Rise and fall of the Beringian steppe bison. Science 306:1561–1565

Schubert M, Ermini L, Der Sarkissian C, Jónsson H, Ginolhac A, Schaefer R, Martin MD 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. doi:10.1038/nprot.2014.063

Soubrier J, Gower G, Chen K, Richards SM, Llamas B, Mitchell KJ, Ho SYW, Kosintsev P, Lee MSY, Baryshnikov G, Bollongino R, Bover P, Burger J, Chivall D, Crégut-Bonnoure E, Decker JE, Doronichev VB, Douka K, Fordham DA, Fontana F, Fritz C, Glimmerveen J, Golovanova LV, Groves C, Guerreschi A, Haak W, Higham T, Hofman-Kamińska E, Immel A, Julien MA, Krause J, Krotova O, Langbein F, Larson G, Rohrlach A, Scheu A, Schnabel RD, Taylor JF, Tokarska M, Tosello G, vander Plicht J, vanLoenen A, Vigne JD, Wooley O, Orlando L, Kowalczyk R, Shapiro B, Cooper A (2016) Early cave art and ancient DNA record the origin of European bison. Nat Commun 7:13158. doi:10.1038/ncomms13158

Appendix 3

Ancient DNA analysis of the extinct North American flat-headed peccary (*Platygonus compressus*).

This appendix contains the publication of a phylogenetic study including the first mitochondrial genomes from American flat-headed peccaries (*Platygonus compressus*) to resolve taxonomy and clarify peccary biogeographic evolution. I contributed to this work by performing the DNA extractions of the ancient peccary samples, and provided edits and comments on the manuscript. This article has been published in Molecular Phylogenetics and Evolution.

Perry, T., **van Loenen, A.L.**, Heiniger, H., Lee, C., Gongora, J., Cooper, A. & Mitchell, K. (2017). Ancient DNA analysis of the extinct North American flat-headed peccary (Platygonus compressus). Molecular Phylogenetics and Evolution, 112, 258-267.

Molecular Phylogenetics and Evolution 112 (2017) 258–267

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Ancient DNA analysis of the extinct North American flat-headed peccary (*Platygonus compressus*)

Tahlia Perry [a,*], Ayla L. van Loenen [a], Holly Heiniger [a], Carol Lee [b], Jaime Gongora [b], Alan Cooper [a], Kieren J. Mitchell [a,*]

[a] Australian Centre for Ancient DNA, School of Biological Sciences, University of Adelaide, Australia
[b] The University of Sydney, Faculty of Science, Sydney School of Veterinary Science, Sydney, NSW 2006, Australia

ARTICLE INFO

Article history:
Received 27 November 2016
Revised 23 February 2017
Accepted 27 March 2017
Available online 28 March 2017

Keywords:
Tayassuidae
Phylogenetics
Biogeography
Molecular dating

ABSTRACT

The geographical range of extant peccaries extends from the southwestern United States through Central America and into northern Argentina. However, from the Miocene until the Pleistocene now-extinct peccary species inhabited the entirety of North America. Relationships among the living and extinct species have long been contentious. Similarly, how and when peccaries moved from North to South America is unclear. The North American flat-headed peccary (*Platygonus compressus*) became extinct at the end of the Pleistocene and is one of the most abundant subfossil taxa found in North America, yet despite this extensive fossil record its phylogenetic position has not been resolved. This study is the first to present DNA data from the flat-headed peccary and full mitochondrial genome sequences of all the extant peccary species. We performed a molecular phylogenetic analysis to determine the relationships among ancient and extant peccary species. Our results suggested that the flat-headed peccary is sister-taxon to a clade comprising the extant peccary species. Divergence date estimates from our molecular dating analyses suggest that if extant peccary diversification occurred in South America then their common ancestor must have dispersed from North America to South America well before the establishment of the Isthmus of Panama. We also investigated the genetic diversity of the flat-headed peccary by performing a preliminary population study on specimens from Sheriden Cave, Ohio. Flat-headed peccaries from Sheriden Cave appear to be genetically diverse and show no signature of population decline prior to extinction. Including additional extinct Pleistocene peccary species in future phylogenetic analyses will further clarify peccary evolution.

1. Introduction

Peccaries are even-toed ungulates belonging to the family Tayassuidae. There are three generally accepted extant species - the collared peccary (*Pecari tajacu*), white-lipped peccary (*Tayassu pecari*), and Chacoan peccary (*Catagonus wagneri*) (Ruvinsky et al., 2011; Taber et al., 2011) - though the collared peccary may in fact represent two or three distinct species (Gongora et al., 2006; Groves and Grubb, 2011). Living peccaries are distributed throughout most of Central and South America, with the range of the collared peccary extending into the southern United States (Fig. 1). However, fossil evidence suggests that the bulk of peccary diversification occurred in North America. The earliest unequivocal fossil peccary from the Big Badlands in North Dakota - *Perchoerus minor* - dates to ~37 million years ago (mya) (Prothero, 2009). Additional extinct peccary species have been described from across North America dating from the Miocene onwards, while the Central American record begins ~10 mya (Johnston, 1935; MacFadden et al., 2010). In contrast, the fossil record of peccaries in South America is more limited and largely restricted to the middle Pliocene (~4 mya) onwards (Gasparini and Ubilla, 2011; Prevosti et al., 2006). The timing and number of peccary dispersals from North America to South America remain uncertain, partially because of conflicting morphological and molecular evidence for the relationships among extant peccary species (Gasparini et al., 2010; Gongora and Moran, 2005; Wetzel, 1977). In addition, the precise relationships between living and extinct peccaries or among extinct peccary species are still unknown. The extinct North American flat-headed peccary (*Platygonus compressus*) is a good

* Corresponding authors at: Department of Genetics and Evolution, School of Biological Sciences, University of Adelaide, Australia (T. Perry); Australian Centre for Ancient DNA, School of Biological Sciences, University of Adelaide, Australia (K.J. Mitchell).

E-mail addresses: tahlia.perry@adelaide.edu.au (T. Perry), kieren.mitchell@adelaide.edu.au (K.J. Mitchell).

http://dx.doi.org/10.1016/j.ympev.2017.03.024
1055-7903/

Fig. 1. Geographic distribution of the extinct flat-headed peccary and the three extant peccary species in the Americas. Grey dots indicate Pleistocene fossil localities yielding remains of the extinct flat-headed peccary; Sheriden Cave, Ohio is marked with a grey cross. Current distributions of collared peccary (blue), white-lipped peccary (orange) and Chacoan peccary (yellow) are reproduced from the IUCN red list database (IUCN Version 2015.4, www.iucnredlist.org). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

candidate for inclusion in a molecular phylogenetic study as its recent extinction at the end of the Pleistocene (approximately 11 thousand years ago; kya) means that it should be possible to extract ancient DNA (aDNA) from its subfossil remains. Incorporating both extinct and extant species in a phylogenetic analysis may help to clarify the evolutionary history of peccaries.

Two main hypotheses have been advanced regarding the biogeography and phylogeny of the extant peccaries, primarily based on palaeontological evidence and their current geographic distributions. The first hypothesis is that the Most Recent Common Ancestor (MRCA) of extant peccaries migrated from North America to South America and diversified during the Pleistocene, with collared and white-lipped peccaries being each other's closest relative (Wetzel, 1977). The second hypothesis is that the living peccary species diverged in North America during the Late Miocene and independently colonised South America, and that the collared and Chacoan peccaries are each other's closest relative (Wright, 1989). A more recent genetic study using the mitochondrial cytochrome *b* (cytb) gene suggested a third hypothesis: that the Chacoan and white-lipped peccary are most closely related, with their divergence occurring after their MRCA colonised South America, approximately 1.7–2.3 mya (Theimer and Keim, 1998). Gongora and Moran (2005) expanded Theimer and Keim's (1998) phylogenetic results using additional mitochondrial data and five nuclear loci, yet suggested that the divergence between Chacoan and white-lipped peccaries occurred during the early Pliocene or Pleistocene, in either North or South America. Despite the growing molecular evidence to the contrary, some researchers still refer to the white-lipped and collared peccaries as sister-taxa, sometimes even considering them congeneric (e.g. Gasparini et al., 2014, 2010).

The taxonomic affinity of the extinct flat-headed peccary is even less clear than the relationships among the extant species. Wetzel (1977) hypothesised that *Platygonus* is the ancestor to *Catagonus* (which comprises several fossil species in addition to the Chacoan peccary) based on shared morphological characters such as high crowned teeth, posterior-set orbits and a long, distinct rostrum. Other morphological studies suggest that the two genera are synonymous (Benirschke et al., 1985; Mayer and Wetzel, 1986; Tonni et al., 2009). The issue is further complicated by the fact that

fossil representatives of both genera have been described dating from the late Miocene to the Pleistocene (Czaplewski, 2012; Hibbard and Riggs, 1949; Prothero and Grenader, 2012; Webb et al., 2008). It is clear from the literature that morphological analyses on their own are not yet sufficiently robust to unequivocally define the relationships among peccary species. Thus, inclusion of the flat-headed peccary in a molecular phylogenetic analysis may help to clarify peccary taxonomy, and consequently contribute towards our understanding of other aspects of peccary evolutionary history and biogeography.

Representatives of *Platygonus* have been described from both North and South America. While systematics of the North American species are debated, Prothero and Grenader (2012) recognised seven: *P. rex* and *P. pollenae* from the late Miocene, *P. bicalcaratus*, *P. texanus* and *P. pearcei* from Pliocene-early Pleistocene, *P. vetus* (= *P. cumberlandensis*) from early-mid Pleistocene along with *P. compressus* from mid-late Pleistocene (referred to throughout as the flat-headed peccary). South American species include: *P. kraglievichi*, *P. scagliai*, *P. marplatensis*, *P. chapadmalensis* and *P. cinctus* all dating from middle Pliocene to early Pleistocene (Gasparini, 2013). The most recent known flat-headed peccary remains (from *P. compressus*) date to approximately 11,000 years ago, after which time it appears to have become extinct, making it the latest surviving *Platygonus* species. The flat-headed peccary's Pleistocene fossil record suggests a widespread distribution from the east to west coast of the United States of America, south to Mexico, and as far north as Canada (Fig. 1) (Beebe, 1980; Czaplewski, 2012; Lucas and Smartt, 1995; Murray et al., 2005). Compared to its congeners, the flat-headed peccary was better adapted for eating grasses, which may have contributed to their survival through the Last Glacial Maximum (LGM) while all other North American peccary species became extinct (Prothero and Grenader, 2012; Wright, 1998, 1993). During the late Pleistocene mass megafaunal extinctions occurred and controversy remains around the exact causes of these extinctions (Barnosky et al., 2004; Cooper et al., 2015; Haynes, 2014). The flat-headed peccary may serve as an important model species for studying the impact of late Pleistocene environmental change on megafauna populations, as their numerous and widespread fossils from across North America may allow the accumulation of significant amounts of genetic data, which can be compared to past vegetation and climate changes.

Abundant flat-headed peccary (*P. compressus*) subfossil remains have been recovered from Sheriden Cave, an exceptional archaeological and palaeontological site in North-Western Wyandot County, Ohio (Fig. 1). One of the most valuable attributes of Sheriden Cave is that the sediment has produced clear, datable stratigraphic layers: a rare finding in the United States of America (Waters et al., 2009). The age of flat-headed peccary subfossils from Sheriden Cave range in age from approximately 11–12.5 kya (Redmond and Tankersley, 2005), representing a transect of approximately 1500 years preceding the mass extinction of megafauna across North America (including the flat-headed peccary). Due to the relatively young age of the Sheriden Cave flat-headed peccary remains, they represent good candidates for aDNA analysis. Traditionally, it has been challenging to obtain informative genetic data from subfossil specimens as aDNA is highly fragmented, damaged, and in low concentration relative to environmental contaminants (Rohland and Hofreiter, 2007). However, high-throughput sequencing (HTS) technologies and new complexity reduction techniques (e.g. hybridisation enrichment) have dramatically increased our ability to recover high quality genetic information from subfossil specimens, even those from suboptimal preservation environments (e.g. hot and humid) (Hofreiter et al., 2015).

Our aim was to confidently resolve the phylogeny of peccaries, including the extinct flat-headed peccary, by sequencing complete mitochondrial genomes for each species. We further aimed to use molecular dating analyses to explore whether diversification of modern peccaries occurred in North or South America. In addition, we performed a preliminary survey of flat-headed peccary genetic diversity at Sheriden Cave and assessed whether diversity and aDNA preservation are adequate for the flat-headed peccary to serve as a future model species for studying megafaunal extinction. To address these aims we used hybridisation enrichment and HTS to generate DNA sequence data for samples of 12 flat-headed peccary specimens and five additional specimens representing the three extant peccary species.

2. Methods

2.1. Site description and ancient specimens

Sheriden Cave (40°58′30″N, 83°27′00″W) is a cave formed during the late Pleistocene by underground waters that widened voids in the dolomite bedrock. The widened region was overlayed with glacial ice 14.8–21 kya (Tankersley, 1997). When the glacier retreated, meltwater formed the cave's entrance. Layers of sediment accumulated in Sheriden Cave approximately 10–13 kya and feature a variety of animal fossils throughout this time period (Redmond and Tankersley, 2005). From the initial discovery of the cave, archaeological excavations ran for an eleven-year period (1990–2001) and flat-headed peccaries were the most abundant fossils found within the cave (McDonald, 1994; Redmond and Tankersley, 2005). In the present study, 12 flat-headed peccary fossils (8 bones and 4 teeth) were selected for aDNA analysis spanning the full range of sediment depths from which peccaries have been recovered in Sheriden Cave, presumably representing up to a 1.5 kya transect from 11–12.5 kya (Tables S1 and S2).

2.2. Ancient DNA extraction

Sample preparation and extraction of our 12 flat-headed peccary samples were performed in the dedicated aDNA laboratory at the Australian Centre for Ancient DNA (ACAD), University of Adelaide. Most samples appeared relatively clean and had surfaces rinsed with ethanol to reduce contamination; however, two samples that appeared degraded (ACAD2864 and ACAD2865) had all external surfaces removed with a Dremel tool and were bleached (5% sodium hypochlorite) before being rinsed with ethanol. Bones and teeth were then broken into small (~1 mm^3) pieces with a hammer for more efficient digestion. All samples were lysed by rotational incubation at 37 °C for 30 min in 4 mL of EDTA pH 8.0. The supernatant from this first EDTA digestion was discarded in order to remove all remaining surface contamination. A second 4 mL of EDTA pH 8.0 and 80 μL of proteinase-K was added to the sample and left overnight at 55 °C under rotation. The released DNA from this second lysis step was bound, washed, and eluted using the silica-based method of Brotherton et al. (2013).

2.3. Modern DNA extraction

A sample of dried skin from a Chacoan peccary (*Catagonus wagneri*; ACAD17698) was obtained from the Smithsonian Institute, Washington D.C., USA. The skin was sliced into small fragments using a scalpel blade to improve digestion efficiency. DNA was extracted with a Qiagen DNeasy Blood & Tissue kit, following the manufacturers' protocol.

DNA extracts from a single white-lipped peccary (*Tayassu pecari*; ACAD19076) and three collared peccaries (*Pecari tajacu*; ACAD19074, ACAD19075 and ACAD19073) were provided by Jaime Gongora, The University of Sydney. Collared peccaries ACAD19074

and ACAD19075 were from Colombia, South America while ACAD19073 was from Mexico, Central America. DNA was extracted from these tissue samples using a standard proteinase K and phenol-chloroform protocol (Sambrook et al., 1989) at the Sydney School of Veterinary Science, The University of Sydney.

DNA extracted from extant peccaries was diluted (10 μL DNA in 120 μL TLE buffer) and underwent shearing at SA Pathology, Adelaide using a Covaris focused-ultrasonicator to fragment DNA into approximately 200 bp lengths suitable for HTS platforms.

2.4. Library preparation

Ancient and modern DNA extracts were subjected to the same library preparation protocol, with the exception that the ancient libraries were prepared in ACAD's dedicated aDNA laboratory while the modern libraries were prepared in the University of Adelaide's Australian Centre for Evolutionary Biology and Biodiversity (ACEBB) labs. All samples were enzymatically repaired and blunt-ended, and truncated Illumina adapters were ligated following the protocol of Meyer and Kircher (2010). Unique barcode sequences were included within the 5′ (P5) and 3′ (P7) adapters to allow downstream identification of samples (Table S3). Libraries were amplified by PCR using primers complementary to the adapter sequences. Each library was split into eight replicate PCRs. Each individual PCR (25 μL) contained 1 × Gold PCR buffer, 2.5 mM $MgCl_2$ or $MgSO_4$, 1 mM dNTPs, 0.5 mM of each primer (Table S4), 0.1 U Amplitaq Gold and 5 μL DNA. Cycling conditions were as follows: 94 °C for 12 min; 13 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 40 s (plus 2 s/cycle); and 72 °C for 10 min. PCR products were purified using AMPure magnetic beads (Agencourt). DNA was eluted in 30 μL TLE buffer and quantified with a Qubit fluorometer (Thermo Fisher). After amplification, five of the flat-headed peccary samples (ACAD2883, ACAD2891, ACAD2895, ACAD2901 & ACAD2902) had insufficient quantities of DNA for hybridisation enrichment and so underwent a second round of amplification (13 cycles; conditions as above), followed by AMPure purification and quantification.

2.5. Hybridisation enrichment

Commercially synthesised 80 mer RNA baits (MYcroarray, MI, USA) were designed using published mitochondrial DNA sequences (Table S5) in order to target a wide range of placental mammals (Mitchell et al., 2016). DNA-RNA hybridisation enrichment was performed according to manufacturer's recommendations (MYbaits protocol v1) with the exception of the incubation step, which we extended to 44 h (3 h at 60 °C, 12 h at 55 °C, 12 h at 50 °C, 17 h at 55 °C). After incubation, magnetic MyOne Streptavidin beads (Life Technologies) were added to bind to the biotinylated RNA baits. The RNA baits were washed twice with 0.1 × SSC and 0.1% SDS (10 min at 50 °C) on a magnetic rack to remove any unbound DNA. The hybridised DNA and MyOne beads were resuspended in 30 μL of water. Full-length Illumina sequencing adapters were then added to the enriched libraries with the use of fusion primers (Table S4) via a final round of "off-bead" PCR (13 cycles; conditions as above). During the first stage of the PCR at 94 °C the RNA baits denature from the library DNA and are subsequently destroyed during the reaction. Following this PCR, the libraries were again purified with AMPure magnetic beads and quantified on TapeStation (Agilent Technologies) to check that no adapter dimers were present.

2.6. Sequencing and data processing

The enriched flat-headed peccary libraries were pooled, diluted to 2nM, and run on an Illumina MiSeq (2 × 150 bp paired-end). All modern species of peccary were pooled, diluted to 0.2nM, and run on an Illumina NextSeq (2 × 150 bp paired-end). After sequencing, reads were demultiplexed via both 5′ and 3′ barcodes using 'sabre' (https://github.com/najoshi/sabre). Adapter sequences were removed and paired-end reads were merged using AdapterRemoval v1.5.2 (Lindgreen, 2012). Low quality bases were trimmed (<Phred20 --minquality 4) and reads shorter than 25 bp were discarded (--minlength 25). Read pairs that successfully merged were retained for mapping.

We mapped reads belonging to the modern peccary libraries to the previously published mitochondrial genome of the collared peccary (*Pecari tajacu* NC_012103) using BWA v0.7.8 (parameters –n 0.01, -l 1024 and –o 2). Mapped reads for each library were used to create a 50% majority consensus sequence. Reads for each library were then re-mapped against their respective consensus sequence. This process was iterated until no additional reads could be mapped. A final round of mapping was then performed at a higher level of stringency (-n 0.04) and a 75% majority consensus sequence (min. read depth = 3×) was generated for each library.

Preliminary mapping analyses revealed that ACAD2848 had the highest endogenous DNA content of the 12 flat-headed peccaries that we sequenced. Consequently, we mapped reads from ACAD2848 against the collared peccary mitochondrial genome (*Pecari tajacu* NC_012103) following the same iterative mapping procedure described above. After 11 iterations, we obtained a final high-quality flat-headed peccary consensus sequence for ACAD2848 (Table S6). All remaining flat-headed peccaries were mapped using this flat-headed peccary sequence as a reference, and final consensus sequences were generated as above (with the exception that a higher stringency round of mapping was not performed). Table S6 details the mapping details for each library including the total reads assigned, number of unique reads mapped, percentage of non-unique reads mapped, length of consensus sequence, percentage of mitochondrial genome covered, mean read length, mean read depth, and standard deviation of read depth.

MapDamage v2.0.2 (Jónsson et al., 2013) was used to assess the patterns of DNA damage across mapped reads for all samples (Table S6) after mapping reads back to the final consensus sequence for each library. Ancient DNA has diagnostic damage patterns due to post-mortem DNA degradation: elevated 5′ C-to-T and 3′ A-to-G mutations, and an elevated frequency of purines (A or G) at the position preceding the beginning of reads. All flat-headed peccary samples showed damage profiles consistent with ancient DNA. Conversely, the modern samples displayed negligible levels of nucleotide misincorporations and no systematic bias towards purines or pyramidines at the position preceding the beginning of reads.

2.7. Phylogenetic analysis

We aligned our flat-headed peccary mitochondrial genomes with comparable data from six collared peccaries (including the previously published sequences NC_012103, JN632682, and JN632683), one white-lipped peccary, one Chacoan peccary, and four outgroups using the ClustalW algorithm in Geneious v8.0.4 (http://www.geneious.com, Kearse et al., 2012). Two outgroup taxa belonged to Suidae, the closest living relatives to peccaries within the suborder Suina (*Sus barbatus* NC_026992 and *Phacochoerus africanus* NC_008830), and two belonged to the ungulate clade Whippomorpha (*Hippopotamus amphibious* NC_000889 and *Orcinus orca* KF164610). PartitionFinder v1.1.1 (Lanfear et al., 2012) was used to determine the optimal partitioning scheme and substitution models for maximum likelihood and Bayesian phylogenetic analyses. We provided PartitionFinder with 32 regions of the mitochondrial genome, including stem and loop positions of the 12S and 16S

rRNAs and the (concatenated) tRNAs, and first and second codon positions of the protein-coding genes. Third codon positions were excluded in order to minimise potential bias arising from substitution saturation.

We generated a maximum likelihood phylogeny using RAxML v8.1.2 (Stamatakis, 2014) under the partitioning scheme suggested by PartitionFinder (Table S7). Our RAxML analysis comprised a maximum likelihood search for the best scoring tree from 1000 bootstrap replicates.

BEAST v1.8.2 (Drummond et al., 2012) was used for phylogenetic analysis and molecular dating following the partitioning scheme and substitution models determined by PartitionFinder (Table S8). Our maximum likelihood analysis revealed that all flat-headed peccary samples formed a strongly supported clade, and that North and South American representatives of the collared peccary formed reciprocally monophyletic clades. Consequently, we subsampled our alignment such that it better conformed to Speciation tree priors implemented in BEAST by including only one flat-headed peccary sample (ACAD2848) and two collared peccary samples (NC_012103 and JN632682) representing the North and South American clades. We implemented a Birth-Death tree prior and a single uncorrelated relaxed lognormal clock model, while allowing relative rates to vary between partitions. The molecular clock was calibrated by constraining the age of the root of the tree – the MRCA of Suina and Whippomorpha – according to a lognormal prior distribution with an offset of 52.4 mya, a mean of 6.77, and a standard deviation of 0.5 (such that 95% of the prior probability fell between 52.4 and 66 mya). The lower bound of 52.4 mya was based on *Himalayacetus*, the oldest known representative of Whippomorpha (Bajpai and Gingerich, 1998; O'Leary and Uhen, 1999). It is difficult to objectively set an upper limit for this node and so we used a conservative figure of 66 mya. This upper bound is based on the split between Perissodactyls (odd-toed ungulates) and Artiodactyls (even-toed ungulates), which most fossil evidence suggests occurred during the Palaeocene (Cifelli, 1981). We also constrained the age of the MRCA of Tayassuidae (peccaries) and Suidae (pigs) using a lognormal prior distribution with an offset of 37 mya, a mean of 14.44, and a standard deviation of 0.5 (such that 95% of the prior probability fell between 37 and 66 mya). The lower limit of 37 mya was based on the earliest undisputed tayassuid fossil: *Perchoerus minor,* from the Big Badlands in North Dakota (Prothero, 2009). Our BEAST MCMC chain was run for 10^8 generations sampling every 10^4 generations, with the first 10% of samples discarded as burnin. Tracer v1.8.2 (Rambaut et al., 2014; http://beast.bio.ed.ac.uk/Tracer) was used to assess the BEAST run, ensuring convergence and ESSs >200. The BEAST analysis was run three times using different starting tree topologies to ensure they were not reaching local optima. Sampled trees and parameter values from each run were combined before summarising the results.

2.8. Flat-headed peccary population analysis

We subsampled our alignment to include only the 12 flat-headed peccary mitochondrial genomes. Network analysis was performed (median-joining algorithm) with PopART v1.7.2 (Bandelt et al., 1999; Leigh and Bryant, 2015). Sample ACAD2865 was removed from the analysis due to an excess of ambiguous sites, though it is notable that this sample possessed a unique single nucleotide polymorphism and thus a different haplotype to the remaining samples. Arlequin v3.5.2.2 (Excoffier and Lischer, 2010) was used to output population statistics such as molecular diversity indices (e.g. nucleotide diversity) and neutrality tests (e.g. Tajima's D and Fu's FS) from the alignment of the 12 flat-headed peccaries.

3. Results

3.1. Phylogenetic results & molecular dating

Our maximum likelihood phylogenetic analysis produced a well-resolved species-level tree (Figs. S1 and S2). The reciprocal monophyly of Suidae and Tayassuidae received unequivocal support (Maximum Likelihood Bootstrap, MLB = 100%), as did a sister-taxon relationship between *Platygonus* and a clade comprising the extant peccary species (Maximum Likelihood Bootstrap, MLB = 100%). Within the extant peccaries, the white-lipped peccary (*Tayassu pecari*) and the Chacoan peccary (*Catagonus wagneri*) appear to be sister-species (node D; MLB = 100%), to the exclusion of the collared peccary (*Pecari tajacu*) (node B; MLB = 99%). Both *Platygonus* and *Pecari* were monophyletic (MLB = 100%), and within *Pecari* we found strong support for reciprocal monophyly between North and South American samples (MLB = 100%; Fig. S1).

The tree produced by our BEAST analyses of the subsampled alignment (Fig. 2) was concordant with our maximum likelihood phylogeny at all equivalent nodes and all ingroup clades received the maximum possible support (Bayesian Posterior Probability, BPP = 1.0). Our molecular dating analyses placed the node age estimates for divergence between flat-headed peccaries and all modern peccaries at 21.7 mya (node A; 95% Highest Posterior Density [HPD] 13.1–31.3 mya); between collared peccaries and white-lipped/Chacoan peccaries at 17.0 mya (node B; 95% HPD 9.7–25.0 mya); between white-lipped and Chacoan peccaries at 7.7 mya (node D; 95% HPD 3.2–12.6 mya); and between North and South American collared peccaries at 5.6 mya (node E; 95% HPD 2.1–9.6 mya). The date for node C, the divergence between the two suid species, is 11.6 mya (95% HPD 4.6–19.5 mya), which is consistent with previously reported results (Gongora et al., 2011).

3.2. Population analysis results

PopART collapsed the 11 sequences (excluding ACAD2865) into five haplotypes using 16 of the 23 polymorphic sites (Fig. 3). The remaining seven polymorphic sites were ambiguously called in at least one sample and could therefore not be considered in the haplotype network analysis. Samples ACAD2902 and ACAD2864 may belong to the same peccary individual, as they collapsed into the same haplotype and belong to the same depth category (80-90 cm). Similarly, ACAD2848 and ACAD2894 (depth = 90-120 cm) may represent the same individual. However, it seems unlikely that the other samples that collapsed into the same haplotypes belonged to the same individuals as they originated from different depths. Interestingly, there appears to be no temporal structure in the haplotype diversity (Fig. 3). Relationships among the 12 flat-headed peccaries in our maximum likelihood phylogenetic analysis (Figs. S1 and S2) recapitulate the haplotype groupings from our network analysis. However, bootstrap values were generally low, reflecting the relatively low number of variable sites. Surprisingly, Fu's FS was significantly negative (−5.92; p-value = 0.002), suggesting population expansion. The less sensitive Tajima's D was not significant (0.13; p-value = 0.60). Nucleotide diversity among the 12 flat-headed peccaries was 0.000466 ± 0.000263. To confirm these results, population statistics were rerun excluding ACAD2865 due to the excess in ambiguous sites, also excluding one each of the samples that potentially belong to the same individuals: ACAD2864 and ACAD2894. Fu's FS remained significantly negative (-2.96; p-value = 0.017), while Tajima's D reported a non-significant positive value (0.82; p-value = 0.84). Nucleotide diversity among the nine flat-headed peccaries was 0.000561 ± 0.000323.

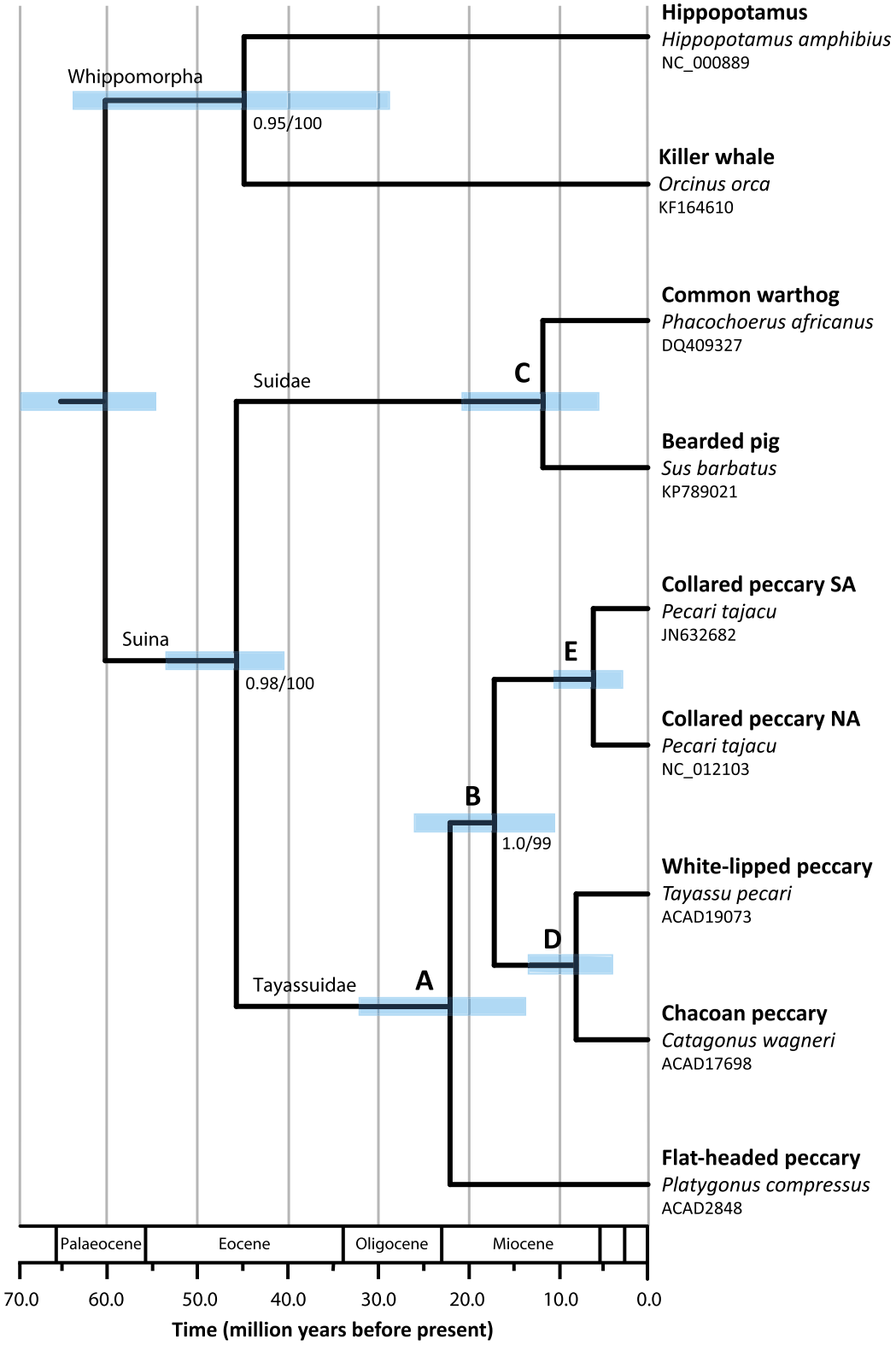


Fig. 2. BEAST maximum clade credibility tree. Mean node ages are given and node bars represent the 95% Highest Posterior Density (HPD). Support values (Bayesian Posterior Probability/Maximum Likelihood Bootstrap) are provided for nodes that did not receive maximum support (1.0/100). Nodes labelled A–E are discussed further in the main text. NA = North America; SA = South America.

4. Discussion

Our analyses demonstrate that the divergence between the extinct flat-headed peccary and a clade comprising the extant South American species occurred in the early Miocene, around 22 mya (Fig. 2). The fossil record further supports a Miocene divergence, as North American remains attributed to *Platygonus* (*P. pollenae* and *P. rex*) have been described from as early as the Hemphilian (~10.3–4.9 mya) (Prothero and Grenader, 2012). In addition, our phylogeny refutes previous suggestions that

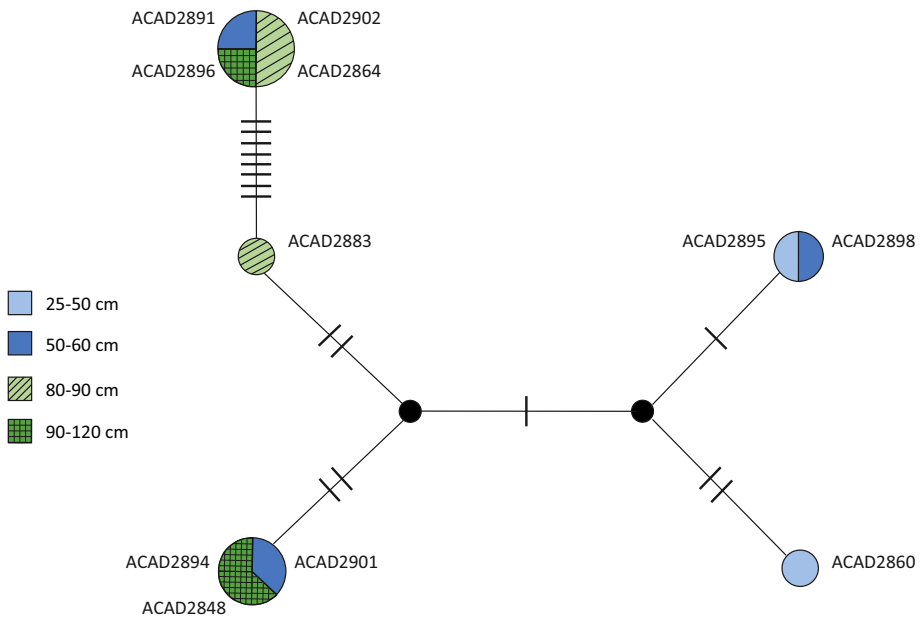


Fig. 3. Haplotype network of flat-headed peccary samples. Haplotypes are labelled with their corresponding identification number. Sizes of spheres represent number of individuals per haplotype. Colour and pattern represents stratigraphic depth of sample; age of fossil increases with depth. Perpendicular lines represent the number of nucleotide changes along each branch. Note that ACAD2865 was not included in this analysis due to an excess of missing data but possessed a unique nucleotide substitution differentiating it from all other samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Platygonus and *Catagonus* are closely related or synonymous. Observed morphological similarities between these two genera are possible examples of parallel evolution, as both possess physical adaptations to living in open, arid environments (Wetzel, 1977). However, it is also possible that both *Platygonus* and *Catagonus* more closely resemble the ancestral peccary morphology compared with the remaining extant peccaries. Uncertain relationships among fossil peccary taxa make these hypotheses about morphological evolution difficult to assess.

Monophyly of the three living peccary species is consistent with the widely accepted hypothesis of a South American diversification of extant peccaries, with the collared and white-lipped peccaries undergoing back-migration into Central and North America (Albert et al., 2004; Gasparini, 2013; Mayer and Wetzel, 1987). However, if this is the case then our age estimates suggest that the MRCA of extant peccaries must have reached South America more than 10 mya. This conflicts with the long-standing notion that Late Cenozoic faunal migrations between North and South America - termed the Great American Biotic Interchange (Marshall et al., 1982) – began only after the Isthmus of Panama formed ~3 mya (O'Dea et al., 2016; Woodburne, 2010). The results of past studies of animals including armadillos, opossums, ground sloths, bears, horses, wolves, and llamas are largely consistent with this timeline (Miller and Carranza-Castañeda, 1984; Prevosti et al., 2009; Webb, 2006). However, it is becoming clear that early pulses of faunal migration may have occurred prior to the formation of the Isthmus of Panama, including taxa such as peccaries, gomphotheres, tapirs, and camelids (Campbell et al., 2010, 2000; Prothero et al., 2014). Consistent with our genetic results, MacFadden et al. (2010) have described a number of peccary specimens (*"Cynorca" occidentale*) from central Panama dating from early-middle Miocene and two extinct genera of peccaries have recently been described (*Sylvochoerus* and *Waldochoerus*) from fossils found in the western Amazon basin dating to ~9 mya (Frailey and Campbell, 2012). Frailey and Campbell (2012) noted similarities of the Miocene peccaries in South America to the collared and white-lipped peccaries and tentatively suggested there may be ancestral-descendant relationships between them.

If the MRCA of living peccaries did not itself migrate to South America, then our divergence date estimates likely reflect an initial North American diversification followed by at least two independent migrations of peccaries to South America. A late Hemphillian age (~5 mya) peccary, *Catagonus brachydontus,* has been described from Mexico, Texas, and Florida (Wright, 1989; Prothero and Grenader, 2012), which may be interpreted as evidence that the common ancestor of *Catagonus* and *Tayassu* inhabited North America. However, it has alternatively been suggested that *C. brachydontus* is most closely related to *Platygonus* and not the South American *Catagonus* (Parisi Dutra et al., 2016). Regardless of the distribution of the MRCA of the living peccaries, some movement of peccaries may have been associated with the formation of the Isthmus of Panama. We confirm that a previously reported split in the mitochondrial diversity of the collared peccary, corresponding to the North and South American populations (Gongora et al., 2006; Gongora and Moran, 2005), occurred 5.57 mya (95% HPD = 2.14–9.56 mya), which is consistent with the timing of land connection. In addition, undisbuted South American peccary fossils belonging to *Platygonus marplatensis* (and Tayassuidae indet.) have been discovered that date to 3–4 mya (Prevosti et al., 2006; Gasparini and Ubilla, 2011). These specimens may reflect the overland migration of an ancestral *Platygonus* taxon into South America, which subsequently diversified into the five recognised South American *Platygonus* species.

In addition to testing biogeographic hypotheses, our data also allowed us to examine the genetic diversity of a North American flat-headed peccary (*P. compressus*) population just prior to their extinction. We anticipated seeing a population with low diversity - possibly reflecting a small population size - which has been observed in other aDNA studies for populations nearing extinction (Barnett et al., 2009; Campos et al., 2010; Shapiro et al., 2004). Instead, the haplotype network (Fig. 3) showed no apparent structure reflecting temporal signal. In fact, the negative value for Fu's FS (−5.92; p-value = 0.002) suggests population expansion. The lack of a signal of population decline in the flat-headed peccaries may reflect favourable local environmental conditions prior to their extinction. Flat-headed peccaries mostly inhabited semi-

arid to arid regions with open woodlands and grasslands (Gasparini et al., 2010; Wetzel, 1977), which would have been widespread in North America during the Last Glacial Maximum. Severe climatic and vegetation changes occurring in North America at the beginning of the Holocene may have caused the rapid extinction of flat-headed peccaries (within only a few generations), which would make a genetic signature of decline difficult to detect. Indeed, Gasparini and Ubilla (2011) suggested that the South American *Platygonus* species became extinct as a result of spreading wet and woodland habitats along with competition from competing *Catagonus* species. Bone artefacts recovered from the subfossil-bearing strata in Sheriden Cave (Redmond and Tankersley, 2005) suggest that human presence may also have played a role in the flat-headed peccary's extinction, perhaps exacerbating the effects of rapid environmental change by disrupting normal population processes (Cooper et al., 2015). Comparing flat-headed peccaries from alternate locations and over a greater time period with climatic and vegetation changes would provide a more comprehensive understanding of the species' demography and diversity preceding extinction.

The flat-headed peccary shows promise as a model species for future investigation into Pleistocene megafaunal extinctions. Sufficient DNA to construct mitochondrial genomes was obtained from *P. compressus* remains, presumably due to the cave environment providing protection from the hot and humid summers common in Ohio. This suggests that similar results may be obtained from analogous cave sites in the contiguous USA, which have previously been considered marginal for DNA preservation. Given the abundance, widespread distribution (Fig. 1), and apparent genetic diversity of flat-headed peccary remains (even at only a single locality), a significant amount of genetic data could potentially be accumulated for this species. These data could then be used to model how this species reacted in the face of extreme climatic and vegetation changes prior to its extinction. As this study used only mitochondrial data, the incorporation of nuclear genes will be important to further test the results. However, so far it has not been possible to sequence nuclear genes from the flat-headed peccary. Future efforts should also be directed towards procuring genetic data from other recently extinct peccaries, such as the South American *Platygonus* and *Catagonus* species or the North American long-nosed peccary (*Mylohyus nasutus*), which also became extinct at the end of the Pleistocene. The inclusion of more extinct species in the phylogenetic analyses will be important for further clarifying the evolution and biogeographic history of peccaries.

Data accessibility

Mitochondrial genome consensus sequences presented in this study are available on GenBank (Accession Numbers: *P. compressus*: KY987551, KY987542, KY987553, KY987554, KY987544, KY987550, KY987548, KY987546, KY987552, KY987547 and KY987549; *P. tajacu*: KY987555, KY987543 and KY987541; *T. pecari*: KY987556; *C. wagneri*: KY987545). Note that sample ACAD2865 had an excess of ambiguous sites and could not be uploaded to GenBank. All sequencing reads, scripts, alignments, analysis input files, and trees are uploaded to the Dryad Digital Repository (doi: http://dx.doi.org/10.5061/dryad.71tv1).

Conflict of interest

The authors declare that there is no conflict of interest.

Acknowledgements

We thank the Cleveland Museum of Natural History (B. Redmond), Cincinnati Museum Center (G. Storrs), and the Smithsonian Institute (D. Lunde, K. Helgen, N. Edmison) for access to specimens in their collections. We are also grateful to the Hendricks family for permitting and supporting the excavation of Sheriden Cave, and to H.G. McDonald for his contributions to the excavation. Funding for this project was provided by the Australian Research Council (ARC). Grid computing facilities were provided by CIPRES (Cyberinfrastructure for Phylogenetic Research). The comments of B. Redmond (CMNH) and H. G. MacDonald (USA Bureau of Land Management) improved the final version of the manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2017.03.024.

References

Albert, S., Ramotnik, C.A., Schmitt, C.G., Edwards, C.W., 2004. Collared peccary range expansion in northwestern New Mexico. Southwest. Nat. 49, 524–528. http://dx.doi.org/10.1894/0038-4909(2004) 049<0524:CPREIN>2.0.CO;2.

Bajpai, S., Gingerich, P.D., 1998. A new Eocene archaeocete (Mammalia, Cetacea) from India and the time of origin of whales. Proc. Natl. Acad. Sci. 95, 15464–15468. http://dx.doi.org/10.1073/pnas.95.26.15464.

Bandelt, H.J., Forster, P., Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37–48.

Barnett, R., Shapiro, B., Barnes, I., Ho, S.Y.W., Burger, J., Yamaguchi, N., Higham, T.F.G., Wheeler, H.T., Rosendahl, W., Sher, A.V., Sotnikova, M., Kuznetsova, T., Baryshnikov, G.F., Martin, L.D., Harington, C.R., Burns, J.A., Cooper, A., 2009. Phylogeography of lions (*Panthera leo* ssp.) reveals three distinct taxa and a late Pleistocene reduction in genetic diversity. Mol. Ecol. 18, 1668–1677. http://dx.doi.org/10.1111/j.1365-294X.2009.04134.x.

Barnosky, A.D., Koch, P.L., Feranec, R.S., Wing, S.L., Shabel, A.B., 2004. Assessing the causes of late Pleistocene extinctions on the continents. Science 306, 70–75. http://dx.doi.org/10.1126/science.1101476.

Beebe, B.F., 1980. Pleistocene peccary, *Platygonus compressus* Le Conte, from Yukon Territory, Canada. Can. J. Earth Sci. 17, 1204–1209. http://dx.doi.org/10.1139/e80-126.

Benirschke, K., Kumamoto, A.T., Meritt, D.A., 1985. Chromosomes of the Chacoan peccary, *Catagonus wagneri* (Rusconi). J. Hered. 76, 95–98.

Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Jane Adler, C., Richards, S.M., Sarkissian, C.D., Ganslmeier, R., Friederich, S., Dresely, V., van Oven, M., Kenyon, R., Van der Hoek, M.B., Korlach, J., Luong, K., Ho, S.Y.W., Quintana-Murci, L., Behar, D.M., Meller, H., Alt, K.W., Cooper, A., Consortium, T.G., 2013. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. Nat. Commun. 4, 1764. http://dx.doi.org/10.1038/ncomms2656.

Campbell, K.E., Frailey, C.D., Romero Pittman, L., 2000. The late Miocene gomphothere Amahuacatherium peruvium (Proboscidea: Gomphotheriidae) from Amazonian Peru: Implications for the Great American Faunal Interchange - [Boletín D 23]. Inst. Geológico Min. Met. - INGEMMET.

Campbell, K.E., Prothero, D.R., Romero-Pittman, L., Hertel, F., Rivera, N., 2010. Amazonian magnetostraigraphy: Dating the first pulse of the Great American Faunal Interchange. J. South Am. Earth Sci. 29, 619–626. http://dx.doi.org/10.1016/j.jsames.2009.11.007.

Campos, P.F., Willerslev, E., Sher, A., Orlando, L., Axelsson, E., Tikhonov, A., Aaris-Sørensen, K., Greenwood, A.D., Kahlke, R.-D., Kosintsev, P., Krakhmalnaya, T., Kuznetsova, T., Lemey, P., MacPhee, R., Norris, C.A., Shepherd, K., Suchard, M.A., Zazula, G.D., Shapiro, B., Gilbert, M.T.P., 2010. Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (*Ovibos moschatus*) population dynamics. Proc. Natl. Acad. Sci. 107, 5675–5680. http://dx.doi.org/10.1073/pnas.0907189107.

Cifelli, R.L., 1981. Patterns of evolution among the Artiodacyla and Perissodactyla (mammalia). Evolution 35, 433–440. http://dx.doi.org/10.2307/2408192.

Cooper, A., Turney, C., Hughen, K.A., Brook, B.W., McDonald, H.G., Bradshaw, C.J.A., 2015. Abrupt warming events drove Late Pleistocene Holarctic megafaunal turnover. Science 349, 602–606. http://dx.doi.org/10.1126/science.aac4315.

Czaplewski, N.J., 2012. Pleistocene peccaries (Mammalia: Tayassuidae) from Western Oklahoma. Southwest. Nat. 57, 112–117. http://dx.doi.org/10.1894/0038-4909-57.1.112.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973. http://dx.doi.org/10.1093/molbev/mss075.

Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564–567. http://dx.doi.org/10.1111/j.1755-0998.2010.02847.x.

Frailey, C.D., Campbell, K.E., 2012. Two new genera of peccaries (Mammalia, Artiodactyla, Tayassuidae) from upper Miocene deposits of the Amazon basin. J. Paleontol. 86, 852–877. http://dx.doi.org/10.1666/12-012.1.

Gasparini, G.M., Ubilla, M., 2011. *Platygonus sp*. (Mammalia: Tayassuidae) in Uruguay (Raigón? Formation; Pliocene–early Pleistocene), comments about its distribution and palaeoenvironmental significance in South America. J. Nat. Hist. 45, 2855–2870. http://dx.doi.org/10.1080/00222933.2011.620716.

Gasparini, G.M., 2013. Records and stratigraphical ranges of South American Tayassuidae (Mammalia, Artiodactyla). J. Mamm. Evol. 20, 57–68. http://dx.doi.org/10.1007/s10914-011-9172-z.

Gasparini, G.M., Rodriguez, S.G., Soibelzon, L.H., Beilinson, E., Soibelzon, E., Missagia, R.V., 2014. *Tayassu pecari* (Link, 1795) (Mammalia, Cetartiodactyla) comments on its South American fossil record, taxonomy and paleobiogeography. Hist. Biol. 26, 785–800. http://dx.doi.org/10.1080/08912963.2013.858247.

Gasparini, G.M., Soibelzon, E., Zurita, A.E., Miño-Boilini, A.R., 2010. A review of the quaternary Tayassuidae (Mammalia, Artiodactyla) from the Tarija Valley, Bolivia. Alcheringa Australas. J. Palaeontol. 34, 7–20. http://dx.doi.org/10.1080/03115510903277717.

Gongora, J., Cuddahee, R.E., Nascimento, F.F.do, Palgrave, C.J., Lowden, S., Ho, S.Y.W., Simond, D., Damayanti, C.S., White, D.J., Tay, W.T., Randi, E., Klingel, H., Rodrigues-Zarate, C.J., Allen, K., Moran, C., Larson, G., 2011. Rethinking the evolution of extant sub-Saharan African suids (Suidae, Artiodactyla) evolution of extant African Suidae. Zool. Scr. 40, 327–335. http://dx.doi.org/10.1111/j.1463-6409.2011.00480.x.

Gongora, J., Morales, S., Bernal, J.E., Moran, C., 2006. Phylogenetic divisions among collared peccaries (*Pecari tajacu*) detected using mitochondrial and nuclear sequences. Mol. Phylogenet. Evol. 41, 1–11. http://dx.doi.org/10.1016/j.ympev.2006.05.015.

Gongora, J., Moran, C., 2005. Nuclear and mitochondrial evolutionary analyses of collared, white-lipped, and Chacoan peccaries (Tayassuidae). Mol. Phylogenet. Evol. 34, 181–189. http://dx.doi.org/10.1016/j.ympev.2004.08.021.

Groves, C., Grubb, P., 2011. Ungulate Taxonomy. JHU Press.

Haynes, G., 2014. North American megafauna extinction: climate or overhunting? In: Smith, C. (Ed.), Encyclopedia of Global Archaeology. Springer, New York, pp. 5382–5390.

Hibbard, C.W., Riggs, E.S., 1949. Upper Pliocene vertebrates from Keefe Canyon, Meade County, Kansas. Geol. Soc. Am. Bull. 60, 829–860. http://dx.doi.org/10.1130/0016-7606(1949) 60[829:UPVFKC]2.0.CO;2.

Hofreiter, M., Paijmans, J.L.A., Goodchild, H., Speller, C.F., Barlow, A., Fortes, G.G., Thomas, J.A., Ludwig, A., Collins, M.J., 2015. The future of ancient DNA: technical advances and conceptual shifts. BioEssays 37, 284–293. http://dx.doi.org/10.1002/bies.201400160.

IUCN, 2015. IUCN Red List of Threatened Species. Version 2015.2. <www.iucnredlist.org>.

Johnston, C.S., 1935. An extension in the range of fossil peccaries. Am. Midl. Nat. 16, 117–119. http://dx.doi.org/10.2307/2419885.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., Orlando, L., 2013. MapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 29, 1682–1684. http://dx.doi.org/10.1093/bioinformatics/btt193.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basican integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649. http://dx.doi.org/10.1093/bioinformatics/bts199.

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. http://dx.doi.org/10.1093/molbev/mss020 (mss020).

Leigh, J.W., Bryant, D., 2015. Popart: full-feature software for haplotype network construction. Methods Ecol. Evol. 6, 1110–1116. http://dx.doi.org/10.1111/2041-210X.12410.

Lindgreen, S., 2012. AdapterRemoval: easy cleaning of next generation sequencing reads. BMC Res. Notes 5, 337. http://dx.doi.org/10.1186/1756-0500-5-337.

Lucas, S.G., Smartt, R.A., 1995. Late Pleistocene peccary from northwestern New Mexico. Southwest. Nat. 40, 293–296.

MacFadden, B.J., Kirby, M.X., Rincon, A., Montes, C., Moron, S., Strong, N., Jaramillo, C., 2010. Extinct peccary "Cynorca" Occidentale (Tayassuidae, Tayassuinae) from the Miocene of Panama and Correlations to North America. J. Paleontol. 84, 288–298. http://dx.doi.org/10.1666/09-064R.1.

Marshall, L.G., Webb, S.D., Sepkoski, J.J., Raup, D.M., 1982. Mammalian evolution and the Great American interchange. Science 215, 1351–1357. http://dx.doi.org/10.1126/science.215.4538.1351.

Mayer, J.J., Wetzel, R.M., 1987. *Tayassu pecari*. Mamm. Species 1. http://dx.doi.org/10.2307/3503865.

Mayer, J.J., Wetzel, R.M., 1986. *Catagonus wagneri*. Mamm. Species 1–5. http://dx.doi.org/10.2307/3503829.

McDonald, H.G., 1994. The late Pleistocene vertebrate fauna in Ohio: coinhabitants with Ohio's Paleoindians. In: The First Discovery of Americaarchaeological Evidence of the Early Inhabitants of the Ohio Area. The Ohio Archaeological Council, Columbus, pp. 23–39.

Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. 2010, pdb.prot5448. doihttp://dx.doi.org/10.1101/pdb.prot5448.

Miller, W.E., Carranza-Castañeda, O., 1984. Late Cenozoic mammals from central Mexico. J. Vertebr. Paleontol. 4, 216–236. http://dx.doi.org/10.1080/02724634.1984.10012005.

Mitchell, K.J., Bray, S.C., Bover, P., Soibelzon, L., Schubert, B.W., Prevosti, F., Prieto, A., Martin, F., Austin, J.J., Cooper, A., 2016. Ancient mitochondrial DNA reveals convergent evolution of giant short-faced bears (Tremarctinae) in North and South America. Biol. Let. 12, 20160062. http://dx.doi.org/10.1098/rsbl.2016.0062.

Murray, L.K., Bell, C.J., Dolan, M.T., Mead, J.I., Jones, C.A., 2005. Late Pleistocene fauna from the southern Colorado Plateau, Navajo County, Arizona. Southwest. Nat. 50, 363–374. http://dx.doi.org/10.1894/0038-4909(2005) 050[0363:LPFFTS]2.0.CO;2.

O'Dea, A., Lessios, H.A., Coates, A.G., Eytan, R.I., Restrepo-Moreno, S.A., Cione, A.L., Collins, L.S., de Queiroz, A., Farris, D.W., Norris, R.D., Stallard, R.F., Woodburne, M.O., Aguilera, O., Aubry, M.-P., Berggren, W.A., Budd, A.F., Cozzuol, M.A., Coppard, S.E., Duque-Caro, H., Finnegan, S., Gasparini, G.M., Grossman, E.L., Johnson, K.G., Keigwin, L.D., Knowlton, N., Leigh, E.G., Leonard-Pingel, J.S., Marko, P.B., Pyenson, N.D., Rachello-Dolmen, P.G., Soibelzon, E., Soibelzon, L., Todd, J.A., Vermeij, G.J., Jackson, J.B.C., 2016. Formation of the Isthmus of Panama. Sci. Adv. 2, e1600883. http://dx.doi.org/10.1126/sciadv.1600883.

O'Leary, M.A., Uhen, M.D., 1999. The time of origin of whales and the role of behavioral changes in the terrestrial-aquatic transition. Paleobiology 25, 534–556.

Parisi Dutra, R., Casali, D.de M., Missagia, R.V., Gasparini, G.M., Perini, F.A., Cozzuol, M.A., 2016. Phylogenetic systematics of peccaries (Tayassuidae: Artiodactyla) and a classification of South American tayassuids. J. Mamm. Evol. 1–14. http://dx.doi.org/10.1007/s10914-016-9347-8.

Prevosti, F., Gasparini, G., Bond, M., 2006. On the systematic position of a specimen previously assigned to Carnivora from the Pliocene of Argentina and its implications for the Great American Biotic Interchange. Neues Jahrb. Geol. Palaontologie-Abh. 242, 133–144.

Prevosti, F.J., Tonni, E.P., Bidegain, J.C., 2009. Stratigraphic range of the large canids (Carnivora, Canidae) in South America, and its relevance to quaternary biostratigraphy The Ensenadan Stage/Age in southern South America Quat. Int. 210, 76–81. http://dx.doi.org/10.1016/j.quaint.2009.06.034.

Prothero, D., 2009. The early evolution of the North American peccaries (Artiodactyla: Tayassuidae). Mus. North. Ariz. Bull. 65, 509–541.

Prothero, D.R., Campbell, K.E., Beatty, B.L., Frailey, C.D., 2014. New late Miocene dromomerycine artiodactyl from the Amazon Basin: implications for interchange dynamics. J. Paleontol. 88, 434–443. http://dx.doi.org/10.1666/13-022.

Prothero, D.R., Grenader, J., 2012. A new primitive species of the flat-headed peccary *Platygonus* (Tayassuidae, Artiodactyla, Mammalia) from the Late Miocene of the High Plains. J. Paleontol. 86, 1021–1031. http://dx.doi.org/10.1666/12-050R.1.

Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J., 2014. Tracer v1.6. Available from <http://beast.bio.ed.ac.uk/Tracer>.

Redmond, B.G., Tankersley, K.B., 2005. Evidence of early paleoindian bone modification and use at the Sheriden Cave site (33WY252), Wyandot County, Ohio. Am. Antiq. 70, 503–526. http://dx.doi.org/10.2307/40035311.

Rohland, N., Hofreiter, M., 2007. Ancient DNA extraction from bones and teeth. Nat. Protoc. 2, 1756–1762. http://dx.doi.org/10.1038/nprot.2007.247.

Ruvinsky, A., Rothschild, M., Larson, G., Gongora, J., 2011. Systematics and evolution of the pig. In: The Genetics of the Pig. CABI Publishing, Wallingford, UK, pp. 1–13.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning. Cold Spring Hardbour Laboratory Press, New York.

Shapiro, B., Drummond, A.J., Rambaut, A., Wilson, M.C., Matheus, P.E., Sher, A.V., Pybus, O.G., Gilbert, M.T.P., Barnes, I., Binladen, J., Willerslev, E., Hansen, A.J., Baryshnikov, G.F., Burns, J.A., Davydov, S., Driver, J.C., Froese, D.G., Harington, C. R., Keddie, G., Kosintsev, P., Kunz, M.L., Martin, L.D., Stephenson, R.O., Storer, J., Tedford, R., Zimov, S., Cooper, A., 2004. Rise and fall of the Beringian Steppe Bison. Science 306, 1561–1565. http://dx.doi.org/10.1126/science.1101074.

Stamatakis, A., 2014. RAxML version 8a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 5 (5). http://dx.doi.org/10.1093/bioinformatics/btu033 (btu033).

Taber, A., Altrichter, M., Harald, B., Gongora, J., 2011. Family Tayassuidae (Peccaries). In: Handbook of the Mammals of the World - Vol 2Hoofed Mammals. Lynx Edicions, Barcelona, Spain, pp. 292–307.

Tankersley, K.B., 1997. Sheriden: A Clovis cave site in eastern North America. Geoarchaeology 12, 713–724. http://dx.doi.org/10.1002/(SICI)1520-6548(199709)12:6<713::AID-GEA9>3.0.CO;2-1.

Theimer, T.C., Keim, P., 1998. Phylogenetic relationships of peccaries based on mitochondrial cytochrome b DNA sequences. J. Mammal. 79, 566–572. http://dx.doi.org/10.2307/1382987.

Tonni, E.P., Soibelzon, E., Cione, A.L., Carlini, A.A., Scillato Yané, G.J., Zurita, A.E., Ríos, F.P., 2009. Preliminar correlation of the Pleistocene sequences of the Tarija valley (Bolivia) with the Pampean chronological standard The Ensenadan Stage/Age in southern South America Quat. Int 210, 57–65. http://dx.doi.org/10.1016/j.quaint.2009.06.015.

Waters, M.R., Stafford Jr., T.W., Redmond, B.G., Tankersley, K.B., 2009. The age of the paleoindian assemblage at Sheriden Cave, Ohio. Am. Antiq. 74, 107–111. http://dx.doi.org/10.2307/25470540.

Webb, S.D., 2006. The Great American Biotic Interchange: patterns and processes. Ann. Mo. Bot. Gard. 93, 245–257. http://dx.doi.org/10.3417/0026-6493(2006)93[245:TGABIP]2.0.CO;2.

Webb, S.D., Hulbert Jr., R.C., Morgan, G.S., Evans, H.F., 2008. Terrestrial mammals of the Palmetto Fauna (early Pliocene, latest Hemphillian) from the central Florida phosphate district. Nat. Hist. Mus. Los Angel. Cty. Sci. Ser. 41, 293–312.

Wetzel, R.M., 1977. The extinction of peccaries and a new case of survival. Ann. N. Y. Acad. Sci. 288, 538–544. http://dx.doi.org/10.1111/j.1749-6632.1977.tb33641.x.

Woodburne, M.O., 2010. The Great American Biotic Interchange: dispersals, tectonics, climate, sea level and holding pens. J. Mamm. Evol. 17, 245–264. http://dx.doi.org/10.1007/s10914-010-9144-8.
Wright, D.B., 1998. Tayassuidae. In: Evolution of Tertiary Mammals of North America. Cambridge University Press, pp. 389–401.
Wright, D.B., 1993. Evolution of sexually dimorphic characters in peccaries (Mammalia, Tayassuidae). Paleobiology 19, 52–70.
Wright, D.B., 1989. Phylogenetic relationships of *Catagonus wagneri*: sister taxa from the Tertiary of North America. In: Advances in Neotropical Mammalogy. Sandhill Crane Press, Gainesville, Florida, pp. 281–308.