Retrieval of DNA from ancient material and the demography of the extinct aurochs

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Conor Rossi

A thesis submitted for the degree of
Doctor of Philosophy

Smurfit Institute of Genetics
Trinity College Dublin

2023

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Signed _____ Date _____

Thesis Summary

Ancient DNA has revolutionised our understanding of animal genomics by providing direct evidence of the demography of ancient wild populations and unravelling the complex domestication process with time-stamped precision. Following over a decade of technological advancements and laboratory innovations, the first part of this thesis explores the retrieval of DNA from ancient and historical sources. These chapters leverage underutilised ancient materials, such mummified skin and coprolite, to test the limits of aDNA survival and reveal local histories of sheep and goat domestication. The second part of this thesis undertakes a large genomic survey of *Bos primigenius* - the extinct ancestors of modern cattle. An ancient dataset is curated with whole genome sequencing of 108 *Bos* individuals sampled from a period spanning 200,000 years. The mitochondrial, Y chromosomal and autosomal past of *Bos primigenius* is interrogated using population genomic analyses to illuminate the evolution, domestication and eventual extinction of this megafaunal Pleistocene animal.

Chapter 3 presents the screening results of 285 ancient and historical animal specimens for endogenous DNA content. A substantial obstacle in the field is the low endogenous DNA content and *postmortem* damage associated with aDNA. These challenges are explored by employing laboratory techniques to maximise the retrieval of endogenous DNA.

Chapter 4 focuses on the remarkable DNA integrity of a ~1,600 year old naturally mummified sheep recovered from Chehrābād, a salt mine in northwestern Iran. The saline taphonomic environment of this sample led to substantial reduction in hallmarks of aDNA damage, including fragmentation and hydrolytic deamination. The profound influence of this environment is further reflected in a metagenomic profile that is dominated by halophilic Archaea, not typical of ancient archaeological material. Finally, the excellent preservation of this sheep facilitated population genomic analyses and domestic trait genotyping, providing insight into Sasanian-period animal husbandry.

Chapter 5 describes a pilot study of ancient DNA retrieval from Levantine goat coprolites. Successful amplification and shotgun sequencing of aDNA from these specimens presented a time series of goat management in the southern Levant. In unprecedented coprolite aDNA preservation, autosomal and mitochondrial genomes are generated that reveal continuity of Levantine goat ancestry from the Neolithic, through the Bronze Age until present-day. The metagenomic landscape of these coprolites is also interrogated, producing a detectable ancient ungulate gut microbiome and signals of past domestic diets.

Chapter 6 attempts to sequence DNA from modern cattle leather using common ancient DNA techniques designed for low-concentration, fragmented and damaged DNA. Very few studies have

reported successful DNA amplification from modern or ancient leather sources, likely due to the harsh tanning process. Successful high-throughput sequencing of the DNA reveals unique damage patterns, including high fragmentation and nucleotide misincorporations due to chromium (III) interactions with DNA. An unusual enrichment of repetitive and mitochondrial sequences is also observed, possibly due to a protective effect of constitutive heterochromatin and the double membrane of the mitochondrial genome.

Chapter 7 begins the second part of this thesis, centred around the extinct *Bos primigenius*. This chapter describes the mitochondrial variation of the aurochs and leverages this information to infer population structure, postglacial expansions and domestication. Extinct outgroup lineages of mitochondria are described for the first time that reveal a possible ancient hybridisation event with an outgroup sub-species of European *Bos primigenius*. Bayesian analysis of Pleistocene Eurasian lineages reveals divergences and subsequent genetic structure that may be associated with climatic oscillations. Holocene variation depicts a collapse in Eurasian mitochondrial structure and expansions associated with key domestication events are highlighted and discussed.

Chapter 8 turns to the male uniparental marker, the Y chromosome. Bayesian analysis further emphasises genetic structure between populations of Pleistocene Eurasian wild lineages. Temporal and geographical surveying reveals a weakly-structured and complex dispersal of Holocene Y chromosomes, suggesting increased male-mobility. The wild dataset is supplemented with modern and ancient domestic bulls to describe multiple complete turnovers in domestic male lineages, revealing localised patterns of early cattle management.

Chapter 9 describes the autosomal variation of *Bos primigenius*. Deep demography history and population structure using MSMC to infer response to Pleistocene climatic changes. Genetic analyses characterise genetic isolation and admixture between different populations through the Last Glacial Maximum when much of the species range was fractured. Aspects of Eurasian glacial refugia and post-glacial expansion are explored in detail, providing possible routes of recolonisation of northern territories. Hybrid zones in Holocene Central Asia are described and leveraged to elucidate long-range migrations of Pleistocene populations. The final stage of the aurochs is explored as intensified human activity drives population to dwindling numbers and eventual extinction. Finally, the genomic legacy of aurochs is described in modern breeds which reveals that local wild incorporations were widespread and varied.

Acknowledgements

First and foremost, I would like to sincerely thank Dan Bradley. The past four or so years of my research have been guided by your mentorship and enthusiasm for the subjects. It is a real privilege to have a fair and supportive supervisor to spearhead this work. Much of the research of this thesis was conducted during the events of a global pandemic and I am very grateful for the flexibility and understanding that was given during that time.

My gratitude is also offered to Valeria Mattiangeli. Since it all began with my fourth year Undergraduate project, you have been such a reassuring and guiding presence in the lab. It has been a saving grace to rely on your expertise throughout this whole PhD. From bone drilling, sequencing demultiplex failures and discussions about leather DNA - thank you for always keeping the whole show on the road.

Kevin Daly - my *"academic big brother"* - thanks for all the patience (a lot of it), guidance, knowledge and advice throughout the years. Our work on the mummy was one of the most instructive projects for this entire thesis and I continue to learn so much from you. Victoria Mullin - my *"academic big sister"* - similarly, you have been so utterly patient and kind over the last few years. It is also such a rare delight to have someone as interested in ancient cows as me. To Mikkel Sinding, I would also like to thank you for recruiting me for the aurochs work which has been such a brilliant opportunity. Your enthusiastic encouragement and vision kept this project going (despite the fact I never truly became "Mr Younger Dryas").

This PhD has benefitted from the wealth of institutional knowledge of many past members of the Bradley lab. I would like to especially thank Lara Cassidy, who started my aDNA journey in 2018 with ancient hares, for always being generous with your time and so encouraging. I would also like to acknowledge Matthew Teasdale, Marta Verdugo and Amelie Schue who have each contributed to this work in significant ways. I appreciate having both Emily Breslin and Bruno Ariano as guides, especially at the start of this work. It has also been a pleasure to work alongside Iseult Jackson - who joined a year after me but still taught me a lot! To the extended population genomics group of Trinity College Dublin (McLaughlin Group, Nakagome Group, Huerta-Sanchez Group, Cassidy Group); it is a privilege to have such an ever-growing community of knowledgeable researchers.

My PhD would not have been possible without the large number of curators, zooarchaeologists and archaeologists who have provided me with ancient animal material to extract DNA. In

particular, I would like to thank Marjan Mashkour for her collaboration and guidance on the sheep mummy chapter. I would also like to thank Steve Rosen and Liora Horwitz for the collaboration in relation to the goat coprolites. Finally, thank you to Erika Rosengren and Claus Kropp for providing me with aurochs material and the many helpful and enjoyable conversations we've had.

To my class of Genetics 2018 - I'm very thankful to have had such a solid group of people for my Undergraduate and I'm very glad to count so many of you among my close friends still. In particular, Ciaran and Tom - it has been very helpful to have you both to share in the trials of PhD life. To my VDP crew - I feel so lucky to have such a supportive and diverse group of people to count as my friends. The "Japlan" crew - what could have been… #Japan2024? Then of course to my VDP Exec crew - I consider our year together as the pinnacle of my time in Trinity. Luke and Kevin - good luck in your academic endeavours; I'll read your theses if you want. Thank you also to my assortment of close friends I've kept since my Trinity days. I think we're now spread over 3 or 4 continents. My "home" crew which I kind of inherited over the last few years - another group that has been lost to emigration..! Special thanks to Jean who was my roomie for about a year in Del Val and bore witness to the odd hours I kept while working. I also want to give a special shout-out to Michael and Rian who are my oldest friends. Thanks for the many, many laughs through the years. Michael in particular thanks for being such a supportive friend/chauffeur. And to Eimhin, thanks for always taking an interest in my work (sarcasm) and for everything else (not sarcasm). I'm very much looking forward to all of our adventures together.

I'd like to express my heartfelt thanks to my family for always encouraging me to pursue what I wanted in life. To my Mom and Dad, for supporting all my endeavours and putting my education first (also to Mom for allowing me to take over her kitchen to write this thesis!). To my sister, for paving the way for us all and being an ever-present guiding force. And to my brother, who is a great friend and probably the one in the family who is the most interested in this work. In the summer of 2022, sadly my Dad passed away after a long illness, borne with courage and resilience. Dad was always amused by my stories about the world of ancient DNA, and so I am dedicating this thesis to him.

Table of Contents

1. Introduction

1.1. A brief history of ancient population genomics

Population genetics first emerged as a field of research in the early 20th century when the use of blood antigens and antibodies demonstrated that the variation in human populations had a molecular basis (Landsteiner 1900). However, it wasn't until the 1970s that the molecule of inheritance, DNA, was directly examined in population genetics through the discovery that restriction enzymes could recognise and cleave DNA sequences at specific recognition sites (Danna and Nathans 1971). Numerous restriction enzymes were subsequently purified and used to perform restriction fragment length polymorphism (RFLP) analysis and construct restriction site maps (Brown and Vinograd 1974). Patterns of cleavage were exploited to describe variation between and within species and populations and statistical frameworks followed in order to describe these phenomena (Nei and Li 1979). Another major development of molecular biology was launched in the late 1970s with the invention of DNA Sanger sequencing (Sanger, Nicklen, and Coulson 1977). Unlike RFLP, Sanger sequencing characterised variation DNA fully and would remain the sequencing technology of choice for nearly 30 years. These innovations led to many discoveries in the field of population genetics, such as estimating the most recent common maternal ancestor of all humans (Cann, Stoneking, and Wilson 1987).

The field of ancient DNA (aDNA) began in 1984 when 229 nucleotides of a museum specimen quagga were successfully sequenced using molecular probes (Higuchi et al. 1984). Using these data, the authors were able to time divergence from the extant sister species zebra, unlocking information that had been previously lost to extinction. While this study was a proof of principle, it was precocious. The authors reported that less than 1% of DNA was extracted from the dried muscle than what would be expected from a fresh sample (Higuchi et al. 1984); a characteristic of aDNA (Pääbo et al. 2004). Furthermore, endogenous DNA of the target organism presented itself in a mixture of microbial and environmental DNA fractions introduced to the tissue in taphonomic processes. These aspects seriously impeded aDNA analysis using Sanger sequencing technologies which required a large amount of target DNA fragments for amplification through bacterial cloning. Thus, the first major advancement of the field of aDNA occurred with the development of the polymerase chain reaction (PCR) (Mullis and Faloona 1987; Saiki et al. 1988) which promised to overcome the extremely low concentrations associated with aDNA.

A number of studies in the late 1980s characterised post-mortem damage patterns that were intrinsic to aDNA (Pääbo 1989). Nucleotide misincorporation caused by hydrolytic deamination of cytosine bases was discovered to be a hallmark of aDNA (Hofreiter et al. 2001). This occurred especially often at overhangs of aDNA fragments, where cytosine residues were converted to uracils (Briggs et al. 2007). During amplification steps of DNA library preparation, proofreading polymerases (such as *Taq* polymerase) misread the uracil as a thymine which results in a C>T misincorporation upon amplification (Hofreiter et al. 2001). aDNA fragments were also revealed to be heavily fragmented, possibly caused by strand cleavage of hydrolytic depuration of amino acid residues and β-elimination (Lindahl 1993). This characteristic short fragment length made ancient samples susceptible to contamination during the PCR step, as PCR primers preferentially amplify longer fragments of modern DNA (Cooper 1997; Stoneking 1995). The pervasiveness of contamination drew scepticism toward aDNA studies and analysis was generally restricted to mitochondrial DNA (mtDNA) which had a higher cellular copy number. A number of extinct species were characterised using ancient mtDNA including the marsupial wolf, Neanderthal and European aurochs (Thomas et al. 1989; Krings et al. 1997; Bailey et al. 1996). However, as biotechnology continued to rapidly advance in the new millennium with the landmark publication of a draft human genome (Lander et al. 2001; Venter et al. 2001), the power of aDNA analysis appeared comparatively limited.

The second major advancement in technology relevant to aDNA analysis was the development of next generation sequencing (NGS) technologies. Prior to NGS platforms, the largest amount of ancient autosomal sequencing data published was ~27,000 base pairs (Noonan et al. 2005); the first data produced by NGS were several orders of magnitude higher, at ~13 million base pairs (Poinar et al. 2006). NGS platforms like Illumina operate by massive parallelisation of short sequencing reads and are serendipitously suited to truncated aDNA fragments; while their high-throughput nature resulted in a higher DNA yields at a fraction of the cost (Margulies et al. 2005; Mardis 2008; Meyer and Kircher 2010). NGS DNA library innovations captured more aDNA fragments and protected against modern contamination through the use of universal adapter ligation (Mardis 2008; Meyer and Kircher 2010). aDNA analysis moved into the genomics era and within the first few years of NGS technologies their power was demonstrated with the publication of three archaic hominid genomes (Green et al. 2010; Rasmussen et al. 2010; Reich et al. 2010). Using these genomic sequences, it was shown that humans split from Neanderthals between 270,000 and 440,000 years ago and that a previously unknown group, the Denisovans, coexisted with Neanderthals in Eurasia. Non-African humans met and hybridised with the Neanderthals and Denosivans as they dispersed through Eurasia,

shaping the genomes of non-Africans today (Meyer et al. 2016). These studies were quickly followed with the publication of dozens of ancient human datasets, shedding light on questions relating to migration that had perplexed archaeologists for decades (Gamba et al. 2014; Jones et al. 2015; Cassidy et al. 2016).

While early studies focused on soft tissue, such as mummified remains, as the substrate for aDNA (Pääbo, Gifford, and Wilson 1988), it was soon shown that bone and teeth were viable targets for aDNA analysis (Hagelberg and Clegg 1991; Hänni et al. 1990). This greatly expanded the samples that could be targeted, as mineralised tissues are much more abundant in archaeological assemblages. However, a fundamental issue of low endogenous content remained in ancient samples which impeded the ability to sequence large datasets of ancient genomes even with NGS technologies. As a result, the sequencing of high-depth genomes was a sporadic occurrence dependent on the discovery of a uniquely well-preserved bone or tooth (Park et al. 2015; Meyer et al. 2012). In 2014, it was discovered that the petrosal part of the temporal bone was an excellent reservoir of aDNA, yielding the highest amount of endogenous DNA compared to any other ancient bone (Gamba et al. 2014). Today datasets comprising hundreds of ancient genomes are available for multiple species (Librado et al. 2021; Allentoft et al. 2022). As the field continues to progress, different substrates are considered for aDNA analysis; such as targeting specific tissues and bone elements to capture ancient pathogens (Margaryan et al. 2018), or use of sediment DNA to reconstruct whole ancient ecosystems (Gelabert et al. 2021; Murchie et al. 2021).

1.2. The Pleistocene Epoch

The Pleistocene Epoch (2,580-11.7 thousand years ago (Ka)) is a period in Earth history that is characterised by glacial episodes that cycled approximately every 100,000 years in a process known as the Quaternary glaciation (Shackleton 2000; Lisiecki and Raymo 2005; Ehlers and Gibbard 2008). The Last Glacial Period (LGP) occurred approximately 115,000 to 11,700 years ago and had several periods of glacial advance and retreat. The cooling peaked in Europe around 20,000 years ago during the Last Glacial Maximum (LGM), when the Weichselian ice sheet expanded across northern Europe, covering it in a permafrost tundra (Svendsen et al. 2004). During this time, European warm-adapted species retreated to southern refugia in the Balkans, Italy and Iberia, resulting in genetic isolation between refuge populations (G. M. Hewitt 1999). The use of aDNA greatly aided the understanding of how species evolved during the Pleistocene by reconstructing population dynamics and phenotypes, such as skin pigment, that are not captured in the fossil record (Hofreiter 2008). In a recent aDNA study, the temporal limit of DNA survival has been tested by successfully

sequencing a sample over one million years old that was preserved in permafrost (van der Valk et al. 2021).

A major occurrence in the Pleistocene was the emergence of *Homo sapiens* in Africa at least 200,000 years ago (Stringer 2016). Genetic evidence suggests that all modern non-African humans are derived from a singular Out-of-Africa (OoA) exodus, occurring approximately 70,000 to 50,000 years ago (Mallick et al. 2016). Earlier migrations from Africa of anatomically modern humans that occurred between 120,000 and 70,000 years ago, perhaps driven by glacial episodes, are thought to have left little to no heritage in modern human genomes (Timmermann and Friedrich 2016). Pleistocene humans employed a hunter-gatherer subsistence strategy (Wojtal and Wilczyński 2015). Human survival and dispersal was deeply intertwined with the climate and during the LGM humans in Europe also retreated into southern refugia, driving genetic bottlenecks (Fu et al. 2016; Posth et al. 2016). Toward the end of the Pleistocene, humans began interacting with fauna in new ways, evidenced by the appearance of the domestic dog in the archaeological record (Larson et al. 2012). This served as a prelude to a much more widespread practice of domestication that shaped human behaviour in the Holocene.

The Late Pleistocene is characterised by a massive global extinction event of terrestrial megafaunal animals (Koch and Barnosky 2006). The drivers of these extinctions is a topic of ongoing debate, although most consider climate change and anthropomorphic pressure to be the most probable causes. In North America and Australia, mass extinctions were sudden and extreme, causing the extinction of 78% and 88% of megafauna, respectively (Koch and Barnosky 2006). As most extinctions generally coincided with the arrival of humans to the region, a framework of overkill of naÏve prey by a new predator was proposed (Burney and Flannery 2005). In Australia, these extinctions may have been further precipitated by human-led habitat destruction through fires (Roberts et al. 2001). However, arguments have been made against an anthropic cause of North American extinctions due to timing inconsistencies and lack of explicit evidence of hunting of many of the lost genera (Meltzer 2020). In Eurasia, megafaunal extinctions were more protracted which has led to wide discussion regarding the impact of humans (Stuart 1991). It has been argued that long-term occupation by members of the *Homo* genus had resulted in a more robust predator-prey dynamic in Eurasia (Koch and Barnosky 2006). As such, a combination of anthropogenic and climate-driven factors has been suggested whereby megafaunal populations, driven to fractured states by climate change, may have been vulnerable to extinction by human hunting populations. For example, the climatic amelioration in post-LGM Eurasia caused population decline and range reduction for woolly

mammoths and were subsequently hunted to extinction by humans in Siberian refugia (Palkopoulou et al. 2013; Vartanyan, Garutt, and Sher 1993). However, some local extinctions of species, such as *Megaloceros* in Ireland, are unequivocally attributed to an ecological restructuring as they occurred prior to human occupation (Barnosky 1986).

1.3. Domestication

The Pleistocene gave way to the Holocene Epoch with an abrupt change to climate approximately 11,700 years ago (Shakun and Carlson 2010). The more favourable climate conditions of the Holocene facilitated a more sedentary lifestyle of hunter-gathers, ultimately leading to multiple independent instances of development of agriculture in a process known as the Neolithic Revolution (Richerson, Boyd, and Bettinger 2001; Diamond 2002). The development of agriculture saw increased human population growth, eventually leading to more complex organisation of society and dramatic anthropomorphic changes to landscapes (Gignoux, Henn, and Mountain 2011; Stephens et al. 2019; Bocquet-Appel 2011). Key to these developments was the domestication of multiple plant and animal species which provided consistent and reliable subsistence for the growing population. These domesticates were introduced to new regions as farming communities migrated from centres of domestication in complex dispersals (Arbuckle et al. 2014). The exact definition of domestication is one of sustained archaeological discussion, with relative pathways and intentionality of the process debated (Vigne 2011; Zeder 2012). Detecting signs of domestication in zooarchaeological remains is challenging when working with only fossilised material, although reduction in body and/or horn sizes and dental variation have been explored as possible markers (Zeder 2006; Evin et al. 2015). Other non-morphological signs of animal management include evidence of seasonal foddering, changes in species range or abundance and signs of harnessing (Zeder 2015; Pitulko and Kasparov 2017).

While the necessity of genetic or phenotypic change to achieve domestication is debated in zooarchaeological terms (Zeder 2015; Vigne 2015; Fuller, Willcox, and Allaby 2011), it is undeniable that ancient genetic analysis has proven key to understanding this complex process (Frantz et al. 2020). The process of animal domestication is a long one, perhaps continuing indefinitely as greater control is exerted on the reproduction of the animal (Larson and Burger 2013). Early selection for traits associated with domestication has been revealed through aDNA analysis, such as robustness in horses and coat colour in goat, horse and pig (Librado et al. 2017; Daly et al. 2018; Frantz et al. 2019). Furthermore, genomic analysis of early goat management in the Zagros Mountains depicted a restricted domestic ancestry, along with signals of reproduction control through young male kill-off and inbreeding practices

(Daly et al. 2021). Indeed, the complete genetic isolation from the wild progenitor species may have been the ultimate step of domestication (Zeuner and Others 1963). Through aDNA analysis, the extent and mechanism by which domesticated species incorporate local wild autosomal variation was revealed to be species specific. For example, the Near Eastern ancestry of the originally domesticated pigs was almost completely replaced with local European wild boar ancestry during the course of pig domestication in Europe (Frantz et al. 2019). In contrast, dogs may have been completely reproductively isolated from their wild progenitors after the establishment of a domestic ancestry thousands of years ago (Bergström et al. 2022). In the case of cattle, it is thought that limited male-mediated introgression may have restocked domestic herds, perhaps for selective advantage (Park et al. 2015; Orlando 2015).

1.4. *Bos primigenius*

Of the four livestock species originally domesticated in the Fertile Crescent, cattle are unique in that their wild progenitors are now extinct in the wild. The extinction of the aurochs is very recent compared to other Pleistocene megafauna, having died off just 400 years ago. Aurochs were larger than modern cattle, with multiple written records describing their highly aggressive and intimidating nature (van Vuure 2005). Indeed, the Bible employs a metaphor of a powerful aurochs to depict the strength of God: “God brings them out of Egypt and is for them like the horns of the wild ox” (Numbers 23:22). This ferocious reputation attracted the attention of hunters, as successful slaughter would earn them admiration and trophies of the horn, heart and skin of the aurochs (van Vuure 2005). While the foundation of aurochs research has been based in the study of cave art depictions, written records and genetic analysis of extant cattle breeds, recent advancements of aDNA have permitted direct genetic characterisation.

The origins of the *Bos* genus remains a debated issue due to a spatially and temporally patchy fossil record. The earliest proposed representative of *Bos* is *Bos acutifrons* which first appeared in the early Pleistocene (2,580 Ka) in South Asia (Wang, Flynn, and Fortelius 2013). However, both a South Asian (Groves 2009) and an African origin (Martínez-Navarro et al. 2014) have been proposed based on morphological analysis of a sparse collection of Pleistocene fossils. *Bos* first appeared in Southern Europe in the Middle Pleistocene, in Spain (700 Ka) (de Carvalho et al. 2022) and at Venosa-Notarchirico in Italy (500–600 Ka) (Martínez-Navarro et al. 2007). The earliest known remains in northern Europe appear sometime later in Britain (424-374 Ka) (Breda et al. 2010). The distribution of aurochs was once considerable; spanning much of Eurasia, North Africa, the Indian subcontinent. The aurochs were a warm-adapted species that likely grazed on grasses in open meadows or along forest margins

(Schulz and KaiSer 2007). In glacial periods, expanding tundra forced aurochs populations to retreat to refugia. During the LGM, European aurochs yielded to encroaching ice sheets by retreating to the southern peninsula (Leonardi et al. 2022). The aurochs refugium of Central and East Asia is not known, although it is thought that most northern regions were too cold to host populations during the LGM based on lack of fossil remains (van Vuure 2005). While the genetic diversity of aurochs is not well-understood, three subspecies have been described; *Bos primigenius primigenius* (Bojanus 1827) in Eurasia, *Bos primigenius namadicus* (Falconer 1859) in South Asia and *Bos primigenius mauritanicus* (Thomas 1881) in North Africa. However, the distinction between *B. p. primigenius* and *B. p. mauritanicus* is not based on any morphological attributes and these two subspecies may be considered as one (Grigson 1980); a clear cranial morphology difference existed between *B. p. primigenius* and *B. p. namadicus*, suggesting that these subspecies diverged long ago (Grigson 1980).

Extant domestic cattle also exist as two subspecies, *Bos taurus indicus and Bos taurus taurus*. The two subspecies of modern cattle have clearly defined morphological and biological differences. Indicine/zebu cattle (*Bos taurus indicus*) have a characteristic humped back with a pronounced dewlap and drooping ears and are well-adapted to arid environments (Grigson 1980). In contrast, taurine cattle (*Bos taurus taurus*) are humpless and generally less drought-resistant. As such it was considered that these subspecies, *B. taurus* and *B. indicus*, were derived from the separate wild aurochs populations of *B. p. primigenius* and *B. p. namadicus*, respectively. This framework was suspected by Charles Darwin in *On The Origin of Species*:

"I should think, from facts communicated to me by Mr Blyth, on the habits, voice and constitution etc. of the humped Indian cattle, that these had descended from a different aboriginal stock from our European cattle; and several competent judges believe these latter may have had more than one wild parent." - (Darwin 1883)

Genetic studies eventually confirmed a deep split between the wild progenitor lineages had occurred around 200,000 years ago (Loftus et al. 1994; Chen et al. 2018). This observation strongly implied two separate domestication events, which was supported by archaeological evidence. Evidence of cattle management in the Fertile Crescent indicated that taurine cattle were domesticated approximately 10,500 years ago, potentially in the Upper Euphrates Valley (Vigne et al. 2017; Arbuckle and Kassebaum 2021). Indicine cattle were independently domesticated later in the Indus Valley, around 8,000 years ago (Vigne et al. 2017; Patel and Meadow 2017). These putative centres of domestication were also supported by compelling genetic evidence (Troy et al. 2001; Achilli et al. 2008); Chen et al. 2010). It is likely that a very

small starting stock of aurochs cows were domesticated in the Fertile Crescent, leading to a restricted taurine maternal lineage (Bollongino et al. 2012; Scheu et al. 2015). Taurine cattle were widely dispersed from the Fertile Crescent into Europe, Africa and Central and East Asia as part of the spread of Neolithic culture and pastoralism, and can be easily traced by the taurine T and Q mitochondrial haplotypes (Cymbron et al. 2005; Arbuckle et al. 2014; Hanotte et al. 2000; Anthony 2010; Cai et al. 2014; Hermes et al. 2019). Indicine cattle were also moved from their centre of domestication to Southern China (Chen et al. 2010). Ancient hybridisation of taurine cattle with indicine occurred in several regions, such as the Near East and East Africa, perhaps for adaptive advantage (Freeman et al. 2004; Mwai et al. 2015; Verdugo et al. 2019).

While the process of cattle domestication has been examined by analysis of aDNA (Edwards et al. 2007; Scheu et al. 2015; Verdugo et al. 2019), comparatively few ancient aurochs genomes have been published. The first fragments of aurochs DNA published were mitochondrial sequences from two Late Pleistocene British aurochs that outgrouped all mitochondrial taurine diversity (Bailey et al. 1996). This provided the first evidence to suggest that the extant diversity of aurochs is not quantifiable with only modern domestic diversity. Later, sporadic local wild introgression was detected with divergent mitochondrail lineages sampled in modern breeds, suggesting multiple wild admixture events (Achilli et al. 2008; Bonfiglio et al. 2010; Mannen et al. 2020). To date, four genomes of Holocene aurochs have been published (Park et al. 2015; Verdugo et al. 2019). These represent aurochs populations of Europe, Southwest Asia and North Africa and displayed considerable genetic diversity and evidence of local wild incorporation. Further sampling of ancient samples is necessary to understand the extent of diversity and legacy of the wild population - especially in Central and East Asia, where no whole genomes have been sampled.

With the available data, the best view of Holocene aurochs population structure may be gleaned from mitochondrial studies which have revealed seven mitolineages termed C, E, I, P, Q, R and T (Figure 1.1). In Southwest Asia, both T (Edwards et al. 2007) and Q matrilines (Verdugo et al. 2019) have been sequenced from early Holocene aurochs. These populations were likely related to the originally domesticated aurochs. In North Africa, a single R mitogenome has been published from a Moroccan aurochs (Verdugo et al. 2019). The P haplotype predominated Late Pleistocene and Holocene Europe, with single divergent E haplotype also present in Holocene Germany (Edwards et al. 2007). While no ancient *B. namadicus* mitogenomes have been published to date, the high diversity of the I haplogroup in the Indus Valley strongly suggests that this was the region of incorporation of the originally

domesticated haplotype (Chen et al. 2010). Early to mid Holocene east Asian aurochs had been shown to harbour a divergent C haplotype (Zhang et al. 2013; Brunson et al. 2016; Cai et al. 2018). However, no autosomal DNA has been published from these aurochs. The current understanding of aurochs autosomal and mitochondrial diversity is summarised in Figure 1.1.

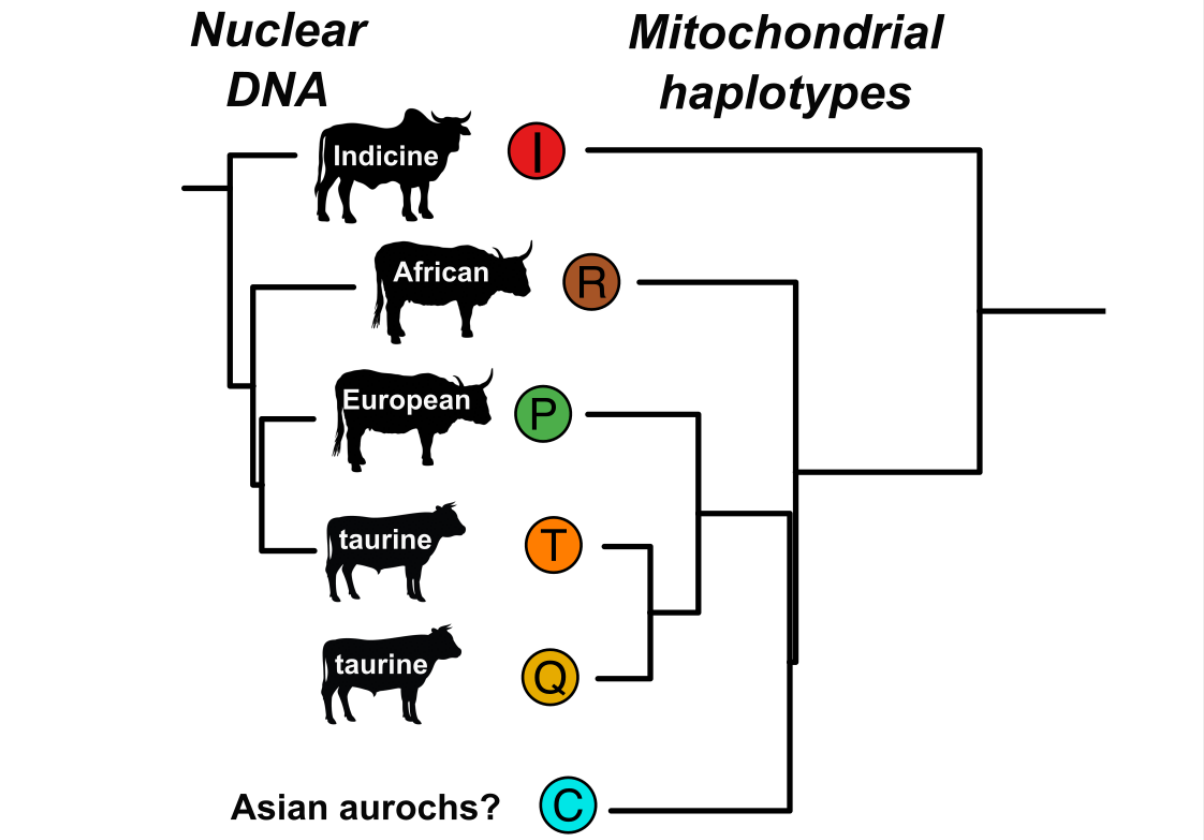


Figure 1.1 - Aurochs autosomal and mitochondrial phylogenies.

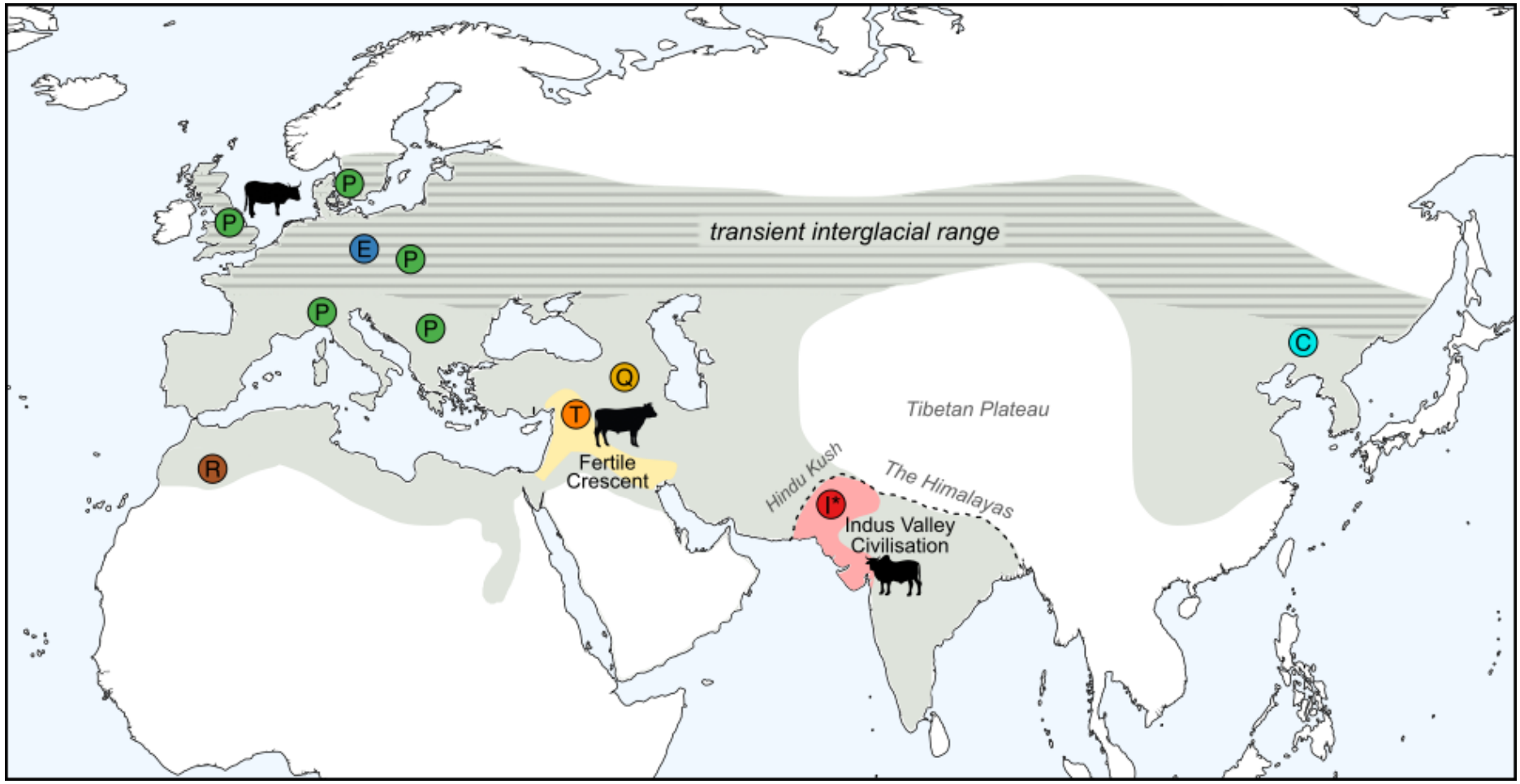


Figure 1.2 - The distribution of aurochs in the Late Pleistocene/Early Holocene. A transient interglacial range is depicted based on the description of fossil finds in (van Vuure 2005); Zhang et al. 2013; Brunson et al. 2016; Cai et al. 2018; Leonardi et al. 2022). This range was unlikely to support long term habitation during glacial periods. The two domestication (Fertile Crescent and Indus Valley Civilisation) centres are depicted. Predicted barriers to *B. namidicus* gene-flow (The Himalayas and Hindu Kush) are represented by a broken line. Ancient mitochondrial haplotype locations are shown on the map based on sequences reported in (Edwards et al. 2007; Zhang et al. 2013; Cai et al. 2018; Brunson et al. 2016; Verdugo et al. 2019). Multiple individuals of the same haplotype are represented together if recovered in close proximity. The I* haplogroup is predicted based on high diversity in modern indicine breeds (Chen et al. 2010).

2. Methods

2.1. Sample Preparation and Drilling

2.1.1. Laboratory contamination control and sample authenticity

All bones and teeth were drilled according to standard protocols to minimise contamination at a dedicated aDNA laboratory in Trinity College Dublin (Pääbo et al. 2004). Briefly, access to the aDNA laboratory was restricted to researchers who had not been in contact with amplified DNA on the same day. All work with unamplified DNA was done in protective hazmat suits, face masks, hairnets, shoe covers, and two pairs of nitrile gloves. Further, separate hazmat suits were used for the 'drilling room' and the 'molecular room' to minimise cross-contamination. Gloves were also periodically changed at different stages of drilling and when handling different samples. Surfaces and equipment were routinely cleaned with absolute (99%) ethanol, 5% sodium hypochlorite solution, and DNA-ExitusPlus. Samples were prepared within a fume hood to prevent environmental contamination. The fume hood was continuously irradiated when not in use.

Blank controls were carried out throughout the drilling, extraction, and library preparation processes. 'Air' controls were done by leaving an open Eppendorf tube in the fume hood for 30 minutes, and 'water' controls were done by adding 500µl H_2O to the mixer mill (Mixer Mill MM 400, Retsch) and shaking for 30 seconds. These controls were brought to sequencing to quantify and characterise any contamination that may occur during drilling and extraction.

2.1.2. Photography

Each sample element was photographed at various angles prior to drilling. At the latter stage of this PhD, a 3D scanner (EinScan Pro HD) was also used to produce 3D scans of bone material. Bones were scanned at a minimum of 6 angles on a rotating scanner. The scans were then aligned and a watertight mesh model was selected. The finished model is suitable for subsequent morphometric analysis.

2.1.3. Drilling of teeth and bones

Prior to drilling, bone and teeth were exposed to UV radiation for 30 minutes on each side. A dentist drill was then used to clean the surface and the exposed element was then irradiated for a further 15 minutes. A saw (Dremel) was used to target and remove the densest part of

the bone for maximum endogenous DNA recovery (Damgaard et al. 2015). For tooth samples, the cementum was targeted as this has been shown to have more DNA recovery (Higgins et al. 2013).

The removed element was then pulverised to a fine powder using a Mixer Mill. In general, multiple runs of pulverisation at a lower intensity were favoured to avoid overheating the bone. Approximately 135mg of the powderised bone was used for DNA extraction. The remaining powder was stored in a refrigerator as spare powder. When possible, a section of intact bone was also kept refrigerated.

2.1.4. Preparation of skin samples

For skin samples, a small section of skin cut was with a decontaminated scalpel to a size of approximately 1x1cm. Prior to extraction, the skin segment was washed with 1ml of H_2O (lab grade and UVed), then with 600µl of Ethanol (99%) and then again with 1ml of H_2O. Each time the supernatant was removed after centrifugation.

2.1.5. Preparation of coprolite samples

The surface of each coprolite pellet was removed using a scalpel to minimise any possible surface-level environmental contamination that occurred during deposition or excavation of the coprolite. The inner core of each individual pellet was then transferred to a clean 2ml Eppendorf tube, at varying masses.

2.1.6. C_{14} Radiocarbon dating

Approximately 1 gram of bone and teeth samples was provided for standard Accelerated Mass Spectrometry radiocarbon dating at ^{14}Chrono Centre (Belfast, United Kingdom). Uncalibrated dates were calibrated using OxCal 4.4 (Ramsey 2009) and IntCal 20 (Reimer et al. 2020).

2.2. DNA Extraction

2.2.1. “Classic” bone and tooth DNA extraction and purification

The classic bone and teeth extraction is based on methods in (Gamba et al. 2014), with modifications (Dabney and Meyer 2019). First, a pre-extraction wash was performed by adding 990µl of 0.5M ethylenediaminetetraacetic acid (EDTA) to each tube containing bone powder. The tubes were then incubated for 30 minutes at 37°C with shaking in an Eppendorf ThermoMixer. The tubes were then spun down in a centrifuge for one minute at 17,000xg and

the supernatant was removed to a clean, labelled Eppendorf tube. This “EDTA1” extraction is stored as a possible source of DNA if the subsequent extraction yields low levels of DNA.

The extraction buffer (1X) was made using 20µl 1M Tris-HCl, 17µl of SDS (2%) and 940µl of 0.5M EDTA. The buffer was exposed to UV light for 30 minutes to ensure that any foreign DNA that may have been in the solution was destroyed. After UV exposure, 13µl of Proteinase K was added to the master mix. 990µl of the extraction buffer was added to each 2ml tube containing bone/tooth powder. The tubes were covered in parafilm and vortexed for 60 seconds. The samples were then incubated at 37°C for 24 hours at 900 rpm.

After incubation for 24 hours, the tube was spun in a centrifuge for 10 minutes at 17,000xg. The supernatant was removed without disturbing the pellet of bone powder and transferred to a clean 1.5ml tubes labelled “WEX1”. In general, the bone/tooth powder dissolved to completion after one wash. However, if there was a substantial bone pellet remaining after centrifugation, a second extraction was carried out. In these circumstances, the WEX1 extract was stored in a freezer at -20°C. The remaining bone powder pellet was resuspended with a freshly made extraction buffer of the same concentrations as above. The samples were incubated at 37°C for a further 24 hours and the resulting “WEX2” is used for subsequent analysis.

To purify the chosen “WEX”, the supernatant was then transferred to a reservoir column (High Pure Viral Nucleic Acid Large Volume Kit, Roche) filled with 13ml of pH 5.5 phosphate buffer (QIAGEN). These tubes were centrifuged in a swing bucket centrifuge (Eppendorf) for 4 minutes at 800xg followed by 2 minutes at 600xg. The High Pure Filter Tube was then removed from the column and placed in a clean collection tube. This tube was dry spun in a centrifuge for 1 minute at 3,000xg. 750µl of PE buffer (Qiagen) was added to the column, and spun down for 1 minute. The flowthrough was discarded and this process was repeated for a total of 2 times. 55µl of EBT (Elution Buffer containing 0.05% Tween) was added to the tube and incubated for 2 minutes before finally centrifuging at 17,000xg.

2.2.2. Bleach bone and teeth extraction

Contaminating microbial DNA reduces the endogenous percentage of ancient samples, creating a considerable economic burden to sequence samples to higher coverage. Previously, it has been suggested that endogenous aDNA is adsorbed to hydroxyapatite located within the bone matrix (Kemp 2005). It has been reported that by applying a pre-extraction wash of bleach to ancient bone samples, higher endogenous levels may be

achieved (Boessenkool et al. 2017). This may make it possible to increase endogenous aDNA levels by removing exogenous microbial DNA on the surface of bone while leaving endogenous aDNA adsorbed to the hydroxyapatite mostly undisturbed. However, there are also conflicting reports in modern samples, that bleach can significantly reduce concentrations of DNA in teeth (Higgins et al. 2013; Kemp 2005).

The modified bleach protocol follows the same protocol as the above except it includes a pre-extraction digest based on (Boessenkool et al. 2017). Prior to the EDTA wash, 990µl of a 0.5% bleach solution was added to each tube of approximately 100mg of bone powder. The tube was vortexed and incubated at room temperature for 15 minutes on a rotator. The tube was then centrifuged briefly at 17,000xg, and the supernatant was removed. 990µl of UVed H_2O was then added to each tube and vortexed until the bone powder was solubilised. The tube was then briefly centrifuged at 17,000xg and the supernatant was once again removed. This H_2O wash was then repeated two more times.

2.2.3. Tissue extraction

DNA was extracted from tissue samples using a modified version of the classic phenol/chloroform DNA extraction protocol (Sambrook, Fritsch, and Maniatis 1989).

The extraction buffer (1X) was made using 387 µl H2O, 12.5µl 4M NaCl, 24.75µl 1M Tris, 1µl 0.5 EDTA, 24.75µl 10% SDS, 2.7µl 0.5M $CaCl_2$ and 50µl fresh Proteinase K. 500µl of the extraction buffer was added to each 2ml tube containing the skin sample. The tubes were covered in parafilm and vortexed for 60 seconds. The samples were then incubated in an Eppendorf ThermoMixer at 56°C for 24 hours at 350 rpm. Each tube was spun down for 60 seconds at 17,000 xg in a centrifuge and the supernatant was removed to a labelled Eppendorf tube (EXT1, Extraction 1). 500µl of fresh extraction buffer was prepared and the extraction was repeated twice more (EXT2 supernatant removed) on the pellet. The final EXT3 of each tissue sample (supernatant and remaining pellet of tissue) was then used for phenol/chloroform DNA purification.

1ml of phenol was added directly into each extraction tube (supernatant and pellet). The tube was vortexed briefly and rotated at room temperature for 15 minutes. The tube was centrifuged at max speed (17,000xg) for 5 minutes. The aqueous layer of the sample tube was then removed from the sample tubes and added to a new 2 mL Eppendorf tube containing 1 mL chloroform+isoamyl alcohol solution (50:1). This tube was briefly vortexed and then rotated at room temperature for 10 minutes. The tube was then centrifuged (17,000xg) for 5 minutes.

Meanwhile, 100µl of sodium acetate solution was added to a newly labelled 2ml tube. The aqueous layer (~1ml) of the reaction tube was added to these tubes containing 100µl of sodium acetate solution. 900µl of refrigerated ethanol (~4°C) was added to each tube. This was then refrigerated for at least three hours. Subsequently, the tube was centrifuged (17,000xg) for 10 minutes. The supernatant was then removed and discarded, before leaving the tubes inverted on a paper tissue for 60-90 minutes, until the ethanol had evaporated. Finally, the DNA at the bottom of the tube was eluted by adding 55µl of EBT (Elution Buffer containing 0.05% Tween) to the dry tube.

2.2.4. Coprolite extraction

An extraction buffer (1X) was made using 387 µl H_2O, 12.5 µl 4M NaCl, 24.75µl 1M Tris, 1µl 0.5 EDTA, 24.75µl 10% SDS, 2.7µl 0.5M $CaCl_2$ and 50µl fresh Proteinase K. 500µl extraction buffer was added to each tube and were vortexed for 60 seconds before being incubated in an Eppendorf ThermoMixer at 56°C for 24 hours at 350 rpm.

Each tube was then centrifuged for 60 seconds at 17,000xg. The tube was then used for a standard Phenol/Chloroform DNA purification, as above.

2.3. Library preparation and amplification

2.3.1. USER Treatment

Treatment of ancient DNA with Uracil-DNA glycosylase (UDG) and Endonuclease VIII has been shown to remove base misincorporations caused by deamination in aDNA (Briggs et al. 2010).

USER® Enzyme (1,000U/ml; Uracil-Specific Excision Reagent, New England BioLabs Inc.) is a mixture of UDG and Endonuclease VIII. It was used on purified DNA to construct UDG-treated libraries. For UDG-treated libraries, 5 µl USER® Enzyme was added to 16.25µl purified DNA and incubated in an Eppendorf ThermoMixer for 3 hours at 37°C, prior to blunt end repair.

2.3.2. Library construction

Libraries of purified DNA were constructed based on the protocol of (Meyer and Kircher 2010) with modifications as reported in (Gamba et al. 2014). Control tubes (21.25µl H_2O starting material) were carried out for each library construction.

Blunt end repair was done using NEB Next End Repair Module (New England BioLabs Inc.). For UDG libraries, blunt end repair was performed with 21.25 µl UDG-treated DNA. Non-UDG-treated libraries were constructed using 6 µl of extracted DNA and 15.25 µl H_2O. All library reactions were scaled to 0.7 for a final reaction volume of 70 ul. The reaction mix was incubated at 25°C for 15 minutes, followed by 5°C for 5 minutes. Purification of the reaction mix was done using QIAQuick MinElute purification kit (QIAGEN) according to the manufacturer's instructions and eluted in 20 µl EBT (10 mM Tris-Cl at pH 8.5, 0.02% Tween-20). Adapter ligation was performed on the eluate using T4 DNA ligase, followed by incubation for 30 minutes at 37°C and purification using QIAQuick MinElute purification kit. Finally, adapter fill-in was performed with *Bst* DNA polymerase, large fragments (New England BioLabs Inc.) and incubated for 30 minutes at 37°C. *Bst* activity was arrested through heat inactivation (20 min at 80°C).

2.3.3. Indexing

Single index Polymerase Chain Reactions (PCRs) were performed using AccuprimeTM *Pfx* Supermix (Invitrogen), primer IS4 (10 µM), and a unique indexing primer (5 µM) detailed in (Meyer and Kircher 2010). 3µl of prepared library was added 1µl of a unique P7 index oligo (5 µM) and 21 µl of amplification master mix (20.5 µl AccuPrime *Pfx* Polymerase (Invitrogen), 0.5 µl primer IS4 (10 µM) for a total volume of 25µl. Dual-index Polymerase Chain Reactions (PCRs) were performed identically, except substituting a unique P5 index oligo (10 µM) in place of the primer IS4.

2.3.4. PCR cycling and DNA quantification

PCR mixes were amplified using 5 min at 95°C, a number of cycles x (15 sec at 95°C, 30 sec at 60°C and 30 sec at 68°C) followed by a final extension of 5 min at 68°C. Amplified product was purified using QIAQuick MinElute purification kit, eluting in 10 µl EB. The DNA concentration of amplified libraries was calculated with the Agilent Tapestation 2200 (Agilent) using a DNA 1000 Kit (Agilent) according to the manufacturer's instructions.

2.3.5. Mitochondrial capture

For mitochondrial sequence enrichment, a 'bait and capture' approach was employed using MYcroarray MYbaits® In-Solution Sequence Capture for Targeted High-Throughput Sequencing tool. This kit was designed with custom RNA baits for domestic species. Five PCR

reactions with unique indexes were produced for each sample selected for mitochondrial capture. Samples were pooled according to DNA concentration and endogenous content for a total of 2000ng of DNA.

2.4. BAM file processing

2.4.1. Quality assessment and read trimming

Sequenced reads which passed Illumina quality filters were returned in a fastq.gz format. FastQC was used on all sequencing files to assess basic quality control metrics (Andrews 2010).

Single-end sequencing reads were trimmed using cutadapt 3.5 (Martin 2011) (cutadapt -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -O 1 -m 30). This trimmed any adapter sequences from the read if an overlap of more than 1 bp was found between the read and adapter sequence. All resulting trimmed reads shorter than 30 bp were then removed.

For paired-end sequencing, adapters were trimmed and collapsed using AdapterRemoval v2.3.0 (Schubert, Lindgreen, and Orlando 2016) (AdapterRemoval --file1 reads_1.fq.gz --file2 reads_2.fq --basename output_paired --trimns --trimqualities --minquality 25 --collapse). The --collapse flag merges mate pairs that overlap by at least 11 bp into a single read. Base quality scores of collapsed reads are then recalculated using the metadata of the paired reads. The --trimns --trimqualities --minquality flags cause stretches of consecutive Ns and/or bases of quality lower than 25 from the 5' and 3' termini to be trimmed. The collapsed fastq file was used in downstream analysis while the truncated read pairs were generally not used.

2.4.2. Taxonomic identification

The approximate species identity of each sample was determined using FastQ Screen (Wingett and Andrews 2018). Several common mammalian reference genomes were combined into a dataset and approximately 100,000 reads of trimmed FastQ files were aligned using BWA (Li and Durbin 2009) 0.7.5a to each reference genome in the dataset. Each read is assigned one of five options. A read which aligns only to one genome in one unique genome is assigned as "one hit, one genome" while reads which align to multiple locations in one genome assembly are "multiple hits, one genome". If a read aligns to two or more genomes it will be assigned as "one hit, multiple genomes" or "multiple hits, multiple genomes" depending on if the read aligns uniquely in the genome or in multiple locations. Finally, if there was no alignment, the read is designated as a "No hits".

Species identification using FastQ Screen is not reliable for species that have recently diverged or for determining if a sample is wild or domesticated. For such cases, two alternative strategies were employed. To determine the species identity of bone fragments from closely related taxa, *Bison* versus *Bos*, a hierarchical cluster dendrogram based on pairwise identity-by-state (IBS) distance was used. Sequencing data from five ancient European *Bos primigenius,* five modern European *Bison bonabus* and a Water Buffalo outgroup were aligned to ARS-UCD1.2 as a reference dataset. A screening sample should share more SNPs with the references of the species it belongs to and therefore cluster with those samples. For some *Bos* and *Capra* samples, mitochondrial haplotyping was employed to determine if the sample was likely wild or domesticated.

2.4.3. Read alignment

All trimmed reads of samples were aligned to their reference genomes using BWA (Li and Durbin 2009) 0.7.5a with relaxed parameters (-l 1024 -n 0.01 -o 2). These flags disable seeding, allow for more mismatches with the reference and allow 2 gap openings, respectively.

The resulting sai files were converted to sam file format using BWA samse (Li and Durbin 2009). Paired-end reads were handled using BWA sampe (Li and Durbin 2009) to collapse sai pairs into one sam file. SAM files were then converted into binary bam files using SAMtools view (Li et al. 2009). The -F4 flag was used with SAMtools view to discard unmapped reads for all bam files. For PE reads, the -f2 flag was also used to remove pairs of reads in which the paired reads were aligned to different chromosomes. The bam files were then sorted using SAMtools sort. Read Groups were added to each BAM file using Picard (version 2.22.1) AddOrReplaceReadGroups (https://broadinstitute.github.io/picard/). Read Groups included information on index and library preparation. Picard (version 2.22.1) was then used to remove optical and PCR duplicates using MarkDuplicates (REMOVE_DUPLICATES=true). Indel were locally realigned using RealignerTargetCreator from GATK (v2.4). BAM files were then filtered for reads of mapping quality less than 30 using SAMtools view. Depth of coverage was calculated using Qualimap v2.1.3 (Okonechnikov, Conesa, and García-Alcalde 2016).

2.4.4. Damage patterns

Bam files were assessed using mapDamage2.0 (Jónsson et al. 2013). mapDamage2.0 calculates read length distributions and substitution patterns at the 5' and 3' ends of aligned read bam files. For USER-treated BAM files, PMDTools was also used to calculate

deamination rates at CpG sites only, which cannot be excised by USER treatment (Skoglund et al. 2014) (--deamination --range 30 --CpG).

Depurination levels were inferred using frequencies of purines close to the 5' end of fragment strand-breaks. The use of *T4* DNA polymerase in library preparation removes 3' overhangs, therefore only 5' ends are considered. Using Qualimap v2.1.3 (Okonechnikov, Conesa, and García-Alcalde 2016), nucleotide frequencies were calculated on filtered bam files at each position of DNA fragments. Following the methods detailed in (Sawyer et al. 2012), approximate depurination rates were calculated.

2.4.5. Molecular sex identification

To identify the molecular sex of each sample, the relationship between the number of reads aligned to each chromosome versus the length of that chromosome was investigated. The ratio of reads aligned to the X chromosome was used as an indicator of the molecular sex of a sample, as males should have approximately 50% coverage compared to autosomes. An in-house R script developed by Dr Eppie Jones, Dr Russell McLaughlin and Dr Matthew Teasdale was used to perform a regression analysis to determine the sex of each sample.

2.5. Population Genomic Analysis

2.5.1. Phylogenetic trees

Phylogenetic tree reconstruction is commonly used to describe the relationships among genetic sequences (Yang and Rannala 2012). Within phylogenetic trees, sequences are represented by nodes connected by branches based on their genetic distance. Studying the relationship between each lineage which can provide insight into speciation, distance and demographic changes. In this thesis, Neighbour-Joining (NJ), Maximum Likelihood (ML) and Bayesian inference tree-building methods are used to answer a range of questions. NJ methods are distance-based and use a Markov chain model of a chosen substitution model to compute pairwise sequence distances. A cluster algorithm then successively assigns individuals to clusters based on taxon distance until a tree is fully resolved with good computational efficiency (Saitou and Nei 1987; Gascuel 2006). In contrast, ML and Bayesian methods are character-based, which compare all sequences simultaneously at one character at a time. The final tree is chosen based on a 'tree score' describing maximum parsimony with the least number of changes. Model parameters in ML methods are unknown fixed constants while parameters are described as random variables with statistical distributions in Bayesian analysis. Bayesian methods of tree-building are implemented in BEAST, which uses Markov

Chain Monte Carlo (MCMC) algorithms to estimate the posterior distributions of the parameters (Drummond and Rambaut 2007). BEAST can also account for uncertainty in sample ages in the form of prior probability distributions. This is useful for samples that are dated by context or samples that are beyond the upper limit of C_{14} dating. However, these added priors may risk overparameterisation and reduction in estimation precision (Molak et al. 2013).

These tree-building methods are “model-based”, using nucleotide substitution models to describe the probability of substitutions observed in the DNA sequences. Simplistic models may underestimate aspects of nucleotide evolution while complex models are vulnerable to overfitting the data, similarly compromising phylogenetic inferences (Sullivan and Joyce 2005). Poor model selection can have significant implications for downstream analysis that rely on phylogenetic conclusions. Most methods of site model selection which are pre-determined by likelihood-based frameworks, like jmodeltest2 (Darriba et al. 2012). In Bayesian analysis, pre-determination of site models forgoes incorporating uncertainty into the site model. This can be addressed in BEAST with bModelTest, which infers and marginalises the site model during the MCMC analysis (Bouckaert and Drummond 2017). Another aspect to consider is partitioning of sequencing data into sites that are assumed to evolve under the same model. Different partitioning schemes can greatly affect tree topology, branch length and node support - sometimes resulting in spurious phylogenetic inferences (Kainer and Lanfear 2015).

2.5.2. Genotypes

The identification of genetic variation in the form of single nucleotide polymorphisms (SNPs), insertion/deletions (indels), or structural variants constitutes a starting point of population genomics studies. Algorithms such as mpileup from SAMtools/BCFtools (Li et al. 2009) the UnifiedGenotyper and HaplotypeCaller from the Genome Analysis Tool Kit (GATK) (McKenna et al. 2010) are widely used for diploid genotype calling on high coverage BAM files. Subsequent haplotypic phasing incorporates linkage-disequilibrium information which can provide more power to infer ancestry (Lawson et al. 2012; Leslie et al. 2015; Hellenthal et al. 2014).

However, accurate diploid genotype calling is not feasible with low coverage data due to inherent uncertainty. For example, if a known biallelic site is covered by one read in a low-depth genome, it is impossible to state with certainty if the site is homozygous or heterozygous. Modelling this uncertainty can be addressed using genotype likelihoods, which incorporate a probabilistic framework of genotypes in low-depth genomes (Nielsen et al. 2011). Genotype likelihoods have proven popular in aDNA analysis to account for the low coverage and ancient

damage that is inherent in most ancient genomes (Gamba et al. 2014). Another approach to low-coverage data analysis is the use of pseudo-haploid calls. Pseudo-haploid calls are generated by randomly selecting a single read aligned to each known variable site. This randomly selected read is then duplicated to form a diploid genotype that is homozygous at the site (Jones et al. 2015). While useful for ultra-low genomes, it has been shown that pseudo-haploidising genomes can introduce bias at heterozygous sites, where chosen reads skew toward those containing the reference allele (Günther and Nettelblad 2019).

While the use of genotype likelihoods and pseudo-haploidisation permits the use of low coverage genomes in population genetic studies, the range of downstream analysis remains restricted by missing data. More recently, the use of imputation has been shown to be effective at upscaling low-coverage genotype likelihood/pseudo-haploid calls to accurate, phased diploid calls (Hui et al. 2020). Imputation leverages linkage patterns of known genotypes are used to extend a subset of genotype calls or genotype likelihoods to genome-wide calls (Browning and Browning 2016). This has been implemented in a range of low coverage ancient human genomes that have been called using pseudo-haploidisation or genotype likelihoods (Gamba et al. 2014; Martiniano et al. 2017; Cassidy et al. 2020). However, to date its use has been limited to ancient humans as imputation usually requires a high quality reference panel of phased haplotypes.

2.5.3. Identity-By-State

Genetic differentiation is the extent of divergence between populations caused by mutation, migration, drift and selection. The extent of divergence is useful for detecting structure within population and often expressed in a single measure of genetic distance, like F_{ST} (Meirmans and Hedrick 2011). Another simple distance measure is identity-by-state (IBS), as implemented in ANGSD (Korneliussen, Albrechtsen, and Nielsen 2014). IBS scores pairs of alleles shared by two identical based on if they are identical (0) or not (1). A pairwise distance score is then calculated by averaging this score over all sites shared by the individuals. A matrix of IBS scores can be computed over many samples to detect population structure and infer patterns of shared ancestry.

2.5.4. Principal Component Analysis

Principal Component Analysis (PCA) is a popular exploratory method to summarise the genetic variation of a group of individuals. Foundational population genetic research first used PCA to visualise a few dozen allele frequencies of human populations on geographic maps (Menozzi, Piazza, and Cavalli-Sforza 1978). PCA computation has since been adapted to handle millions

of loci at an individual level (Patterson, Price, and Reich 2006). Its use has become ubiquitous in population genetic studies as an efficient method of detection of structure in a group of individuals. PCA condenses multidimensional data into fewer synthetic variables known as Principal Components (PCs) or eigenvectors. The largest amount of variance, or eigenvalue, is explained by PC1. PCs can be used as the axes of graphical representation to describe individuals in continuous axes rather than classifying them into discrete populations.

While genomic data from ancient samples can be used in the principal component computation, high quality ancient genotype calls are rare. Therefore, it is common practice to "project" the low-depth ancient genomes onto a PC space computed using a reference dataset of high-quality SNP calls from modern genetic variation. Programs like Smartpca and LASER adopt least-squares-based and Procrustes analysis to project samples onto such a reference PC space (Patterson, Price, and Reich 2006; Wang et al. 2014). The uncertainty associated with low coverage ancient data can also be compensated using genotype likelihoods as implemented in PCAngsd (Meisner and Albrechtsen 2018).

2.5.5. ADMIXTURE

Unsupervised model-based clustering methods are commonly used in population genomics to assign an individual a proportion of ancestry to *K* number of genetically distinct source populations. This concept was first introduced in STRUCTURE which uses an iterative Bayesian approach to identify source populations from genetic data and assigns individuals to that population based on the variation observed (Pritchard, Stephens, and Donnelly 2000). Using a maximum likelihood framework, ADMIXTURE calculates population parameters much more rapidly using a fast numerical optimisation algorithm (Alexander, Novembre, and Lange 2009). The results of ADMIXTURE analysis are typically visualised in a barchart form.

The value of *K* is sometimes chosen based on hitherto known aspects of the studied populations. For example, *K=3* is often used in cattle studies based on BovineHD Genotyping BeadChip SNPs as this *K* value is well-characterised in distinguishing between indicine, European taurine and African taurine ancestry (Bovine HapMap Consortium et al. 2009). It is more common in exploratory work to trial a range of potential *K* values to evaluate the most appropriate using a number of likelihood methods. For example, the ADMIXTURE software offers a cross-validation procedure to select an appropriate K. This is achieved by partitioning all genotypes into roughly equally-sized portions. These portions are masked, in turn, and ADMIXTURE attempts to predict the ancestral components for the masked individuals using

the unmasked portions. A “cv error” is then computed by averaging squares of the deviance residuals for the binomial model, with a lower score signifying a better result.

2.5.6. *f* statistics and *D* statistics

f statistics are a popular group of statistics to measure shared genetic drift between two (f_2), three (f_3) or four (f_4) populations to make inferences on shared evolutionary history. First introduced by (Reich et al. 2009) and formalised in (Patterson et al. 2012), *f* statistics have proven to be a versatile and powerful toolkit to answer a range of phylogenetic questions by representing genetic drift as branch length. The shared genetic drift between two populations (P_1 and P_2) may therefore be represented by the shared branch length, $f_2(P_1,P_2)$. The three population test, $f_3(P_1,P_2; P_3)$, measures the branch length of P_3 and the internal node of all populations. This test is widely used to measure the amount of shared drift between P_1 and P_2 in outgroup f_3 analysis, where P_3 is an outgroup common to P_1 and P_2. Finally, f_4 corresponds to the internal branch length between the two bifurcating populations: $f_4(P_1,P_2; P_3,P_4)$. Applications of *f* statistics include testing treeness of a group of populations, detecting the number of founder populations (Lazaridis et al. 2014), and detecting and quantifying admixture proportions from ancestral populations (Haak et al. 2015).

A variation of the f_4 test known as Patterson’s *D statistic,* or the ABBA/BABA test, is a widely implemented statistic that can test the goodness of a four-genome tree. This statistic was first developed to detect and quantify the genetic contribution of Neanderthals to modern non-African human populations (Green et al. 2010). It was later extended and formalised to be robust in a variety of demographic scenarios for admixture in populations or closely related species (Durand et al. 2011). Given four taxa (P_1, P_2, P_3, P_4), the test considers the ancestral and derived alleles at a number of biallelic sites. At sites where P_3 and P_4 have different alleles, the outgroup genome (P_4) is used to define the ancestral allele (“A”) while the ingroup genome P_3 defines the derived allele (“B”). The number of discordant sites of P_1 and P_2 are then used to calculate the *D statistic*. The number of sites that P_1 has the “A” allele and P_2 has the “B” allele are denoted as nABBA, while the number of sites that H1 has the “B” allele and P_2 has the “A” allele are denoted as nBABA. The *D statistic* is then calculated by the following equation:

$$D = \frac{nABBA - nBABA}{nABBA + nBABA}$$

The null hypothesis is that the four-genome tree is correct and no introgression has occurred. Under this scenario, nABBA ≈ nBABA, resulting in a *D statistic* that is statistically indistinguishable from 0. If introgressive hybridisation occurred between P_1 or P_2 and P_3, an excess of ABBA or BABA sites between two taxa occurs. This results in either a positive or negative D statistic, implying gene flow between these branches. Alternatively, an excess of ABBA/BABA sites can occur simply due to the incorrect topology of the tree. The significance of the result can be assessed using a calculated Z-score.

2.5.7. TreeMix

TreeMix is a graph-based algorithm that describes the relationship between populations using allele-frequency data (Pickrell and Pritchard 2012). TreeMix constructs a ML tree of all populations and models admixture between branches or tips in the form of migration edges. With a view to define how well the model fits the data, TreeMix calculates the residual covariance between each pair of populations divided by the average standard error. Positive residuals in a pair indicate that the model underestimates the observed covariance which suggests while negative residuals suggest that the model overestimates the observed covariance in a population pair. Migration edges are added iteratively to explain the largest residuals of the tree, thus increasing the model likelihood.

2.5.8. Runs of Homozygosity

Runs of homozygosity (ROH) are large stretches of a genome with little or no heterozygous sites due to the inheritance of pairs of haplotypes that are identical by descent (IBD). The patterns of ROHs can reveal demographic events hidden in a genome (Ceballos et al. 2018). The length of individual ROHs can indicate when and how the inbreeding events occurred due to recombination reducing the length of these regions over generations. ROHs are of particular interest to aDNA studies as they can provide insight into species decline and extinction, the past cultural human practices and the domestication process (Daly et al. 2021; Cassidy et al. 2020; Palkopoulou et al. 2015). Detecting ROHs in ancient genomes generally requires high-quality genotype calls due to the potential of false heterozygosity introduced by ancient damage. Alternatively, ROHan provides accurate ROH detection in genomes as low as 5X using a Bayesian Hidden Markov Model (HMM) that estimates both heterozygosity and ROH (Renaud et al. 2019). By sampling the minimum and maximum values of three MCMC chains (each time using a different estimate for heterozygosity), ROHan also provides confidence intervals of ROH estimates.

2.5.9. Sequentially Markovian coalescent methods

Orthologous DNA sequences are related by a branching structure known as a genealogy. Generating genealogical trees as a Markov process was first presented by (McVean and Cardin 2005). Here, the authors simplified modelling the coalescence time of many genealogical trees across the genome using a hidden Markov model and leveraging the distribution to deduce demography. This concept of genealogical tree as a Markovian process was later implemented in PSMC (Pairwise Sequential Markovian Coalescent) to infer the demographic history of a population using a single diploid genome (Li and Durbin 2011). In a given diploid genome, there may be millions of independently evolving biallelic loci. Using a HMM, PSMC infers the most recent common ancestor (TMRCA) of different segments of the genome by estimating the coalescence time of the two alleles. It constructs a TMRCA distribution across the genome which it uses to deduce the demographic history of the population since Ne is inversely proportional to the rate of coalescence. For example, a smaller *Ne* is inferred at a time when many loci coalesce and vice versa. The density of heterozygosity in stretches of DNA corresponds to the age of the coalescent event.

PSMC was later extended to account for multiple sequences in Multiple Sequentially Markovian Coalescent (MSMC and MSMC2) (Schiffels and Wang 2020). The incorporation of multiple sequences greatly increased the power to infer more recent demographic events. In addition to a demographic history PSMC/MSMC has also been used to detect population divergences. Cessation of gene flow may be inferred by determining the point at which PSMC curves diverge in trajectory. While a wealth of information is available by using these models, some caution is necessary as it has been shown that a structured population can result in false inflation of effective population size (*Ne*) (Mazet et al. 2016).

2.5.10. Metagenomic analysis

Metagenomics is the sequencing and characterisation of all genetic material recovered from environments. While aDNA studies usually target the endogenous content of archaeological material, there is a growing awareness of value in the exogenous aDNA which may detect ancient pathogens and diets and characterise microbiome evolution (Fellows Yates et al. 2021). Metagenomic studies of modern data often employ metabarcoding strategies to sequence universal genes to capture that can be taxonomically assigned and analysed without much computational effort (Chua et al. 2021). However, the amplicons of metabarcoding strategies are usually too large for use in aDNA studies and so ancient metagenomics analysis is usually restricted to the use of shotgun sequencing reads (Harbert 2018). This can result in a substantial, and sometimes intractable, computation burden which requires careful selection

of appropriate databases and taxonomic assigners. Taxonomic assignment of shotgun metagenomic reads can be performed using alignment-based search tools (Nayfach et al. 2016; Altschul et al. 1990), using Burrows-Wheeler transform mapping algorithms (Kim et al. 2016), or ultrafast k-mer pseudoalignment approaches (Wood, Lu, and Langmead 2019).

Most microbes coexist with each other in complex communities known as microbiomes that can be characterised to shed light on aspects of health, evolution and diet (Davenport et al. 2017). A popular microbial source tracking approach is SourceTracker, which uses Bayesian statistics and Gibbs sampling to estimate the proportion of sequencing reads of samples (sinks) originating from selected metagenomes (sources) (Knights et al. 2011).

3. Preparation, Screening and Sequencing of Ancient and Historic Materials

3.1. Introduction

The genomics revolution ushered in an era of unprecedented data generation within the field of aDNA. In the last decade, NGS technologies and various laboratory innovations have transformed the field into a high-throughput data science. In what is a seismic shift in power, aDNA studies moved from PCR-based analysis of a select loci to whole genome population analysis (Sirak, Sawchuk, and Prendergast 2022; Poinar et al. 2006). Today, high quality genomes are being published at a considerable rate and it is not uncommon for a single aDNA paper to publish hundreds of newly sequenced ancient individuals (Librado et al. 2021; Allentoft et al. 2022). Each year, the temporal limits of aDNA are wound further back through application of these methods (van der Valk et al. 2021). These advances have lent novel clarity to topics of migration, domestication, extinction and our own history as a species. Despite the last decade of advancements to aDNA analysis, obstacles remain that have challenged researchers since the foundation of the field. Perhaps the largest is the low endogenous DNA content associated with aDNA (Damgaard et al. 2015). Targeted sampling and innovations in aDNA extraction and purification methods have improved success rates. However, successful studies are rare in arid climates unconducive to aDNA preservation despite extensive sampling (Sirak, Sawchuk, and Prendergast 2022). Unsuccessful samples are rarely reported which may create an inflated rate of success in aDNA. Even in more favourable climates, for every published sample there are likely several with unusable levels of endogenous DNA.

It is standard practice to screen ancient material for endogenous DNA content prior to investing extensive resources into deep sequencing. While screening rarely results in sufficient coverage for downstream population genetic analysis, it does provide enough information to calculate the endogenous DNA content of the sample. Screening is also an important step to validate the authenticity of a sample before investing in deeper sequencing as contamination estimates and age-related damage patterns may be generated from only a few thousand reads.

3.2. Aims

1. To extract DNA from ancient animal material and create DNA libraries suitable for Next Generation Sequencing.
2. To sequence DNA libraries and taxonomically identify samples and estimate endogenous DNA content.
3. To investigate aDNA damage patterns and molecular sex.
4. To inform a high-depth sequencing strategy for successful samples.

3.3. Materials and methods

3.3.1. Materials

In total, 285 historical and ancient samples were screened for endogenous DNA content (Table 3.2). Of these samples, 186 were processed from original material provided to Trinity College Dublin. An additional 42 of the bone and teeth samples were extracted from pre-drilled bone powder and 30 historical samples were processed from pre-extracted DNA. 27 *Bovini* samples were bioinformatically screened from sequencing data provided by Lund University.

The majority of samples were mineralised tissue (246/285) with soft tissue samples (10/285) and coprolites (22/285) making up the remainder.

3.3.2. Laboratory methods

Photography, sample preparation and drilling are detailed in Section 2.1. 3D scanning was done using an EinScan Pro HD. DNA extraction and sequencing library construction are detailed in Section 2.2 and 2.3, respectively.

DNA concentration was quantified using an Agilent 2200 TapeStation (Agilent Technologies) and/or an Invitrogen Qubit Fluorometer (Thermo Fisher Scientific). Libraries were amplified with 12 PCR cycles as standard. If overamplification of the libraries was detected on the TapeStation trace, PCR libraries were redone with lowered PCR cycling. In cases with insufficient DNA concentration, PCR libraries were also redone with raised PCR cycles.

3.3.3. Low depth sequencing screening strategy

Samples were screened for approximate DNA concentration by generating sequencing reads on Illumina NGS sequencing platforms (Illumina). As standard, at least 100,000 sequencing

reads were generated to provide an adequate number of reads for taxonomic identification and endogenous content percentage.

172 of the 285 samples were screened on Illumina's MiSeq platform (Appendix Table 3.1). Typically, dozens of uniquely indexed samples were multiplexed and screened together in one sequencing run to maximise economic efficiency. 10ng of each DNA library were pooled, diluted, and sequenced together. The libraries were sequenced with 1% PhiX control at 1x63 bp at TrinSeq, the Trinity Genome Sequencing Laboratory, Trinity College Dublin. 82 samples were screened on an S1 or S4 Flow Cell using Illumina's NovaSeq 6000 platform. 2-4% of a sequencing lane was dedicated to screening samples, resulting in approximately one million reads per sample. Each sample was dual-indexed to mitigate the higher risk of index-jumping. The libraries were sequenced at 2x50 bp and 2x100 bp at TrinSeq, the Trinity Genome Sequencing Laboratory, Trinity College Dublin.

3.3.4. Read Processing

Samples were demultiplexed using bcl2fastq and the sequencing reads were returned in fastq.gz format. The quality of the sequencing reads was assessed with FastQC and trimmed according to Section 2.4. Only the collapsed reads were used, unless otherwise stated.

3.3.5. Taxonomic identification

Samples were taxonomically identified using FastQ Screen, as detailed in Section 2.4.2. (Wingett and Andrews 2018). The samples were screened against a database of eight genome assemblies representing eight genera (Table 3.1). For Bovini samples, taxon identity was determined by computing an identity-by-state matrix with high quality genomes of five aurochs, five wisent and an outgroup using ANGSD 0.930 (Korneliussen, Albrechtsen, and Nielsen 2014).

Table 3.1 - The genome assemblies used in FastQ Screen.

Genus	Representative species	Genome build used
Alces	*Alces alces*	GSC_moose_1.0
Bos	*Bos taurus*	ARS-UCD1.3
Canis	*Canis familiaris*	CanFam3.1
Capra	*Capra hircus*	ARS1.2
Equus	*Equus caballus*	EquCab3.0
Homo	*Homo sapiens*	GRCh38.p12

Ovis	*Ovis aries*	Oar_v4.0
Sus	*Sus scrofa*	Sscrofa11.1

3.3.5. Read alignment

After taxonomic identification, sequencing reads were aligned to the representative genome in Table 3.1 using BWA 0.7.5 (Li and Durbin 2009) with relaxed parameters (Section 2.4.3.). For samples that could not be genetically identified due to poor DNA preservation, sequencing reads were aligned to the genome corresponding to the archaeological taxonomic ID.

3.3.6. Endogenous content quantification

The endogenous DNA content was calculated for each sample using the following formula:

$$\%\ \boldsymbol{Endogenous\ content} = \frac{\boldsymbol{total\ number\ of\ aligned\ QC\ reads} \times \mathbf{100}}{\boldsymbol{Total\ number\ of\ trimmed\ reads}}$$

3.3.7. Molecular sex identification

An in-house R script was used to assign molecular sex to samples with a sufficient number of aligned reads (Section 2.4.5.). The script was modified to account for the karyotype of each genome assembly.

3.3.8. Damage assessment

Age-related damage patterns were computed using mapDamage2.0 and PMDtools (Section 2.4.4.). Read length distributions and median read lengths were calculated using an in-house script.

3.3.9. Deep sequencing strategy

After determining endogenous content, a sequencing plan was developed for each sample. The deep sequencing strategy was designed based on the desired genome coverage of each sample. The number of reads for each sample was calculated using the equation below.

$$\boldsymbol{reads\ required} = \frac{\boldsymbol{desired\ genome\ coverage} \times \%\ \boldsymbol{endogenous\ content} \times \boldsymbol{average\ endogenous\ read\ length}}{\boldsymbol{Total\ nucleotides\ of\ reference\ genome} \times \mathbf{100}}$$

For some samples, the endogenous DNA content was too low for high depth sequencing. Some relevant low endogenous *Bovini* samples were included for mitochondrial capture (discussed in Chapter 7).

3.4. Results

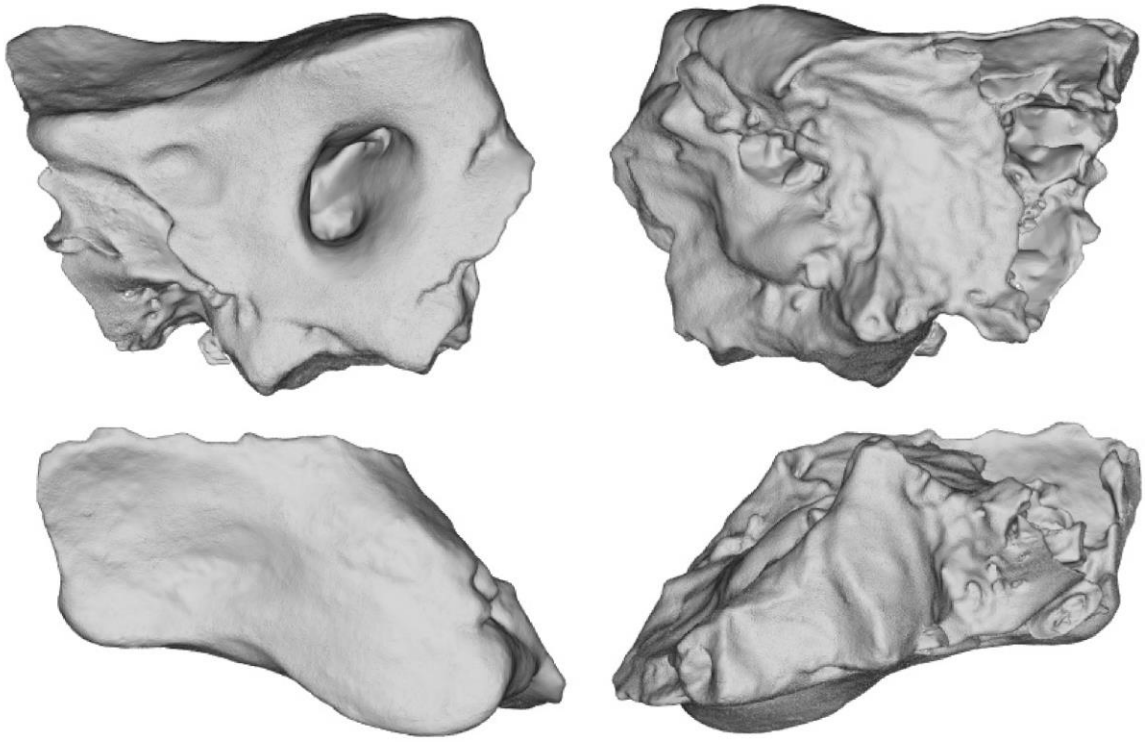


Figure 3.1 - A projected 3D scan of an aurochs petrous bone using EinScan Pro HD. This scan can be used for 3D printing and downstream morphometric analysis.

3.4.1. Taxonomic identification

223 of the 285 samples (~78%) screened produced a sufficient number of reads and endogenous DNA content to determine genus identity. For 215 higher endogenous samples, FastQ Screen was sufficient to determine the sample at the genus level (Figure 3.2). However, FastQ Screen may lack resolution for lower endogenous samples, especially for Genera that are closely related (i.e. *Ovis* and *Capra*). In cases when the FastQ Screen was not conclusive, samples taxonomy was designated N/A.

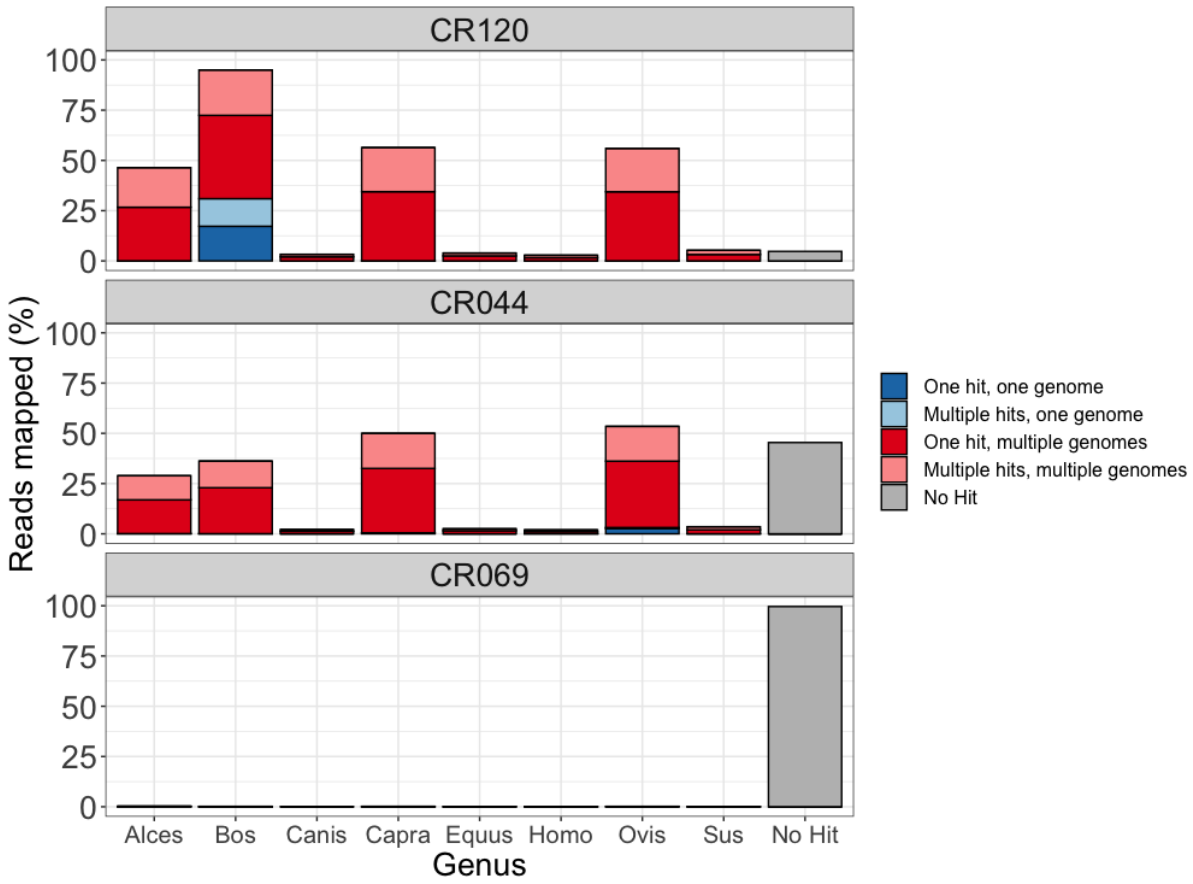


Figure 3.2 - FastQ Screen results displaying a representative selection of samples of varying endogenous DNA preservation. (A) Displays the resulting FastQ Screen for sample CR120, a Mesolithic aurochs sample with high endogenous DNA content. (B) displays the results for CR044, a Neolithic domestic sheep sample. (C) displays the results for a sample that was unidentifiable. While the archaeological taxonomic ID was that of Capra, more than 99.5% of reads did not align to any of the mammalian genomes in the screening database.

Pairwise identity-by-state (IBS) distances using very low coverage genomes were used to determine the likely Genus ID for 8 ambiguous *Bovini* bone fragments. As few as 600 endogenous reads were sufficient to cluster samples with the correct species (Figure 3.3). These results were later corroborated by mitochondrial haplotyping (Chapter 7).

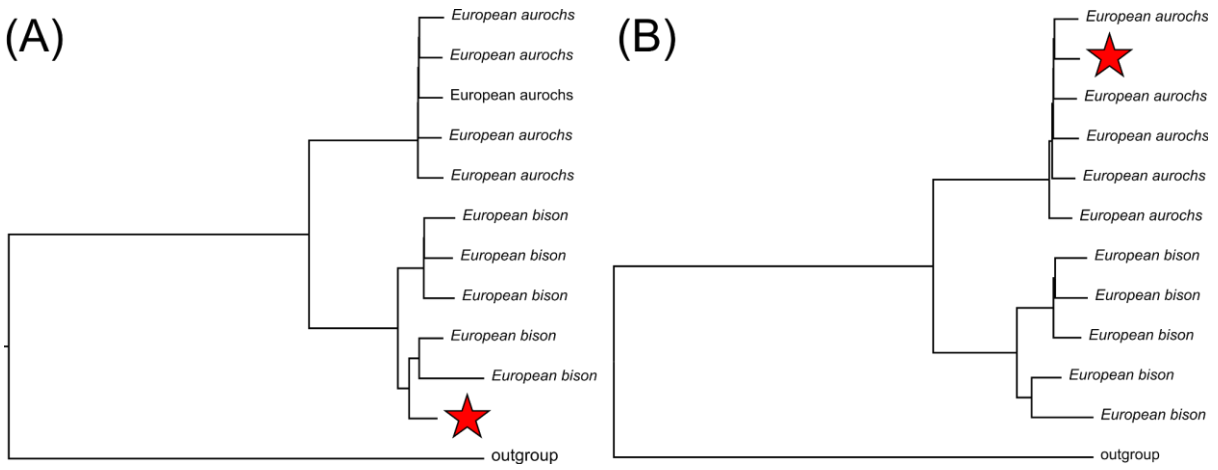


Figure 3.3 - Pairwise identity-by-state Neighbour-Joining trees to test the likely taxonomic identity of low-coverage *Bovini* samples. The trees were constructed with 5 high coverage European aurochs genomes and 5 high coverage European bison genomes, with a Water Buffalo used as an outgroup. (A) displays ER007 clustering within the European bison clade while (B) displays ER005 clustering within the European aurochs clade.

In total, samples from 8 different genera were identified (Figure 3.4). Of the 223 identified samples, 8 (~3.6%) samples had a genetic identification that differed from the archaeological identification. Additionally, there was taxonomic identification for 70 samples with an absent or ambiguous archaeological assignment (i.e. "*Caprinae*", "*Bovine*", or "small ruminant").

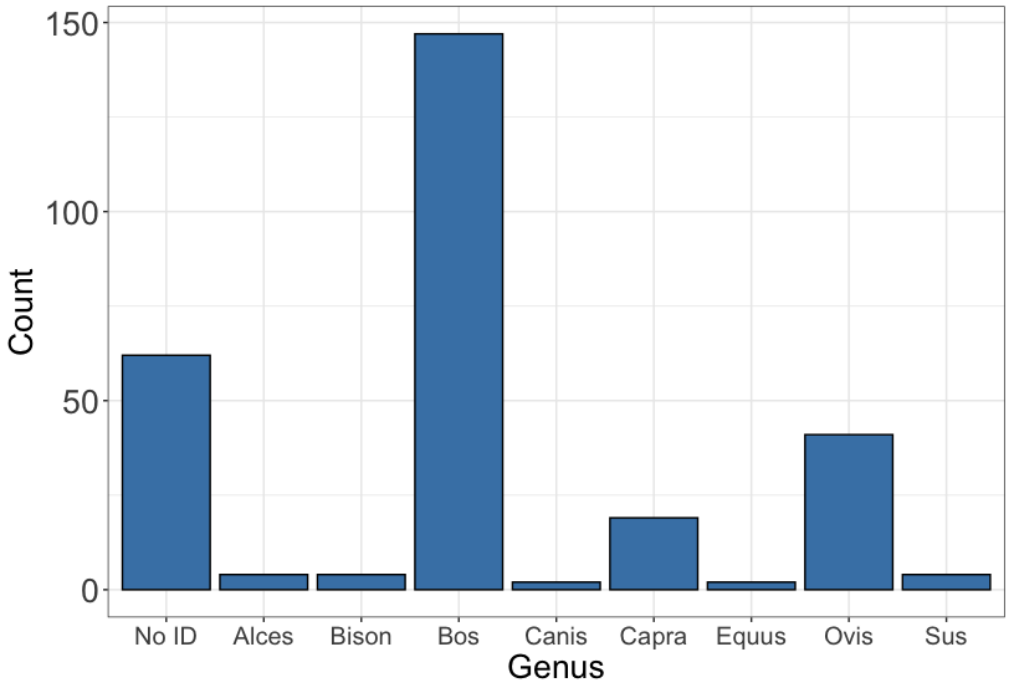


Figure 3.4 - Summary of Genus identification using FastQ Screen and Pairwise IBS NJ trees.

3.4.2. Endogenous DNA content

Endogenous DNA contents are displayed in Figure 3.5 and summarised in Table 3.2. Extended information regarding sample preparation, sequencing platform and alignment are detailed in Appendix Table 3.1.

Of the ancient mineralised materials, the petrous bone yielded the highest levels of endogenous DNA content (mean of 25.28%). This was followed by tooth samples, which had an average of 6.62% endogenous DNA. While the Metatarsal/Phalanx samples yielded the lowest average endogenous content (0.4%), non-petrous and non-tooth ancient samples generally yielded very low endogenous DNA levels. 60 of the 81 (74.07%) ancient non-petrous bone samples yielded endogenous DNA content of <1%, while only 5 yielded above 5% endogenous DNA. On average, historical soft and hard tissue performed similarly; means of 16.03% and 16.89%, respectively. The preservation of coprolites is discussed in greater detail in Chapter 5.

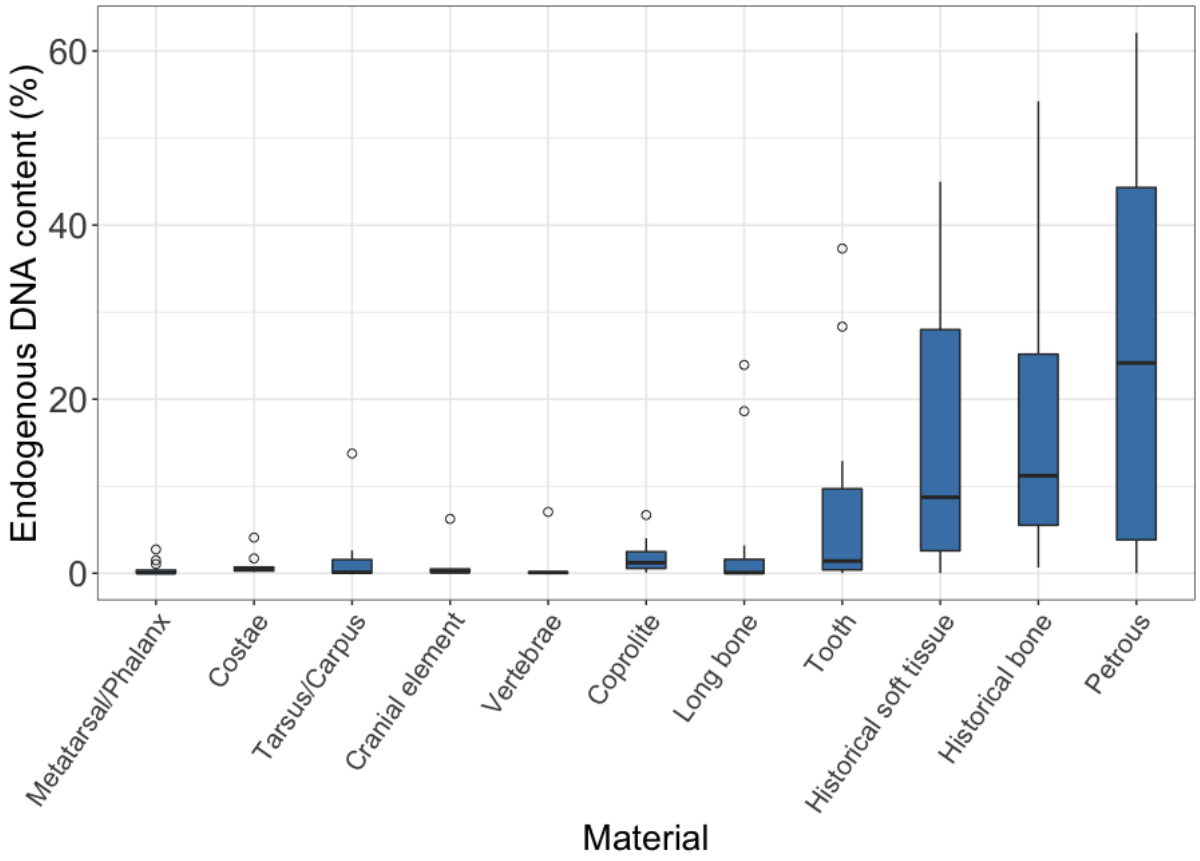


Figure 3.5 - Endogenous DNA content per material type. Endogenous content detailed in Table 3.2. Only samples with at least 10,000 sequenced reads are included as endogenous content can be inflated in samples with low read totals. Midline bars denote the median.

The relationship between endogenous content yield and factors associated with aDNA decay were investigated using only data from petrous samples (n=117). The correlation between endogenous DNA content and latitude and age were both non-significant (p = 0.99 and 0.95, respectively) (Figure 3.6 and Figure 3.7).

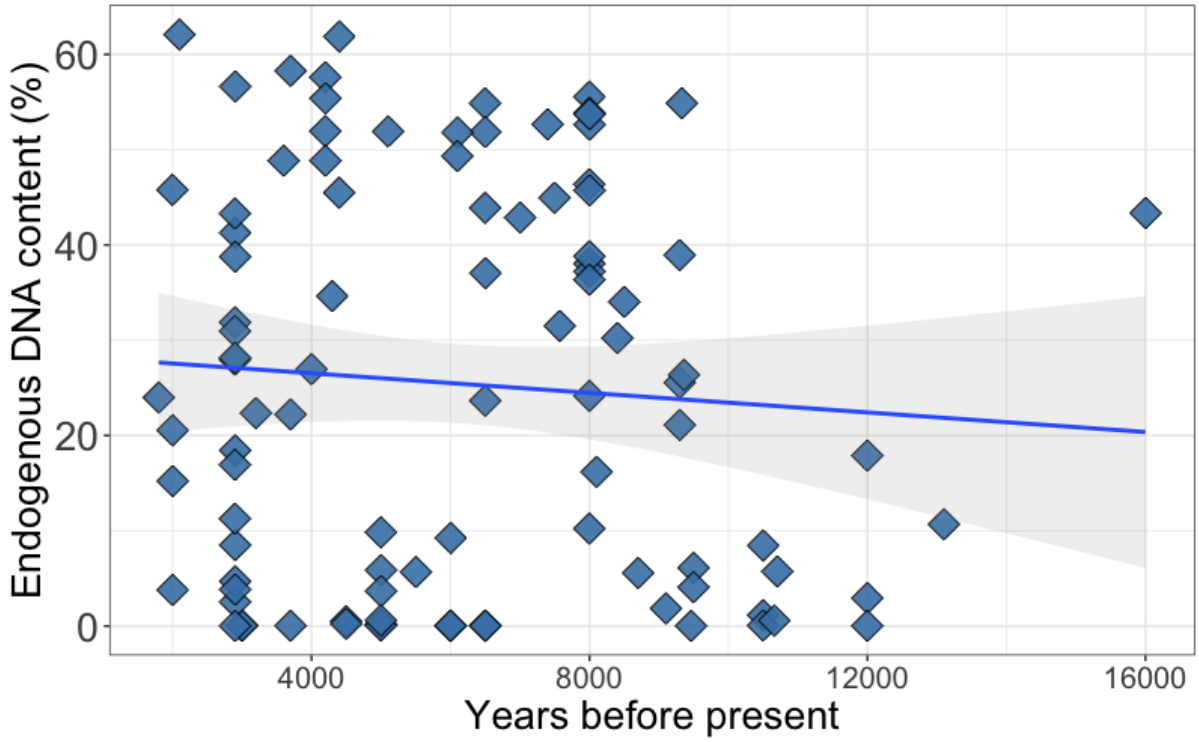


Figure 3.6 - Endogenous DNA content of 99 petrous samples with associated contextual or C14 dating.

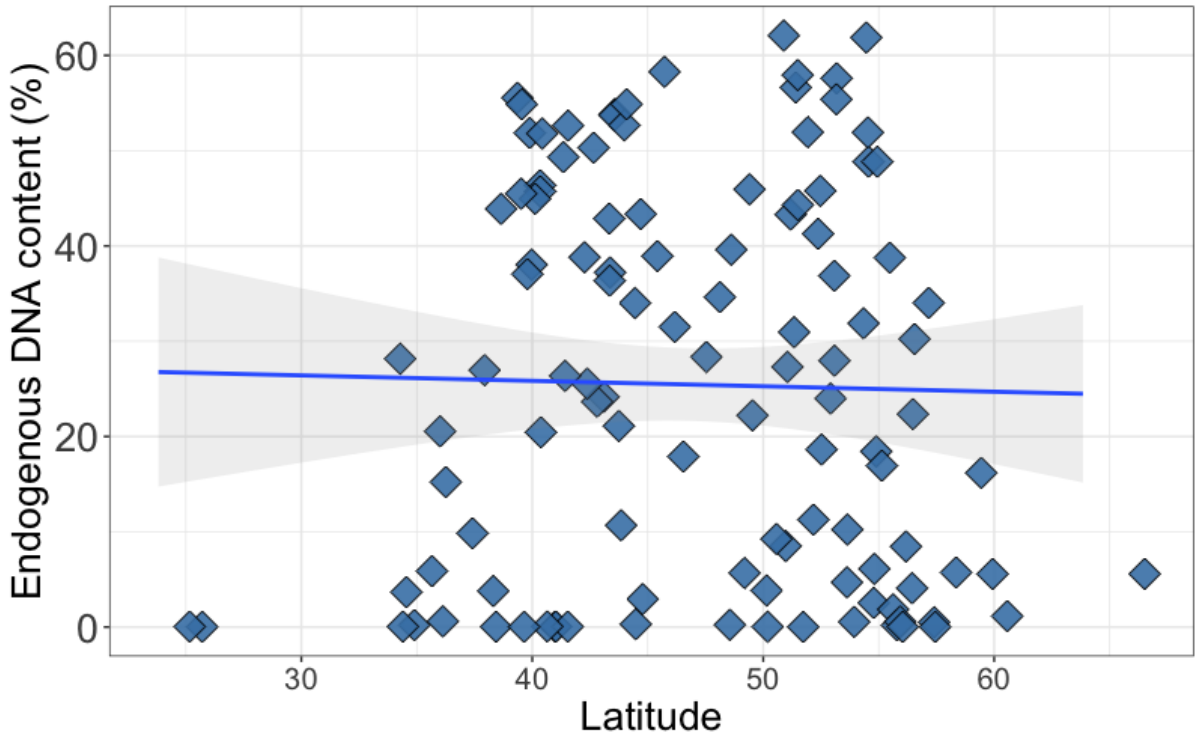


Figure 3.7 - Endogenous DNA content of 114 petrous samples with known deposition.

3.4.3. Extraction methods

DNA from 7 previously drilled samples were extracted in an attempt to improve endogenous content. The samples were originally drilled 2-4 years prior to the re-extraction and stored as bone powder at 4°C.

Four samples were re-extracted using the “classic” extraction method while three samples were re-extracted using the bleach extraction method. The endogenous DNA content decreased in three of the four samples re-extracted using the classic extraction. Meanwhile, endogenous DNA content improved substantially in two of the three samples extracted using the bleach extraction method, while the remaining sample (MV240) yielded almost identical endogenous content.

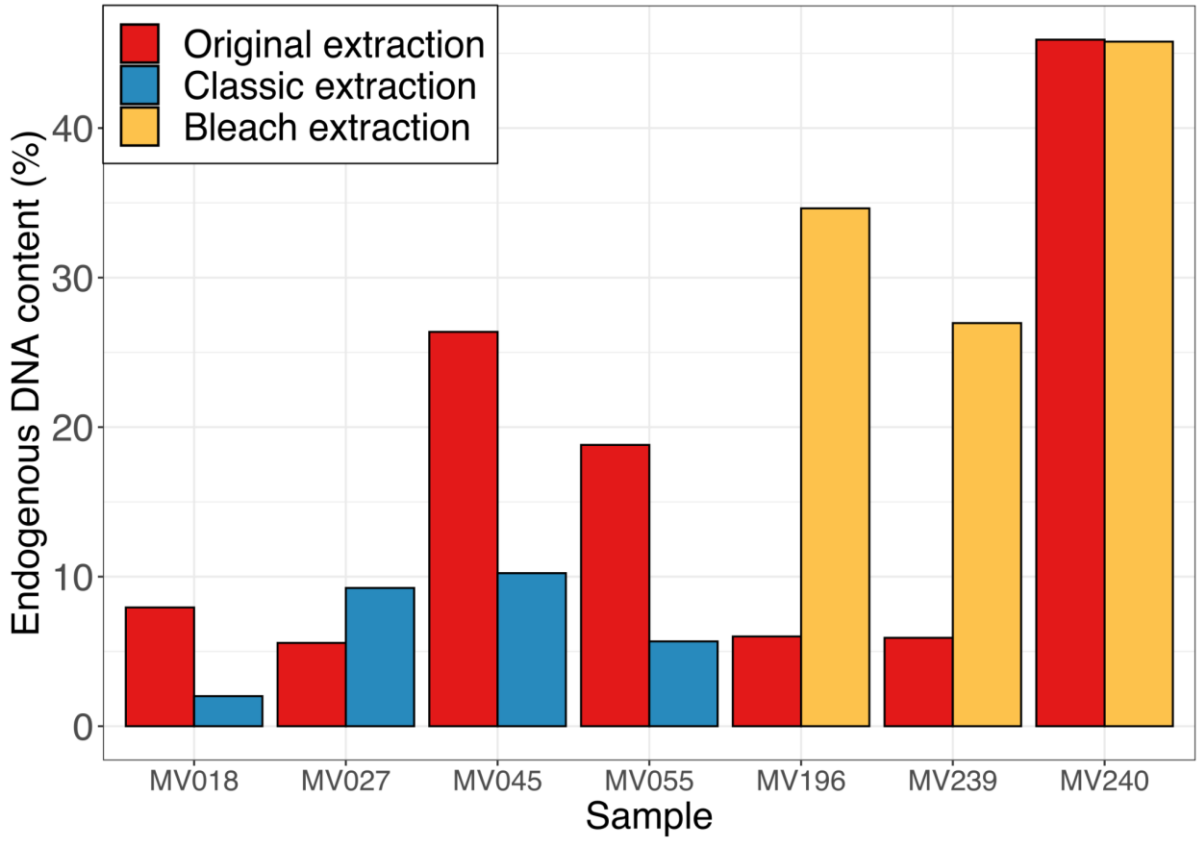


Figure 3.8 - Endogenous DNA content for the original extraction and the re-extraction from bone powder for 7 samples. The original extractions were done in Trinity College Dublin between the years of 2016 and 2018.

Separately, both a classic extraction and a bleach extraction were performed on the bone powder from three samples. For all samples, higher endogenous DNA content was recovered using the bleach extraction (Figure 3.9A). In one case the endogenous DNA content increased by ~4.7 times. With the increase in endogenous content, the DNA concentration also decreased substantially in each sample (Figure 3.9B).

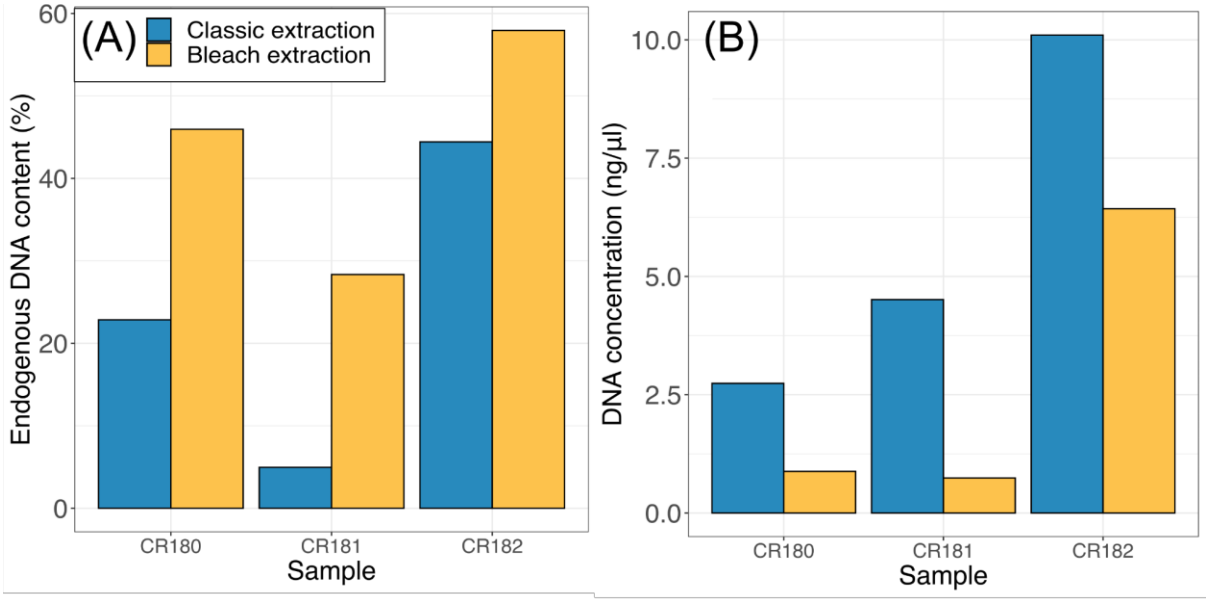


Figure 3.9 - (A) The endogenous DNA content of three samples when extracted using the classic extraction method and the bleach extraction method. (B) The DNA concentration of each sample when extracted using the classic extraction method and the bleach extraction method. In each case, the DNA yield fell in concentration.

3.4.4. Sexing

Sexing of samples was successful for 166 of the 223 genetically identified samples.

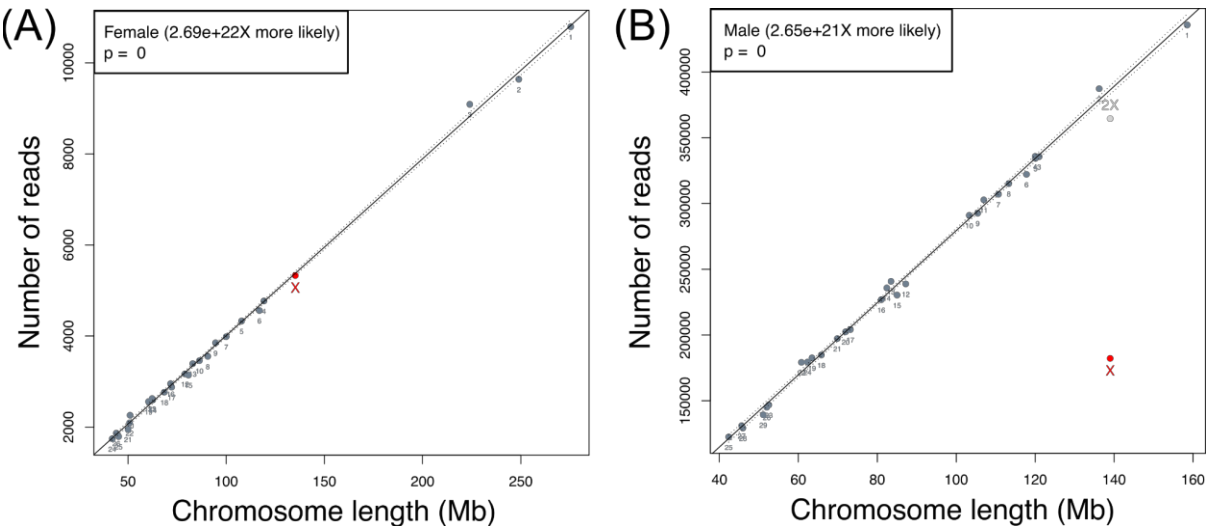


Figure 3.10 - Example sexing plots for a (A) female *Bos* sample and (B) a male *Capra* sample.

In total 78 female and 88 male animals were sexed, suggesting a balanced sex ratio. However, when the animals were subdivided into “wild” and “domestic” categories, a sex bias emerged. Of the 80 wild animals, ~71% (57) were male. Conversely, ~64% (55) of domestic animals were determined to be female.

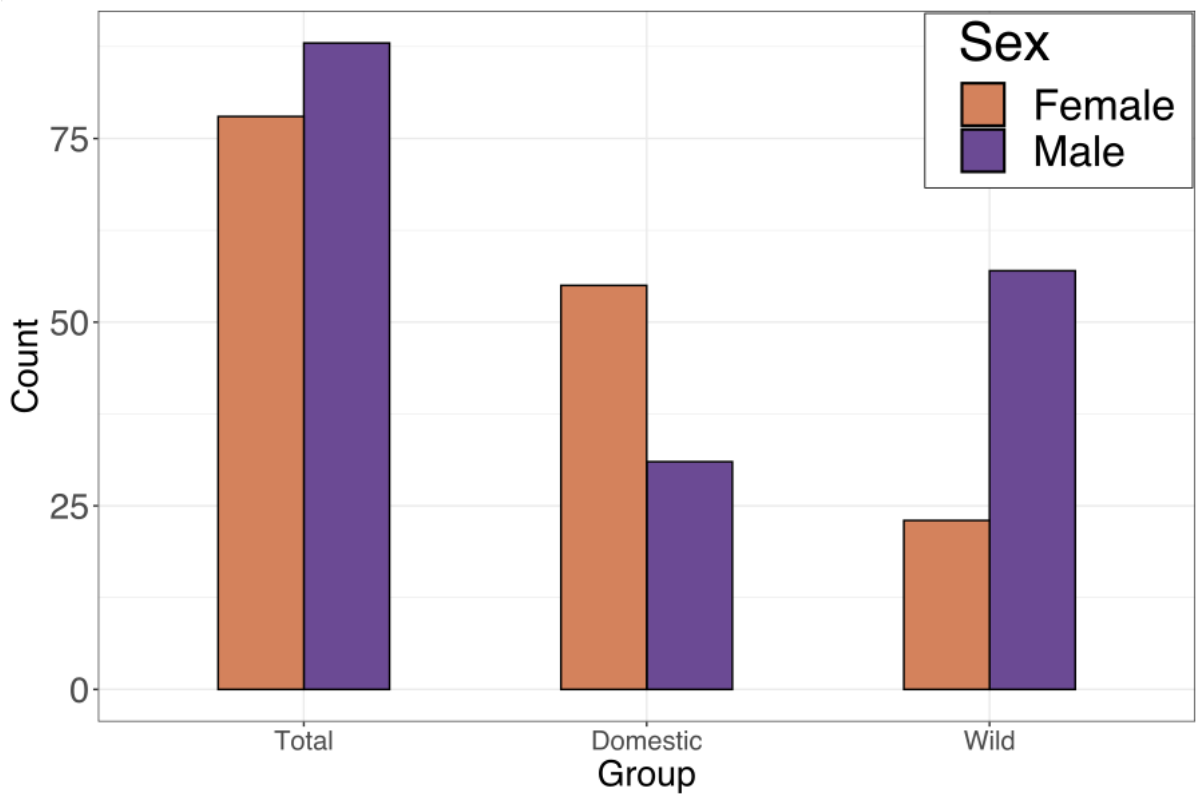


Figure 3.11 - Sex ratios of all animals screened and based on domestic or wild.

3.4.5. DNA damage patterns

Deamination patterns associated with aDNA were assessed using mapDamage and PMDTools. An example of the deamination patterns of a non-USER and a USER-treated library is shown in Figure 3.12. This figure displays how deamination patterns at CpG sites are retained post-USER treatment.

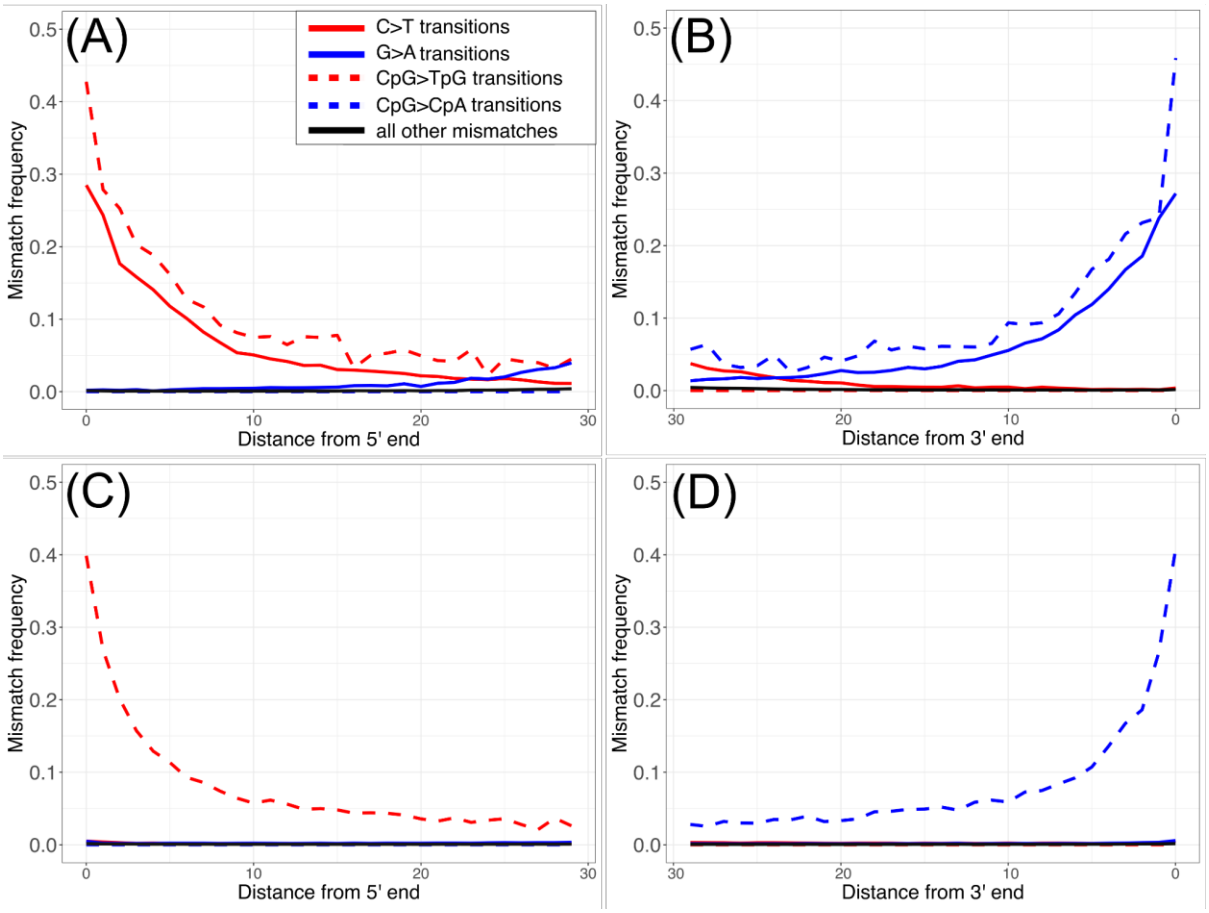


Figure 3.12 - Deamination and mismatch patterns in CR120, a mesolithic aurochs, pre- and post-USER treatment computed by PMDTools. (A) Mismatches prior to the 5' end in a non-USER-treated DNA library. C>T transitions reach a maximum at the base just prior to the 5' end at a frequency of approximately ~0.3. C>T transitions at CpG sites also peak here at a frequency of over 0.4. G>A transitions and all other mismatches are in very low frequency. G>A transitions begin to rise after the 20bp mark. (B) Mismatches prior to the 3' end in a non-USER-treated DNA library. Here, C>T transitions are low frequency while G>A reaches a maximum of ~0.3 in all sites and over 0.4 in CpG sites. (C) Mismatches prior to the 5' end in a USER-treated DNA library. Here, C>T transitions are drastically reduced compared to the non-USER-treated DNA library but C>T transitions remain at CpG sites. (D) Mismatches prior to the 3' end in a USER-treated DNA library. The G>A transitions are once again drastically reduced while they remain in high frequency at CpG sites.

A selection of deamination patterns of higher coverage aurochs genomes is displayed in Figure 3.13. Generally, USER treatment was highly efficient at removing C>T transitions at the 5' end of sequenced reads and the corresponding G>A transitions at the 3' ends. However some samples retained low levels of deamination at read ends even post-USER treatment e.g. Bar1-CR119 (Figure 3.13).

Median aligned read lengths were calculated for each sample and are reported in Table 3.2. The shortest median read length reported was in the petrous bone sample Rhi8-CR187 and in historical skin sample CR193. The first is among the oldest samples screened during this PhD and the second the most recent. The longest median read length for a sample screened in this dataset was from a Neolithic cattle tooth from Denmark (62bp). Median read length is non-significantly positively correlated with endogenous DNA content (R^2= 0.0486, p= 0.479436) and negatively correlated with age (R^2= 0.0760, p= 0.311944).

A selection of read length distributions are displayed in Figure 3.14. These samples have a peak frequency in the range of 30-60 bp except for Fal7, which is a relatively young (~1,000 years old) non-USER-treated aurochs from a site in Germany. A mesolithic Spanish sample Gal2-CR170 displays peak periodicity in the frequency of observed reads. This periodicity of around 10 bases is thought to be related to the length of DNA wrapping of a single histone.

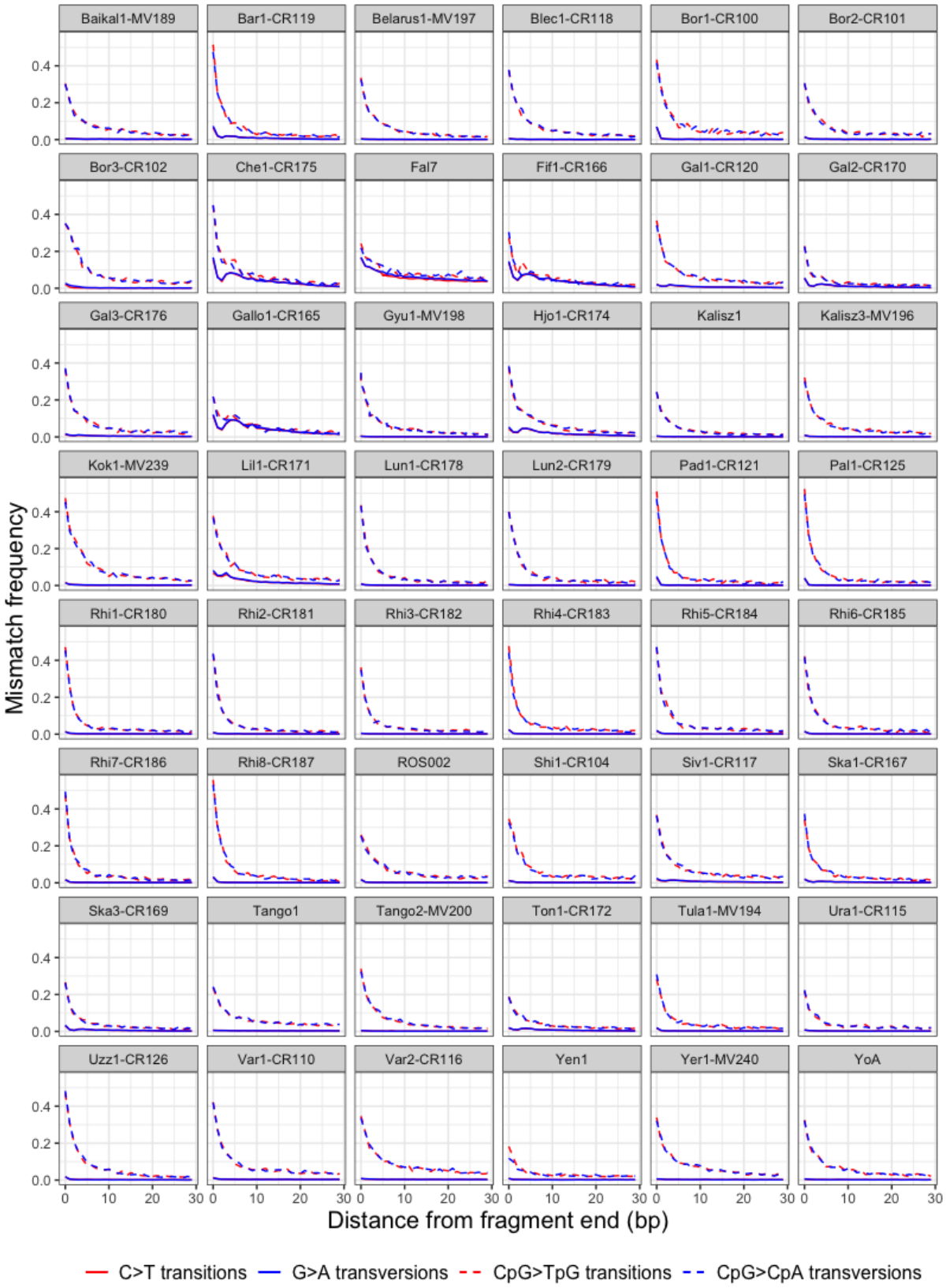


Figure 3.13 - Deamination patterns for a selection of higher-coverage aurochs genomes. All samples consist solely of USER-treated libraries except for Fal7, which is non-USER-treated and Che1-CR175, Fif1-CR176, Gallo1-CR165, Hjo1-CR174, and Lil1-CR171 which are merged BAM files of both USER-treated and non-USER-treated libraries.

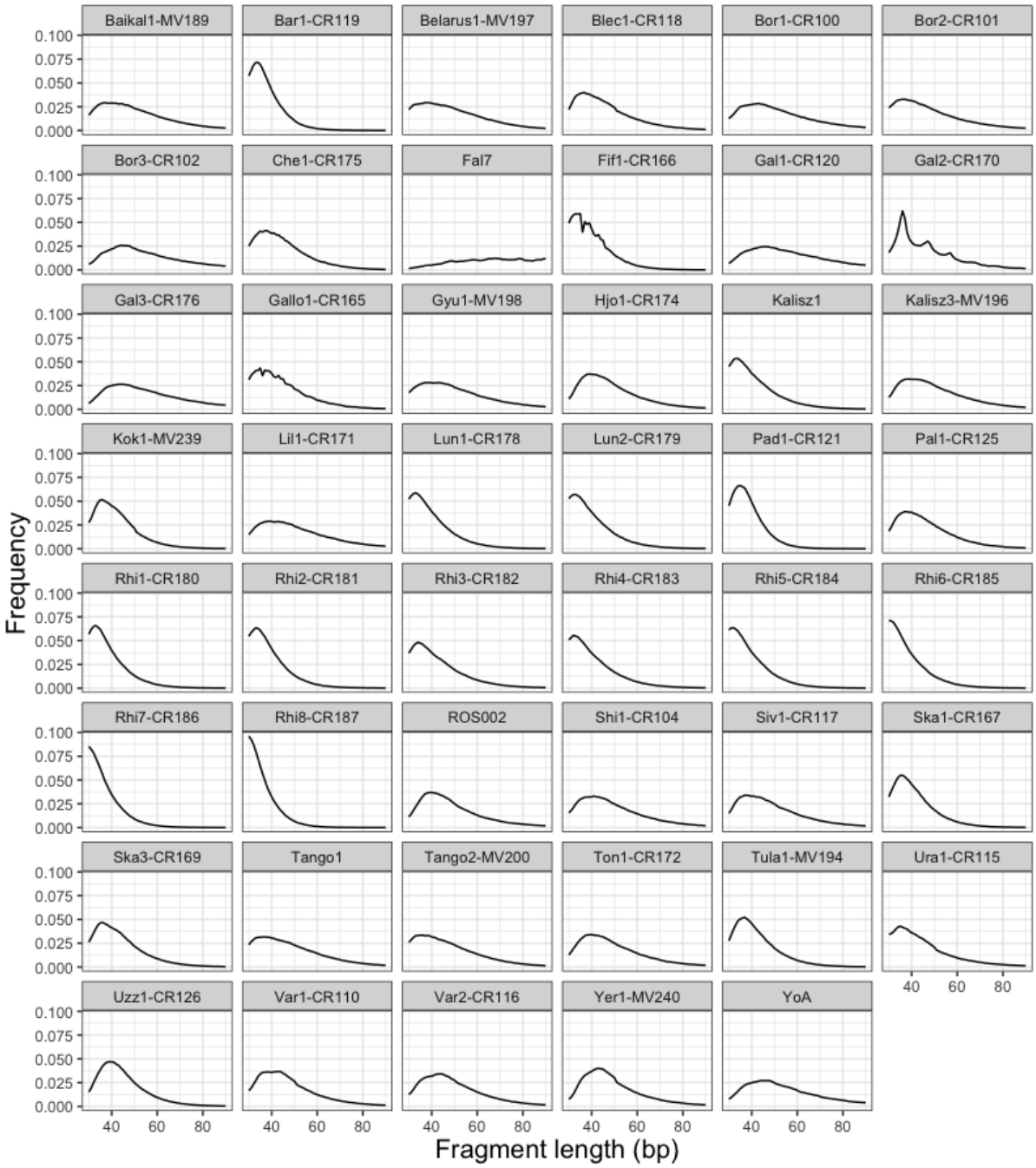


Figure 3.14 - Read length distribution for a selection of higher-coverage aurochs genomes.

3.4.6. Sequencing

In total, over 35 billion raw paired-end sequencing reads were produced during the course of this PhD. A common obstacle in higher-depth sequencing was clonality. While rarely an issue in screening runs, some PCR libraries reached duplication levels as high as 70% when higher numbers of reads were sequenced (e.g. Uzz-CR126). Clonality was positively correlated with the number of reads sequenced across samples (Figure 3.15).

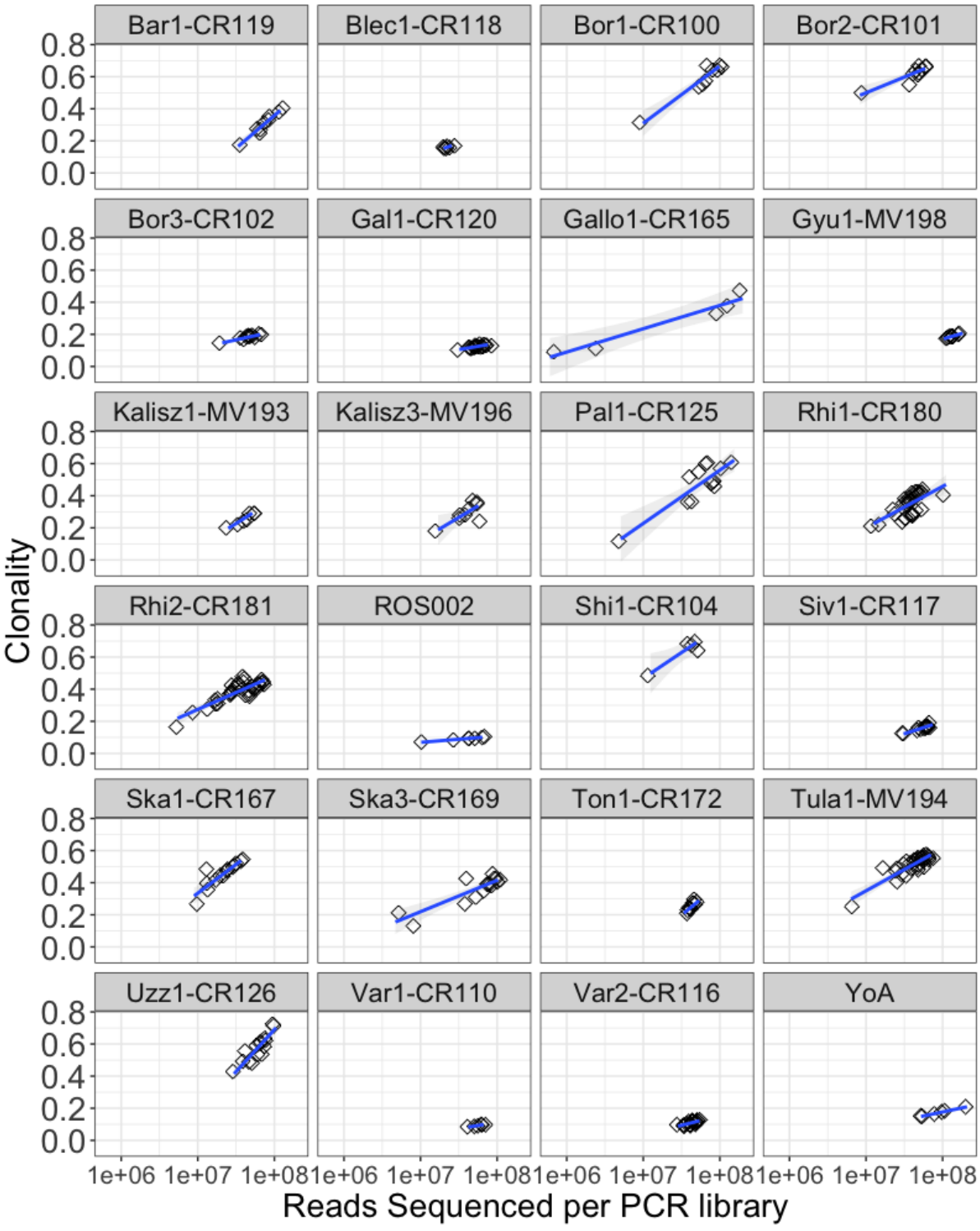


Figure 3.15 - Clonality versus number of reads sequenced per PCR library. A positive trend is observed across all samples with increasing the number of reads sequenced increasing clonality.

The relationship between clonality and DNA concentration was investigated. Clonality was computed for a range of different PCR libraries sequenced to approximately 50M reads. The

clonality was weakly negatively correlated with DNA concentration after 12 PCR cycles as quantified by TapeStation 2100 (R^2= 0.35, p= 0.0682). However, when excluding samples above 1 ng/ul, a strongly negative correlation was observed (R^2=0.73, p= 0.0050).

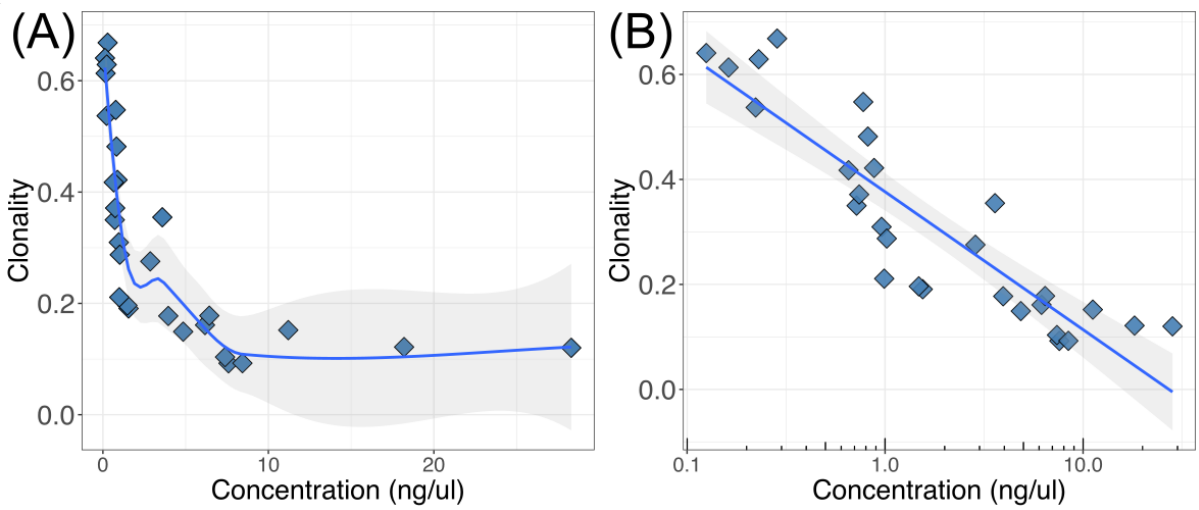


Figure 3.16 - The clonality of PCR libraries that were sequenced with approximately 50M reads (n=28). The clonality drops rapidly for samples of concentration at least 1 ng/ul (A). This is seen clearly when the X axis is log transformed (B). Each point on the plot represents unique libraries that were sequenced to ~50M reads on Illumina's NovaSeq platform.

3.5. Discussion

3.5.1. Endogenous DNA content

The objective for screening of ancient biological materials is to facilitate deep sequencing for downstream population genomic analysis. Previously, the petrous bone was discovered to be an excellent reservoir of endogenous aDNA, and is now the preferred substrate for aDNA analysis (Gamba et al. 2014b; Hansen et al. 2017). The findings presented in this Chapter corroborate this, as the petrous bone was largely found to have the highest average endogenous content. For the most part, other bone types yielded consistently lower endogenous DNA content.

There are well-documented issues with DNA preservation in arid, high-temperature climates (Rizzi et al. 2012). This was observed in our dataset, as no bone samples from Israel and UAE yielded endogenous DNA content above 1%. Despite this, latitude (as a proxy for temperature) was not correlated with sample preservation in the petrous bone samples. This was further evidenced in the consistently poor preservation in samples from Sweden, despite a relatively

cool and stable mean temperature. Recent models of post-mortem DNA decay suggest that aDNA damage is dictated by multiple processes and not easily predicted (Keighley et al. 2021). Soil pH, salinity or precipitation may have led to poorer preservation and underlie the difficulty of predicting aDNA survival. Therefore, it is advisable to initially sample a small proportion of the material available from any new site.

The collection of samples screened includes a novel dataset in the emerging field of "museomics". So called historical DNA (hDNA) has been defined as DNA derived from samples less than 200 years old, not originally archived for DNA analysis (Raxworthy and Smith 2021). hDNA collections can expand the understanding of extinction genetics, domestication and recent genetic erosion in endangered species and this particular collection will aid in the understanding of the poorly resolved *Bovini* tribe (Sinding et al. 2021). While most aDNA techniques can be directly applied for hDNA screening, some adjustments are necessary (Raxworthy and Smith 2021; McDonough et al. 2018). For example, the concentration of unamplified hDNA far exceeded what is generally encountered with aDNA, which prompted dilution of starting material and an adjustment of PCR cycling. This is likely due to the reduced time within which DNA degradation and leaching can take place in historical tissues compared to ancient samples. However, despite their recent age, the historical bone and tissue samples still yielded mixed levels of endogenous DNA and ancient petrous bones still outperformed historical samples in average endogenous DNA yields. In addition, damage levels in historical samples displayed considerable variability with some skin samples having extremely short median fragment lengths. This small dataset emphasises the limited understanding of DNA decay dynamics beyond mineralised tissue.

3.5.2. Deep-sequencing strategy

In genomic studies, bioinformatic analytic techniques dictate the minimum average genome coverage. Some analyses like IBS trees are possible with low average coverage (~0.01X), while others, like PSMC, necessitate high coverage genotype calls (>20X). While the endogenous DNA content of a sample was perhaps the most important factor in our deep sequencing strategy, desired genome coverage per sample was decided on a case-by-case basis. For example, some well-preserved samples offered little new information because the geographic and temporal origin of the sample was already well-represented in the dataset. For samples of interest for which shotgun sequencing was not feasible, mtDNA enrichment was attempted (Chapter 7).

The low-depth sequencing screening strategy meant that many samples could be economically and rapidly screened together. However, deep sequencing exposed a weakness to this strategy. At the screening stage, only ~100,000 reads were sequenced per PCR library which generally yielded very few duplicate reads. Clonality issues arose at the deep-sequencing stage when the number of reads sequenced was increased to ~50 million reads per PCR library. Clonality had varying effects depending on the sample (Figure 3.X), but resulted in a drastic underperformance in desired genome coverage for some. For example, the mesolithic Italian aurochs CR126 had an endogenous content of 45% and duplication rate of <1% when screened on a MiSeq run with ~80,000 reads. However, when PCR libraries were deep-sequenced at ~50M reads per PCR library, duplication rates reached as high as 70%. Ultimately, this resulted in a lower final genome coverage than expected. In every sample, a lower number of reads per PCR library was positively correlated with a lower clonality. A strategy to combat high clonality could therefore be to increase the number of PCR libraries per sample to lower the number of reads sequenced per reaction. In a small dataset, we found that bleached samples had a lower DNA concentration compared to when they were non-bleached. This is likely due to endogenous DNA complexity being lessened by the more aggressive extraction. As there is a strong correlation between DNA concentration and clonality in deep-sequenced samples, caution should be advised when considering potentially more destructive aDNA extraction methods. New methods of aDNA extraction should be continuously monitored, validated and revised if necessary.

The inconstant nature of available endogenous DNA content was further emphasised when several samples were re-extracted from old bone powder. Using the same extraction method, a substantial drop in endogenous DNA was observed for three of four samples. It is possible that the increased surface volume of the powderised bone accelerated DNA decay leading to a substantial reduction in endogenous DNA over the 2-4 years the powder was kept refrigerated. In contrast, four of five samples re-extracted using the bleach extraction method yielded substantial improvement.

3.5.3. Sex biases in aDNA datasets

In this dataset, a bias toward females emerged in domestic assemblages (approximate ratio of 2:1). A bias toward adult females in bone assemblages of ancient domestic cattle, sheep, and goat has been well-documented (Manning et al. 2015; Salque et al. 2013). This is likely the result of a practice of killing young males in domestic cattle, sheep and goats. Early male kill-off has been a hallmark of animal management for dairying animals since earliest farming and persists in modern agricultural practices (Baker and Worley 2019; Payne 1973). Very few

adult males are needed to produce offspring for the next generation, while preserving adult females maximises both dairying and breeding potential. It has been hypothesised that the preservation of young male remains may be worse due to unfused elements, resulting in the overrepresentation of females (Daly et al. 2021). Interestingly, sex biases in domestic assemblages do not always skew female. For example, there is a sex bias for males in horses from Bronze Age sites likely due to greater strength of stallions for transportation (Fages et al. 2020).

The predominance of male samples in the wild dataset is intriguing. A possible explanation may be due to the social structure of the aurochs, which make up most of the animals in this dataset. While little is known about behaviour of this animal due to its extinction in the 1600s, observations of other wild members of the *Bovini* tribe suggest a pattern of reproductive behaviour (van Vuure 2005). The social structure of wisent, gaur and banteng consists of cows, calves and young bulls living together year round. Older bulls tend to be isolated from the herd - living alone or in small groups of males for much of the year, only returning to the female herd at mating season. The isolated, wandering aurochs bulls may have perished in more diverse locations than their female counterparts, increasing the possibility for favourable preservation. A second possibility is that the larger size of aurochs makes positive identification easier. There was a significant sexual dimorphism in size of male and female aurochs (van Vuure 2005). Due to the reduced size, female aurochs may be generally less commonly identified as aurochs in assemblages.

Table 3.2 - Summary of 285 ancient and historical samples screened. Context is reproduced as received from the curator unless the sample is radiocarbon dated. * = sample dated by context; ^ sample radiocarbon dated indirectly.

Internal Lab ID	Alias	Material	Country	Site / Region	Context	Archaeological assignment	Extraction Type	Genetic assignment	Endog %	Sex
CR001		petrous	Netherlands	Enkhuizen-Kadijken	Bronze Age*	*Bos taurus*	Classic	*Bos*	2.53	F
CR002		petrous	Netherlands	Enkhuizen-Kadijken	Bronze Age*	*Bos taurus*	Classic	*Bos*	41.28	F
CR003		petrous	Netherlands	Enkhuizen-Kadijken	Bronze Age*	*Bos taurus*	Classic	*Bos*	31.88	M
CR004		petrous	Netherlands	Enkhuizen-Kadijken	Bronze Age*	*Bos taurus*	Classic	*Bos*	30.95	F
CR005		petrous	Netherlands	Enkhuizen-Kadijken	Bronze Age*	*Bos taurus*	Classic	*Bos*	4.69	F
CR006		petrous	Netherlands	Enkhuizen-Kadijken	Bronze Age*	*Bos taurus*	Classic	*Bos*	43.30	M
CR007		petrous	Netherlands	Enkhuizen-Kadijken	Bronze Age*	*Bos taurus*	Classic	*Bos*	18.43	M
CR008		petrous	Netherlands	Enkhuizen-Kadijken	Bronze Age*	*Bos taurus*	Classic	N/A	0.01	N/A
CR009		petrous	Netherlands	Stede Broec	Bronze Age*	*Bos taurus*	Classic	*Bos*	3.84	F
CR011		petrous	Netherlands	Stede Broec	Bronze Age*	*Bos taurus*	Classic	*Bos*	16.93	F
CR012		petrous	Netherlands	Stede Broec	Bronze Age*	*Bos taurus*	Classic	*Bos*	8.50	F
CR013		petrous	Netherlands	Stede Broec	Bronze Age*	*Bos taurus*	Classic	*Bos*	27.96	F
CR014		petrous	Netherlands	Stede Broec	Bronze Age*	*Bos taurus*	Classic	*Bos*	56.64	F
CR015		petrous	Netherlands	Stede Broec	Bronze Age*	*Bos taurus*	Classic	*Bos*	38.77	F
CR016		petrous	Netherlands	Stede Broec	Bronze Age*	*Bos taurus*	Classic	*Bos*	11.27	F
CR017		petrous	Lebanon	Tell Fadous	Bronze Age*	*Bos taurus*	Classic	*Bos*	28.16	F
CR018		petrous	Malta	Kordin Temples	Neolithic*	*Bos taurus*	Classic	*Bos*	0.05	N/A
CR019		petrous	Malta	Kordin Temples	Neolithic*	*Bos taurus*	Classic	*Bos*	0.60	F

CR020		petrous	United Arab Emirates	Baynunah	8000 BCE*	*Camelus dromedarius*	Classic	N/A	0.02	N/A
CR021		petrous	United Arab Emirates	Baynunah	8000 BCE*	*Camelus dromedarius*	Classic	N/A	0.05	N/A
CR022		petrous	England	Gloucester	Bronze Age*	*Bos taurus*	Classic	*Bos*	57.60	F
CR023		petrous	England	Gloucester	Bronze Age*	*Bos taurus*	Classic	*Bos*	55.40	F
CR024		petrous	England	Gloucester	Bronze Age*	*Bos taurus*	Classic	*Bos*	51.94	F
CR025		petrous	England	Gloucester	Bronze Age*	*Bos taurus*	Classic	*Bos*	48.84	M
CR026		tooth	Denmark	Syltholm	Neolithic (4500-2600 BCE)*	*Bos taurus*	Classic	*Bos*	37.30	F
CR027		petrous	Denmark	Syltholm	Neolithic (4500-2500 BCE)*	*Ovis aries*	Classic	*Ovis*	0.50	F
CR028		petrous	Denmark	Syltholm	Neolithic (4500-1300 BCE)*	*Bos taurus*	Classic	*Bos*	53.22	F
CR029		petrous	Denmark	Syltholm	Neolithic (4800-3400 BCE)*	*Bos taurus*	Classic	*Bos*	51.92	M
CR030		petrous	Iran	Shahre-Qumis	Parthian Empire*	*Bos*	Classic	*Bos*	3.78	M
CR031		talus	Iran	Tepe Naderi	Bronze Age/Parthian Empire*	*Bos*	Classic	*Bos*	2.14	M
CR032		petrous	Iran	Shahre-Qumis	Parthian Empire*	Small ruminant	Classic	*Capra*	20.54	F
CR033		petrous	Iran	Shahre-Qumis	Parthian Empire*	Small ruminant	Classic	*Capra*	15.22	F
CR034		petrous	Spain	Cova de l'avellaner	Early Neolithic*	*Caprinae*	Classic	*Ovis*	38.04	M
CR035		petrous	Spain	Cova de l'avellaner	Early Neolithic*	*Caprinae*	Classic	*Ovis*	55.55	F
CR036		petrous	Spain	Cova de l'avellaner	Early Neolithic*	*Caprinae*	Classic	*Ovis*	52.63	F
CR037		petrous	Spain	Cova de l'avellaner	Early Neolithic*	*Caprinae*	Classic	*Ovis*	37.22	M
CR038		petrous	Spain	Cova de l'avellaner	Early Neolithic*	*Caprinae*	Classic	*Ovis*	46.37	F
CR039		petrous	Spain	Cova de l'avellaner	Early Neolithic*	*Caprinae*	Classic	*Capra*	53.85	F

CR040		petrous	Spain	Cova de l'avellaner	Early Neolithic*	*Caprinae*	Classic	*Ovis*	24.16	F
CR041		petrous	Spain	Cova de l'avellaner	Early Neolithic*	*Caprinae*	Classic	*Ovis*	53.72	F
CR042		petrous	Spain	Cova de l'avellaner	Early Neolithic*	*Caprinae*	Classic	*Capra*	45.73	M
CR043		petrous	Spain	Cova de l'avellaner	Early Neolithic*	*Caprinae*	Classic	*Ovis*	38.81	M
CR044		petrous	Spain	Cova de l'avellaner	Early Neolithic*	*Caprinae*	Classic	*Ovis*	36.35	M
CR045		petrous	Spain	Cueva del Toro	Neolithic*	*Caprinae*	Classic	*Ovis*	54.87	M
CR046		petrous	Spain	Cueva del Toro	Neolithic*	*Caprinae*	Classic	*Ovis*	43.89	F
CR047		petrous	Spain	Cueva del Toro	Neolithic*	*Caprinae*	Classic	N/A	0.05	N/A
CR048		petrous	Spain	Cueva del Toro	Neolithic*	*Caprinae*	Classic	N/A	0.03	N/A
CR049		petrous	Spain	Cueva del Toro	Neolithic*	*Caprinae*	Classic	*Ovis*	51.86	F
CR050		petrous	Spain	Cueva del Toro	Neolithic*	*Caprinae*	Classic	*Sus*	37.05	F
CR051		petrous	Spain	Cueva del Toro	Neolithic*	*Caprinae*	Classic	N/A	3.61	N/A
CR052		petrous	Spain	Cueva del Toro	Neolithic*	*Caprinae*	Classic	*Ovis*	23.65	F
CR053		tooth	Italy	Montale	Bronze Age (1450-1400 BCE)*	*Ovis aries*	Bleach	*Ovis*	1.44	F
CR054		tooth	Italy	Montale	Bronze Age (1300-1200 BCE)*	*Ovis aries*	Bleach	*Ovis*	1.36	F
CR055		tooth	Italy	Montale	Bronze Age (1500-1450 BCE)*	*Ovis aries*	Bleach	*Ovis*	7.37	M
CR056		tooth	Italy	Montale	Bronze Age (1450-1400 BCE)*	*Ovis aries*	Bleach	*Ovis*	0.44	M
CR057		tooth	Italy	Montale	Bronze Age (1300-1200 BCE)*	*Ovis aries*	Bleach	*Ovis*	0.38	F
CR058		tooth	Italy	Montale	Bronze Age (1400-1350 BCE)*	*Ovis aries*	Bleach	*Ovis*	0.49	M
CR059		tibia	Germany	Schwabmünchen	Bronze Age*	*Ovis aries*	Bleach	N/A	0.00	N/A
CR060		calcaneus	Germany	Schwabmünchen	Bronze Age*	*Ovis aries*	Bleach	*Ovis*	2.63	M

CR061		tibia	Germany	Götschenberg	Bronze Age*	*Ovis aries*	Bleach	*Ovis*	2.14	M
CR062		calcaneus	Germany	Götschenberg	Bronze Age*	*Ovis aries*	Bleach	*Capra*	13.75	M
CR063		metatarsus	Germany	Götschenberg	Bronze Age*	*Ovis aries*	Bleach	*Ovis*	0.27	M
CR064		femur	Germany	Kelheim	Bronze Age*	*Ovis aries*	Bleach	*Ovis*	18.62	F
CR065		metatarsus	Germany	Kelheim	Bronze Age*	*Ovis aries*	Bleach	N/A	0.04	N/A
CR066		metacarpal	Germany	Kelheim	Bronze Age*	*Ovis aries*	Bleach	*Ovis*	0.07	N/A
CR067		cranial element	Scotland	Skara Brae, Orkney	Neolithic*	*Bos primigenius*	Bleach	*Bos*	0.09	N/A
CR068		astragalus	Israel	Abel Beth Maacah	Iron Age*	*Capra*	Bleach	N/A	0.04	N/A
CR069		astragalus	Israel	Abel Beth Maacah	Iron Age*	*Capra*	Bleach	*Capra*	0.08	N/A
CR070		astragalus	Israel	Abel Beth Maacah	Iron Age*	*Capra*	Bleach	*Capra*	0.11	N/A
CR071		astragalus	Israel	Abel Beth Maacah	Iron Age*	*Capra*	Bleach	*Capra*	0.57	N/A
CR072		astragalus	Israel	Abel Beth Maacah	Iron Age*	*Capra*	Bleach	N/A	0.02	N/A
CR073		coprolite	Israel	Ramon Crater	Late Neolithic (6000-6200 BCE)	*Capra*	Coprolite	*Capra*	4.03	N/A
CR074		coprolite	Israel	Ramon Crater	Late Neolithic (6000-6200 BCE)	*Capra*	Coprolite	N/A	4.82	N/A
CR075		coprolite	Israel	Ramon Crater	Bronze Age (3000 BCE)*	*Capra*	Coprolite	N/A	1.02	N/A
CR076		coprolite	Israel	Ramon Crater	Bronze Age (3000 BCE)*	*Capra*	Coprolite	*Capra*	1.18	F
CR077		coprolite	Israel	Ramon Crater	Recent (1700 CE)*	*Capra*	Coprolite	*Capra*	1.38	F
CR078		coprolite	Israel	Ramon Crater	Recent (1700 CE)*	*Capra*	Coprolite	*Capra*	0.58	F
CR079		coprolite	Israel	Ramon Crater	Late Neolithic (6000-6200 BCE)	*Capra*	Coprolite	*Capra*	2.86	F
CR080		coprolite	Israel	Ramon Crater	Late Neolithic (6000-6200 BCE)	*Capra*	Coprolite	N/A	0.84	N/A

CR081		coprolite	Israel	Ramon Crater	Late Neolithic (6000-6200 BCE)	*Capra*	Coprolite	*Capra*	0.08	N/A
CR082		coprolite	Israel	Ramon Crater	Late Neolithic (6000-6200 BCE)	*Capra*	Coprolite	N/A	0.36	N/A
CR083		coprolite	Israel	Ramon Crater	Late Neolithic (6000-6200 BCE)	*Capra*	Coprolite	N/A	0.43	N/A
CR084		coprolite	Israel	Ramon Crater	Bronze Age (3000 BCE)*	*Capra*	Coprolite	*Capra*	1.25	F
CR085		coprolite	Israel	Ramon Crater	Bronze Age (3000 BCE)*	*Capra*	Coprolite	N/A	0.26	N/A
CR086		coprolite	Israel	Ramon Crater	Bronze Age (3000 BCE)*	*Capra*	Coprolite	*Capra*	0.89	N/A
CR087		coprolite	Israel	Ramon Crater	Bronze Age (3000 BCE)*	*Capra*	Coprolite	*Capra*	0.49	F
CR088		coprolite	Israel	Ramon Crater	Bronze Age (3000 BCE)*	*Capra*	Coprolite	N/A	0.59	N/A
CR089		coprolite	Israel	Ramon Crater	Bronze Age (3000 BCE)*	*Capra*	Coprolite	*Capra*	2.35	N/A
CR090		coprolite	Israel	Ramon Crater	Recent (1700 CE)*	*Capra*	Coprolite	N/A	1.35	N/A
CR091		coprolite	Israel	Ramon Crater	Recent (1700 CE)*	*Capra*	Coprolite	N/A	0.39	N/A
CR092		coprolite	Israel	Ramon Crater	Recent (1700 CE)*	*Capra*	Coprolite	*Capra*	6.69	F
CR093		coprolite	Israel	Ramon Crater	Recent (1700 CE)*	*Capra*	Coprolite	N/A	1.03	N/A
CR094		coprolite	Israel	Ramon Crater	Recent (1700 CE)*	*Capra*	Coprolite	N/A	0.72	N/A
CR095		tibia	Germany	Kelheim	Bronze Age*	*Caprinae*	Bleach	*Ovis*	0.44	M
CR096		humerus	Germany	Kelheim	Bronze Age*	*Caprinae*	Bleach	*Ovis*	23.93	F
CR097		humerus	Germany	Kelheim	Bronze Age*	*Caprinae*	Bleach	*Ovis*	0.29	N/A
CR098		metacarpal	Germany	Götschenberg	Bronze Age*	*Caprinae*	Bleach	N/A	0.03	N/A
CR099		humerus	Germany	Götschenberg	Bronze Age*	*Caprinae*	Bleach	*Ovis*	0.43	F
CR100	Bor1	tooth	Kazakhstan	Borly	6740 - 6521 cal yrBP	*Bos primigenius*	Bleach	*Bos*	9.96	M
CR101	Bor2	tooth	Kazakhstan	Borly	4913 - 5099 cal yrBP	*Bos primigenius*	Bleach	*Bos*	8.82	M
CR102	Bor3	tooth	Kazakhstan	Borly	5115 - 4925 cal yrBP	*Bos primigenius*	Bleach	*Bos*	11.10	F

CR103	Shau1	tooth	Kazakhstan	Shauke	3983 - 3772 cal yrBP	*Bos primigenius*	Bleach	*Bos*	1.12	F
CR104	Shi1	tooth	Kazakhstan	Shiderty	Early Bronze Age*	*Bos primigenius*	Bleach	*Bos*	28.33	F
CR105		jaw bone	China	Kongni ditch	Mesolithic*	*Bos primigenius*	Bleach	*Bos*	6.25	F
CR106		tooth	China	Kongni ditch	Mesolithic*	*Bos primigenius*	Bleach	*Bos*	8.93	F
CR107	UKR11	tooth	Ukraine	Starobels	9000 - 8000 yrBP*	*Bos primigenius*	Bleach	*Bos*	0.21	F
CR108	UKR07	tooth	Ukraine	Starobels	9000 - 8000 yrBP*	*Bos primigenius*	Bleach	*Bos*	0.17	F
CR109	UKR08	tooth	Ukraine	Starobels	9000 - 8000 yrBP*	*Bos primigenius*	Bleach	*Bos*	0.24	N/A
CR110	Var1	petrous	Kazakhstan	Varfolomeevka	7641 - 7500 cal yrBP	*Bos primigenius*	Bleach	*Bos*	31.51	M
CR111		tooth	Kazakhstan	Karatomar	Unknown	*Bos primigenius*	Bleach	*Bos*	0.15	M
CR112		petrous	Russia	Verkhnyaya Alabuga	Unknown	*Bos primigenius*	Bleach	*Equus*	0.54	F
CR113		petrous	Kazakhstan	Varfolomeevka	Unknown	*Bos primigenius*	Bleach	*Bos*	0.30	N/A
CR114		tooth	Kazakhstan	Botai	Unknown	*Bos primigenius*	Bleach	*Equus*	12.46	F
CR115	Ura1	cranial element	Kazakhstan	Uralsk	Unknown	*Bos primigenius*	Bleach	*Bos*	0.49	F
CR116	Var2	petrous	Kazakhstan	Varfolomeevka	7486 - 7238 cal yrBP	*Bos primigenius*	Bleach	*Bos*	52.67	M
CR117	Siv1	petrous	Bulgaria	Vratsa	6258 - 6006 cal yrBP	*Bos primigenius*	Bleach	*Bos*	51.80	M
CR118	Blec1	petrous	Bulgaria	Vratsa	6258 - 6006 cal yrBP^	*Bos primigenius*	Bleach	*Bos*	49.33	M
CR119	Bar1	petrous	Italy	Barche di Solferino	13246 - 13005 cal yrBP	*Bos primigenius*	Bleach	*Bos*	10.68	M
CR120	Gal1	petrous	Spain	Galicia	9548 - 9113 cal yrBP	*Bos primigenius*	Bleach	*Bos*	54.88	M
CR121	Pad1	petrous	Italy	Padova	12267 - 11844 cal yrBP	*Bos primigenius*	Bleach	*Bos*	17.90	M
CR122		petrous	Italy	Padova	Mesolithic*	*Bos primigenius*	Bleach	*Bos*	2.94	M
CR123		petrous	Sweden	Mjölby	10765 - 10147 cal yrBP	*Bos primigenius*	Bleach	*Bos*	1.15	M

CR124		petrous	Italy	Grotta Romanelli	Late Upper Palaeolithic*	*Bos primigenius*	Bleach	*Bos*	0.06	M
CR125	Pal1	petrous	Italy	Palidoro	16287 - 15849 cal yrBP	*Bos primigenius*	Bleach	*Bos*	43.35	F
CR126	Uzz1	petrous	Italy	Grotta dell'Uzzo	7593 - 7480 cal yrBP	*Bos primigenius*	Bleach	*Bos*	44.95	M
CR127		petrous	Italy	Grotta della Madonna	Late Upper Palaeolithic*	*Bos primigenius*	Bleach	N/A	0.03	N/A
CR128		petrous	Italy	Torbiera	Bronze Age*	*Bos primigenius*	Bleach	*Bos*	0.24	M
CR129		jaw bone	France	Normandy	Roman Era*	*Bos primigenius*	Bleach	*Bos*	0.16	M
CR130		petrous	Sweden	Stora Slågarp	10570 - 9909 cal yrBP	*Bison bonasus*	Bleach	*Bison*	1.85	M
CR131		metacarpal	Sweden	Öbacken	Unknown	Bovine	Bleach	N/A	0.01	N/A
CR132		carpel	Sweden	Tågerup	8723 - 8375 cal yrBP	*Bos primigenius*	Bleach	*Bos*	1.88	M
CR133		metatarsal	Sweden	Bjäreshög	9471 - 9144 cal yrBP	Bovine	Bleach	*Bos*	1.05	M
CR134		tibia	Sweden	Rönneholm	8763 - 8430 cal yrBP	*Bos primigenius*	Bleach	*Bos*	3.20	M
CR135		calcaneus	Sweden	Häljarp	1864 - 1708 cal yrBP	*Bos primigenius*	Bleach	*Bos*	2.20	F
CR136		carpel	Sweden	Sjöholmen	Unknown	Bovine	Bleach	N/A	0.01	N/A
CR137		vertebrae	Sweden	Sunnansund	Unknown	Bovine	Bleach	N/A	0.03	N/A
CR138		humerus	Sweden	Sunnansund	Unknown	Bovine	Bleach	N/A	0.01	N/A
CR139		tooth	Sweden	Smedstorps	8590 - 8410 cal yrBP	Bovine	Bleach	N/A	0.66	N/A
CR140		vertebrae	Sweden	Smedstorps	9540 - 9425 cal yrBP	Bovine	Bleach	*Bos*	0.12	F
CR141		metacarpal	Sweden	Smedstorps	Mesolithic*	Bovine	Bleach	N/A	0.04	N/A
CR142		metatarsal	Sweden	Smedstorps	Mesolithic*	Bovine	Bleach	N/A	0.02	N/A
CR143		radius	Sweden	Genarp	9467 - 9013 cal yrBP	Bovine	Bleach	N/A	0.03	N/A
CR144		humerus	Sweden	Genarp	9527 - 9031 cal yrBP	Bovine	Bleach	N/A	0.02	N/A

CR145		humerus	Sweden	Gyllebo	11152 - 10404 cal yrBP	Bovine	Bleach	N/A	0.01	N/A
CR146		vertebrae	Sweden	Barsebäck	9905 - 8137 cal yrBP	Bovine	Bleach	N/A	0.05	N/A
CR147		costae	Sweden	Frörum	9423 - 9033 cal yrBP	*Bos primigenius*	Bleach	*Bos*	0.72	M
CR148		radius	Sweden	Lund	10119 - 9537 cal yrBP	*Bos primigenius*	Bleach	*Bos*	2.06	M
CR149		costae	Sweden	Lund	10119 - 9433 cal yrBP	*Bos primigenius*	Bleach	*Bos*	0.63	M
CR150		tooth	Sweden	Tranås-Esperöd	Mesolithic*	Bovine	Bleach	*Canis*	1.89	M
CR151		costae	Sweden	Tranås-Esperöd	Mesolithic*	Bovine	Bleach	*Bos*	0.16	N/A
CR152		costae	Sweden	Tranås-Esperöd	9691 - 9497 cal yrBP	Bovine	Bleach	*Bos*	0.43	M
CR153		costae	Sweden	Landskrona	9655 - 9447 cal yrBP	Bovine	Bleach	*Bison*	0.27	N/A
CR154		costae	Sweden	Slimminge	9553 - 9129 cal yrBP	Bovine	Bleach	*Bos*	0.20	M
CR155		costae	Sweden	Slimminge	9904 - 9488 cal yrBP	Bovine	Bleach	*Bos*	0.42	M
CR156		tooth	Sweden	Östra Grevie	Unknown	Bovine	Bleach	N/A	0.01	N/A
CR157		long bone	Sweden	Slabälta	11166 - 10603 cal yrBP	Cervidae	Bleach	*Alces*	1.59	M
CR158		humerus	Sweden	Ringsjöholm	Unknown	Bovine	Bleach	N/A	0.01	N/A
CR159		astragalus	Sweden	Ringsjöholm	Unknown	Bovine	Bleach	*Bos*	0.03	F
CR160		astragalus	Sweden	Ringsjöholm	Unknown	Bovine	Bleach	N/A	0.01	N/A
CR161		vertebrae	Sweden	Unknown	Unknown	Bovine	Bleach	*Bos*	0.09	N/A
CR162		metacarpal	Sweden	Övre Ålebäck	11190 - 10415 cal yrBP	Bovine	Bleach	N/A	0.04	N/A
CR163		petrous	Scotland	Fordyce, Banffshire	Mesolithic*	*Bos primigenius*	Bleach	*Bos*	0.19	M
CR164		petrous	Scotland	Unknown	Mesolithic*	*Bos primigenius*	Bleach	N/A	0.02	N/A
CR165	Gallo1	petrous	Scotland	Galloway	3764 - 3479 cal yrBP	*Bos primigenius*	Bleach	*Bos*	48.84	M

CR166	Fif1	petrous	Scotland	Newburgh, Fife	10634 - 10277 cal yrBP	*Bos primigenius*	Bleach	*Bos*	8.46	M
CR167	Ska1	petrous	Sweden	Skåne	9610 - 9207 cal yrBP	Bovine	Bleach	*Bos*	30.23	M
CR168	Ska2	petrous	Sweden	Skåne	10294 - 9896 cal yrBP	Bovine	Bleach	N/A	1.73	N/A
CR169	Ska3	petrous	Sweden	Skåne	9731 - 9508 cal yrBP	Bovine	Bleach	*Bos*	34.03	M
CR170	Gal1	petrous	Spain	Galicia	9481 - 9100 cal yrBP	*Bos primigenius*	Bleach	*Bos*	21.10	M
CR171	Lil1	petrous	Denmark	Randers	4484 - 4228 cal yrBP	*Bos primigenius*	Bleach	*Bos*	61.88	F
CR172	Ton1	petrous	Denmark	Jejsing	3279 - 3074 cal yrBP	*Bos primigenius*	Bleach	*Bos*	22.35	F
CR173	Fre1	petrous	Denmark	Frederiksborg, Alsønderup	10825 - 10496 cal yrBP	*Bos primigenius*	Bleach	*Bos*	0.58	M
CR174	Hjo1	petrous	Denmark	Hjørring, Tofte Bæk	8245 - 8053 cal yrBP	*Bos primigenius*	Bleach	*Bos*	16.18	M
CR175	Che1	petrous	Bulgaria	Chernomorets	7123 - 6818 cal yrBP	*Bos primigenius*	Bleach	*Bos*	42.88	M
CR176	Gal3	petrous	Spain	Galicia	9538 - 9109 cal yrBP	*Bos primigenius*	Bleach	*Bos*	38.94	M
CR177	Gal1	petrous	Spain	Galicia	9548 - 9113 cal yrBP	*Bos primigenius*	Bleach	*Bos*	25.55	M
CR178	Lun1	petrous	Denmark	Lundby	9744-9525 / 9045-8656 cal yrBP^	*Bos primigenius*	Bleach	*Bos*	6.11	F
CR179	Lun2	petrous	Denmark	Lundby	9744-9525 / 9045-8656 cal yrBP^	*Bos primigenius*	Bleach	*Bos*	4.08	M
CR180	Rhi1	petrous	Germany	Upper Rhine Valley	47000+ cal BP	*Bos primigenius*	Bleach	*Bos*	45.95	M
CR181	Rhi2	petrous	Germany	Upper Rhine Valley	47009 - 45312 cal yrBP	*Bos primigenius*	Bleach	*Bos*	28.35	M
CR182	Rhi3	petrous	Germany	Upper Rhine Valley	Pleistocene*	*Bos primigenius*	Bleach	*Bos*	57.94	M
CR183	Rhi4	petrous	Germany	Upper Rhine Valley	Pleistocene*	*Bos primigenius*	Bleach	*Bos*	44.31	M
CR184	Rhi5	petrous	Germany	Upper Rhine Valley	Pleistocene*	*Bos primigenius*	Bleach	*Bos*	36.87	M
CR185	Rhi6	petrous	Germany	Upper Rhine Valley	Pleistocene*	*Bos primigenius*	Bleach	*Bos*	39.61	M

CR186	Rhi7	petrous	Germany	Upper Rhine Valley	Pleistocene*	*Bos primigenius*	Bleach	*Bos*	27.31	M
CR187	Rhi8	petrous	Germany	Upper Rhine Valley	Pleistocene*	*Bos primigenius*	Bleach	*Bos*	18.63	M
CR188		vertebrae	Russia	Siberia	800 yrBP*	*Bison*	-	*Alces*	7.05	M
CR189		bone	Borneo	Central Borneo	Historical	*Bos javanicus*	Bleach	*Bos*	0.67	F
CR190		bone	Borneo	Borneo	Historical	*Bos javanicus*	Bleach	*Bos*	1.26	M
CR191		bone	Borneo	Borneo	Historical	*Bos javanicus*	Bleach	*Bos*	2.97	M
CR192		bone	Borneo	Borneo	Historical	*Bos javanicus*	Bleach	*Bos*	5.36	M
CR193		skin	Myanmar	Lepangang	Historical	*Bos javanicus*	Tissue	*Bos*	6.00	M
CR194		skin	Myanmar	Lepangang	Historical	*Bos javanicus*	Tissue	N/A	0.00	N/A
ER001		phalanx	Sweden	Ageröd	Mesolithic*	Bovine	-	N/A	0.01	N/A
ER002		calcaneus	Sweden	Haväng	9021 - 8723 cal yrBP	Bovine	-	N/A	0.04	N/A
ER003		costae	Sweden	Haväng	8970 - 8540 cal yrBP	Bovine	-	*Bos*	4.10	M
ER004		metatarsal	Sweden	Haväng	10705 - 10294 cal yrBP	Bovine	-	*Bos*	0.36	F
ER005		costae	Sweden	Stehag	8991 - 8645 cal yrBP	Bovine	-	*Bos*	1.70	M
ER007		petrous	Sweden	Skåne	10158 - 9489 cal yrBP	*Bison bonasus*	-	*Bison*	5.73	F
ER008		petrous	Sweden	Lommaryd	11180 - 10382 cal yrBP	*Bison bonasus*	-	*Bison*	5.57	M
ER009		petrous	Sweden	Sjörup	11072 - 10246 cal yrBP	*Bison bonasus*	-	N/A	0.02	N/A
ER010		metatarsal	Sweden	Haväng	10731 - 10430 cal yrBP	Bovine	-	*Bos*	2.74	F
ER011		bone	Sweden	Ageröd	Mesolithic*	Bovine	-	N/A	0.02	N/A
ER012		calcaneus	Sweden	Ageröd	Mesolithic*	Bovine	-	*Bos*	0.04	N/A
ER013		humerus	Sweden	Ageröd	Mesolithic*	Bovine	-	N/A	0.00	N/A

ER014		metacarpal	Sweden	Ageröd	Mesolithic*	Bovine	-	N/A	0.00	N/A
ER015		costae	Sweden	Ageröd	Mesolithic*	Bovine	-	N/A	0.00	N/A
ER016		calcaneus	Sweden	Ageröd	Mesolithic*	Bovine	-	*Bos*	0.05	N/A
ER017		calcaneus	Sweden	Ageröd	Mesolithic*	Bovine	-	N/A	0.01	N/A
ER018		humerus	Sweden	Ageröd	Mesolithic*	Bovine	-	N/A	0.01	N/A
ER019		humerus	Sweden	Ageröd	Mesolithic*	Bovine	-	N/A	0.01	N/A
ER020		metatarsal	Sweden	Ageröd	Mesolithic*	Bovine	-	N/A	0.02	N/A
ER021		humerus	Sweden	Ageröd	Mesolithic*	Bovine	-	N/A	0.00	N/A
ER022		phalanx	Sweden	Jussberg	8973 - 8544 cal yrBP	Bovine	-	N/A	0.04	N/A
ER023		phalanx	Sweden	Jussberg	8318 - 7935 cal yrBP	Bovine	-	*Bos*	1.49	F
ER024		metacarpal	Sweden	Västra Tollstrad	Mesolithic*	Bovine	-	*Alces*	0.64	M
ER025		cranial element	Sweden	Västra Tollstrad	8984 - 8593 cal yrBP	Bovine	-	*Bos*	0.10	F
ER026		long bone	Sweden	Västra Tollstrad	Mesolithic*	Bovine	-	*Alces*	1.61	M
ER027		carpel	Sweden	Västra Tollstrad	8999 - 8645 cal yrBP	Bovine	-	*Bos*	0.41	F
ER028		carpel	Sweden	Västra Tollstrad	8993 - 8645 cal yrBP	Bovine	-	*Bos*	0.25	F
MV018		petrous	North Macedonia	Govrlevo	Neolithic*	*Bos taurus*	Bleach	*Bos*	2.01	N/A
MV027		petrous	Switzerland	Sutz-Lattrigen VII Hafen	3800 - 3850 BCE*	*Bos taurus*	Bleach	*Bos*	9.24	F
MV045		petrous	Türkiye	Menteşe	Neolithic*	*Bos taurus*	Bleach	*Bos*	10.24	M
MV055	Kalisz1	petrous	Netherlands	Schipluiden	Neolithic (3500 BCE)*	*Bos taurus*	Bleach	*Bos*	5.67	F
MV193	Kalisz2	petrous	Poland	Kalisz	4487 - 4229 cal yrBP	*Bos primigenius*	Bleach	*Bos*	26.36	M
MV194	Tula1	petrous	Russia	Tula	40746 - 42460 cal yrBP	*Bos primigenius*	Bleach	*Bos*	45.49	M

MV196	Kalisz3	petrous	Poland	Kalisz	4477 - 4168 cal yrBP	*Bos primigenius*	Bleach	*Bos*	34.63	M
MV239	Kok1	petrous	Uzbekistan	Kok Tepe	5000-3500 yrBP*	*Bos taurus*	Bleach	*Bos*	26.96	M
MV240	Yer1	petrous	Uzbekistan	Yerqorqan	2100 yrBP*	*Bos taurus*	Bleach	*Bos*	45.78	F
NK1		bone	Cambodia	Phnom Koulen	Historical	*Bos sauveli*	-	*Bos*	5.67	M
NK2		bone	Cambodia	Phnom Koulen	Historical	*Bos sauveli*	-	*Bos*	25.14	M
4169		petrous	Thailand	Mekrong	Historical	*Bos gaurus*	-	*Bos*	0.40	N/A
4476		bone	Cambodia	Unknown	Historical	*Bos gaurus*	-	*Bos*	26.40	M
4477		bone & cartilage	Cambodia	Unknown	Historical	*Bos gaurus*	-	*Bos*	51.21	F
4478		cartilage	Cambodia	Unknown	Historical	*Bos javanicus*	-	*Bos*	10.87	M
AH067		petrous	Georgia	Unknown	Unknown	*Caprinae*	Classic	*Ovis*	20.44	M
AH068		petrous	Georgia	Unknown	Unknown	*Caprinae*	Classic	*Ovis*	34.01	M
AH071-1		petrous	Georgia	Unknown	Unknown	*Caprinae*	Classic	*Ovis*	38.61	M
AH071-2		petrous	Georgia	Unknown	Unknown	*Caprinae*	Classic	*Ovis*	50.31	M
AH074		petrous	Uzbekistan	Dzharkutan	Bronze Age*	*Caprinae*	Classic	*Ovis*	10.57	F
AH075		petrous	Uzbekistan	Dzharkutan	Bronze Age*	*Caprinae*	Classic	N/A	0.52	N/A
AH076-1		petrous	Uzbekistan	Dzharkutan	Bronze Age*	*Caprinae*	Classic	N/A	0.05	N/A
AH076-2		petrous	Uzbekistan	Dzharkutan	Bronze Age*	*Caprinae*	Classic	N/A	0.05	N/A
AH099		petrous	Hungary	Szazhalombatta	Bronze Age*	*Caprinae*	Classic	*Bos*	14.65	M
AH100		petrous	Hungary	Szazhalombatta	Bronze Age*	*Caprinae*	Classic	N/A	0.20	N/A
AH101		petrous	Hungary	Szazhalombatta	Bronze Age*	*Caprinae*	Classic	N/A	0.04	N/A

AH102-1		petrous	Hungary	Szazhalombatta	Bronze Age*	*Canis*	Classic	*Canis*	58.28	M
AH103		petrous	Hungary	Szazhalombatta	Bronze Age*	*Caprinae*	Classic	*Bos*	22.21	F
AH106		petrous	Finland	Karleby	Unknown	*Caprinae*	Classic	*Ovis*	5.58	F
AH107		petrous	Türkiye	Unknown	Neolithic*	*Caprinae*	Classic	N/A	0.47	N/A
AH108		petrous	Lebanon	Tell Fodous	Bronze Age*	*Caprinae*	Classic	*Ovis*	5.87	F
AH109		petrous	Netherlands	Enkhuizen	Bronze Age*	*Caprinae*	Classic	*Ovis*	23.92	F
AH110		petrous	Netherlands	Jelsum	Roman (12 BCE - 300 CE)*	*Caprinae*	Classic	*Ovis*	23.99	M
AH111		petrous	Netherlands	Jelsum	Iron Age (250-12 BCE)*	*Caprinae*	Classic	*Ovis*	62.08	F
AH112		petrous	Malta	Santa Verna	Neolithic*	*Caprinae*	Classic	*Ovis*	3.66	N/A
AH113		petrous	Malta	Santa Verna	Neolithic*	*Caprinae*	Classic	N/A	0.20	N/A
AH114-1		petrous	Malta	Santa Verna	Neolithic*	*Caprinae*	Classic	*Ovis*	9.84	F
CN1147		tooth	Thailand	Bandon Province	Historical	*Bos javanicus*	-	*Bos*	12.90	M
CN1264		cartilage	Thailand	Unknown	Historical	*Bos javanicus*	-	*Bos*	0.13	N/A
CN1576		cartilage	Thailand	Bandon Province	Historical	*Bos javanicus*	-	*Bos*	44.98	F
CN1577		cartilage	Thailand	Bandon Province	Historical	*Bos javanicus*	-	*Bos*	42.95	F
CN1578		cartilage	Thailand	Bandon Province	Historical	*Bos javanicus*	-	*Bos*	2.58	M
CN1587		bone	Thailand	Bandon Province	Historical	*Bos gaurus*	-	*Bos*	25.78	F
CN1588		bone	Thailand	Bandon Province	Historical	*Bos gaurus*	-	*Bos*	16.20	F
CN1589		cartilage	Thailand	Bandon Province	Historical	*Bos gaurus*	-	N/A	0.02	N/A
CN1590		bone	Thailand	Bandon Province	Historical	*Bos gaurus*	-	*Bos*	10.35	M
CN1591		cartilage	Thailand	Bandon Province	Historical	*Bos gaurus*	-	*Bos*	8.73	M

CN1592		bone	Thailand	Bandon Province	Historical	*Bos javanicus*	-	*Bos*	25.20	F
CN3353		bone	Afghanistan	Kabul	Historical	*Bos indicus*	-	*Bos*	11.21	M
CN3402		bone	Thailand	Bangkok	Historical	*Bos indicus*	-	*Bos*	23.71	M
CN526		bone	India	Gulma, Sikkim Terai	Historical	*Bos gaurus*	-	*Bos*	1.84	M
CN527		bone	India	Gulma, Sikkim Terai	Historical	*Bos gaurus*	-	*Bos*	54.23	F
CN528		bone	India	Gulma, Sikkim Terai	Historical	*Bos gaurus*	-	*Bos*	17.81	M
CN529		bone	Bhutan	Tondoo, Bhootan Dooars	Historical	*Bos gaurus*	-	*Bos*	36.57	F
CN643		cartilage	Indonisia	Java	Historical	*Bos javanicus*	-	*Bos*	27.99	M
CN867		bone	India	India	Historical	*Bos gaurus*	-	*Bos*	2.44	M
CN903		bone	Indonisia	Probolinggo	Historical	*Bos javanicus*	-	*Bos*	8.74	M
CN904		bone	Indonisia	Probolinggo	Historical	*Bos javanicus*	-	*Bos*	8.94	M
CN906		bone	Indonisia	Probolinggo	Historical	*Bos javanicus*	-	*Bos*	17.67	F
CN907		bone	Indonisia	Probolinggo	Historical	*Bos javanicus*	-	*Bos*	9.11	F
ROS001	ROS001	astragalus	Kazakhstan	Roshchinskoe	5539 - 5372 cal yrBP	*Bos primigenius*	Bleach	*Bos*	0.22	M
ROS002	ROS002	astragalus	Kazakhstan	Roshchinskoe	5116 - 4926 cal yrBP	*Bos primigenius*	Bleach	*Bos*	1.57	F
WB001		ulna	Ireland	Ballycahane Lower	Mesolithic*	*Sus*	Bleach	*Sus*	0.06	N/A
WB002		ulna	Ireland	Ballycahane Lower	Mesolithic*	*Sus*	Bleach	*Sus*	0.01	N/A
WB004		humerus	Ireland	Ballycahane Lower	Mesolithic*	*Sus*	Bleach	*Sus*	0.90	N/A

4. Exceptional ancient DNA preservation and fibre remains of a Sasanian saltmine sheep mummy in Chehrābād, Iran

This chapter is adapted from the paper “Exceptional ancient DNA preservation and fibre remains of a Sasanian saltmine sheep mummy in Chehrābād, Iran” (Rossi et al. 2021), published in *Biology Letters*.

4.1. Introduction

In 1993, a remarkably preserved human body dating to the ~1700 yrBP was discovered in the Douzlākh salt mine near Chehrābād village in the Zanjan Province of northwest Iran (Aali et al. 2012; Aali and Stöllner 2015; Stöllner, Aali, and Kashani 2020). A total of 8 “Salt Men” have been identified at the mine (Öhrström et al. 2016; Nasab et al. 2019), several retaining keratinous tissues such as skin, hair, and both endo- and exoparasites, despite dating to the Achaemenid (550-330 BCE, 2500-2280 yrBP) and Sasanian (224-651 CE, ~1700-1300 yrBP) periods. The mine, also known as Chehrābād, was active in various periods and its archaeological refilling layers represent an extraction history that ranged from the 6th century BCE to 20th century CE. In addition to the “Salt Men”, textiles, leather objects, and animal remains have been discovered (Mashkour 2014; Mashkour, Fathi, and Davoudi 2020), likely preserved by high salinity and low moisture content of the mine. Isotopic, genetic, and lipid analyses have been reported for this material (Aali et al. 2012), and studies have been carried out to characterise genomic DNA survival (Warinner and Bouwman 2015).

These human and animal remains are examples of natural mummification - the spontaneous desiccation of soft tissue by a dry environment that rapidly dehydrates soft tissue before decay begins (David 2008). Mummification has been suggested as a mechanism that may sufficiently preserve keratinised tissue for aDNA sequencing (David 2008). The effects of age-related damage in aDNA are well documented and include base misincorporation at strand overhangs, fragmentation and low endogenous content (Dabney, Meyer, and Pääbo 2013). Both deamination and depurination, associated with post-mortem transition error and DNA fragmentation, respectively, require water as a substrate (Dabney, Meyer, and Pääbo 2013). Ancient DNA from Chehrābād, a highly saline, anhydrous environment, presents an opportunity to investigate potential differences in nucleotide degradation resulting from this unusual

taphonomic context. This study centres on a mummified sheep leg that was recently discovered in a large mining gallery in Douzlākh saltmine of Chehrābād during archaeological excavations (Aali and Stöllner 2015). The specimen was likely deposited during refilling activities in the 4-5th centuries CE after the reopening of the gallery in the Early Sasanian period (2nd-3rd centuries CE) and following its initial collapse between 405-380 BCE.

The leg was possibly discarded during food preparation activities, as both sheep and goat were likely used as provisioning for Sasanian-period miners (Nezamabadi et al. 2013). By this time, sheep were an established commodity for their meat and secondary products such as wool fibre. The derived woolly fleece was widespread in Southwest Asia by 4th millennium BCE and showed regional specialisation by the 3rd millennium BCE (Arbuckle and Hammer 2019). Recently, the derived "woolly" phenotype has been suggested to be a result of a ~1.5 kbp insertion of an antisense *EIF2S2* retrogene into the *IRF2BP2* 3' untranslated region (UTR) (Demars et al. 2017). This insertion allele was shown to be recessive to the ancestral allele associated with "hairy" coat and produced a chimeric RNA-RNA transcript, possibly interacting with the genuine sense *EIF2S2*. The fat-tailed sheep phenotype appeared in the 3rd millennium BCE, possibly selected to adapt to harsh, nutrient-poor conditions in Central Asia (Dong et al. 2020). Recently, the platelet derived growth factor D (*PDGFD*) gene was proposed as a gene controlling this phenotype due to high genetic differentiation between fat- and thin-tailed sheep breeds. Both woolly and fat-tailed sheep are depicted in the Early Bronze Age Mesopotamia but the spread of these phenotypes may have been uncoupled, and occurred via distinct processes (Pöllath, Schafberg, and Peters 2019; Vila et al. 2021). Genetic investigation of the ancient Southwest Asian sheep, including key domestic trait characterisation, may reveal important aspects of animal husbandry during the Sasanian Empire.

4.2 Aims

1. To characterise damage patterns in ancient DNA sequenced from mummified tissue and draw comparisons to other published ancient tissues.
2. To explore the metagenomic landscape of the mummified tissue.
3. To investigate the ancestry of the sheep in the context of modern variation.
4. To genotype alleles associated with derived domestic traits, integrating insights from microscopy of tissue and fibre remains.

4.3. Materials and methods

4.3.1. Materials

A sample of the mummified sheep skin (MUM2) from sheep leg 4305 was used in this study. Two Iranian sheep bone samples (Khor1 and Azer2) of approximately similar ages were also analysed for comparison (Table 1). MUM2 was directly radiocarbon dated at the ^{14}CHRONO Centre (Queen's University Belfast). OxCal 4.3.2 (Ramsey 2009) was used to calibrate its age (95.4% confidence interval) using (Reimer et al. 2013).

Figure 4.1 - Sheep leg (4305) in the archaeological context of layer 31091 (left) and recent condition after cleaning (two angles, centre, right). Photos: Deutsches Bergbau-Museum Bochum/RUB/AMZ, Thomas Stöllner (left), N. Tehrani (centre, right).

4.3.2. DNA extraction, library construction and sequencing

DNA extraction and initial screening sequencing library construction were performed by Dr Kevin Daly and Fionnuala McDaid. Details are reported in (Rossi et al. 2021).

Both USER and non-USER treated libraries of MUM2 were prepared and sequenced. Three rounds of sequencing were done: single-end and paired-end shotgun sequencing of USER-

treated libraries on Illumina HiSeq 2500 (Macrogen), and single-end shotgun sequencing for non USER-treated libraries on Illumina MiSeq (TrinSeq). The USER-treated library was sequenced on a HiSeq 2500 Illumina platform (100bp SE and 100bp PE) via a commercial sequencing company (Macrogen, Republic of Korea) while the non-UDG-treated library was sequenced on a MiSeq Illumina platform (50bp SE) at TrinSeq (Trinity College Dublin, Ireland) with Phi X control at 1%.

4.3.3. Read-processing

Trimmed sequencing reads were aligned to OviAri3.1 using bwa 0.7.5 with relaxed parameters. Post-alignment read processing was done according to Section 2.4.

4.3.4. Damage assessment

Bam files were assessed using mapDamage2.0 (Jónsson et al. 2013). mapDamage2.0 calculates read length distributions and substitution patterns at the 5' and 3' ends of read fragments of aligned bam files. mapDamage2.0 also outputs a number of damage parameters estimated in a Bayesian manner including δS, the cytosine deamination probability in single strand context. Published sequencing data of other ancient samples were retrieved from NCBI GenBank in order to contextualise the damage patterns of our sample with published data (Appendix Table 4.1-3).

Depurination levels were inferred using frequencies of purines close to the 5' end of fragment strand-breaks. The use of *T4* DNA polymerase in our library preparation removes 3' overhangs, therefore only 5' ends are considered. Only non-UDG libraries were used to calculate these levels as USER excises uracil residues in DNA fragments, creating new strand breaks. Using Qualimap v2.1.3 (Okonechnikov, Conesa, and García-Alcalde 2016), nucleotide frequencies were calculated on filtered bam files at each position of DNA fragments. Following the methods detailed in (Sawyer et al. 2012), approximate depurination rates were calculated.

4.3.5. Metagenomic analysis

For all metagenomic analysis, all reads originating from the host organism and human contamination were removed. To do this, all trimmed reads were aligned to the sheep genome (oviAri3.1) and the human genome (hg38) using methods detailed above. Unmapped reads were retained using SAMtools view -f 4 and converted to fastq files using SAMtools fastq. The retained reads were then deduplicated using PrinSeq (Schmieder and Edwards 2011).

Unmapped, deduplicated reads were searched against the Kraken2's standard database using default settings (Wood, Lu, and Langmead 2019). This database includes sequences of bacterial, archaeal and viral genomes. To calculate Class abundances, microbial abundances were recalculated using Bracken (Lu et al. 2017) est_abundance.py, with a minimum threshold of 0.1% of total classified reads. SourceTracker2 was used to further define the metagenomic composition of MUM2. DNA sequences of various relevant metagenomes (salt-rich, live sheep skin, soil, tissue decomposers, and laboratory reagents) were downloaded from NCBI (Table 4.1). The sequences were trimmed using AdapterRemoval v2.1.1 and taxonomic assignment for these reference sequences was done by Kraken2, as detailed above. The kraken reports were merged and converted to a BIOM table using kraken-biom (https://github.com/smdabdoub/kraken-biom) and the human taxid (9606) was removed. Bacterial species abundances were also generated using MIDAS (Nayfach et al. 2016b).

Table 4.1 - Metagenomic sources used for SourceTracker2.

Name	N	Reference
Tissue decomposers	20	(Emmons et al. 2020)
Salt-rich environment	31	(Robinson et al. 2015)
Sheep skin	12	(Ross et al. 2018)
Soil	20	(Moreira-Grez et al. 2019)
Laboratory reagents	35	(Salter et al. 2014)

4.3.6. Mitochondrial analysis

All UDG-treated sequencing data was aligned to the domestic sheep mitochondrial reference, AF010406.1 (retrieved from NCBI GenBank) using an identical method as to the nuclear genome alignment. Consensus fasta sequences were generated using ANGSD (Korneliussen, Albrechtsen, and Nielsen 2014) (angsd -doFasta 2 -doCounts 1 -setMinDepth 3 -minQ 20 -minMapQ 30). With these flags, base calls must have a minimum quality of 20 and a minimum MAQ of 30 to be considered. In the event that there are more than one base at a given site, the most common base is chosen. For a site to be called a minimum depth of 3 is required.

Ten modern sheep whole genome sequences and one urial sheep whole genome sequence published in (Meadows, Hiendleder, and Kijas 2011) were retrieved from NCBI GenBank. Multiple sequence alignments were performed with MUSCLE (Edgar 2004). A maximum-likelihood phylogeny (100 bootstraps) was generated. The HKY85 substitution model was

selected for tree-building using jmodeltest2 (Darriba et al. 2012; Guindon and Gascuel 2003). Alignments and maximum-likelihood phylogeny constructions were implemented in SeaView v. 5.0.1 (Edgar 2004; Gouy, Guindon, and Gascuel 2010; Guindon et al. 2010).

4.3.7. SNP calling

The ovine SNP50 HapMap dataset used for the analyses described was provided by the International Sheep Genomics Consortium and obtained from www.sheephapmap.org in agreement with the ISGC Terms of Access. This dataset is genotyped on the Illumina OvineSNP50 Genotyping BeadChip (Illumina), covering >54,000 SNPs evenly spaced over the sheep genome. 1503 genotyped individuals are included in this dataset, representing modern breeds from Africa, the Americas, Europe, SW Asia and Asia (Kijas et al. 2012). The genomic positions of the OvineSNP50 were updated for the OviAri3.1 genome build and SNPs flipped to the forward strand by Dr Matthew Teasdale. Using PLINK v1.9 (Chang et al. 2015), these SNPs were filtered down to 44,223 sites using a MAF filter of ≤ 0.5.

The coverage of our sample was too low (approx. 4X) for accurate diploid genotype calling. To circumvent this issue, the pileup tool in GATK (McKenna et al. 2010) v3.7 was used to report sample base calls for each read at each position of the filtered SNP list (44,223 SNP positions). A minimum base quality of 30 was required to be considered and sites with three or more different bases present were removed. A single base call was then chosen at random from each site and duplicated to create a pseudohaploid homozygous genotype at that position for each individual. In total 43,027 SNPs were called for MUM2.

Using PLINK v1.90, the pseudohaploidised MUM2 genotype calls were merged with the modern sheep breeds dataset. In cases where the allele of MUM2 matched neither of the alleles given in the sheep breed position, the SNP position was flipped using PLINK v1.90 and attempted to be merged again. Any remaining triallelic SNPs were removed. The final merged dataset was used for f_3 statistics, ADMIXTURE and TreeMix analysis.

4.3.8. Autosomal analysis

Projection PCA using Procrustes analysis was performed using LASER v2.03 (Wang et al. 2015). The PCA reference space and projection transformation were constructed using a subsample of our modern breeds dataset. This subsample contained one representation from all modern breeds, chosen randomly. Using the OvineSNP50 reference panel, pileup files were generated for MUM2, Khor1 and Azer2 using SAMtools mpileup (-q 30 -Q 20). Each sample

was projected onto this PCA space and averaged across ten independent runs. The uncertainty based on the standard deviation * 2 for each sample was plotted on the PCA.

f_3 statistics were computed to quantify the amount of genetic drift shared between MUM2 and modern sheep breeds. Three Asiatic mouflon individuals that were genotyped on the OvineSNP50 Genotyping BeadChip were used as an outgroup. Using PLINK v1.9, these were merged with the full dataset of modern breeds and MUM2. Using PLINK v1.9, a filter of 0 for missingness was applied to this merged dataset. The dataset was then filtered for linkage disequilibrium (LD) between markers, with the command --indep-pairwise 50 5 0.5. Outgroup f_3 values were not estimated for Khor1 and Azer2 due to the lower number of called SNPs. Using this complete dataset, the f_3 statistics were computed in the form of f_3(MUM2, modern breed; Asiatic Mouflon) with ADMIXTOOLS (Patterson et al. 2012) (Appendix Table 4.4). f_3 statistics were plotted using approximate geographic origins of breeds with R using the sf package (Pebesma 2018).

To perform ADMIXTURE analysis, the genotyped breed dataset was downsampled using PLINK v1.9 to include approximately 90 individuals from each Asia, Europe and Africa. Using PLINK v1.9, a filter of 0 for missingness was applied to this merged dataset. The dataset was then filtered for linkage disequilibrium (LD) between markers, with the command --indep-pairwise 50 5 0.5. Finally, transitions were removed using PLINK v1.9 and VCFtools (Danecek et al. 2011). This fully filtered dataset contained 31,261 sites. Unsupervised ADMIXTURE was explored with a range of hypothetical ancestral populations, *K*s (2-20). ADMIXTURE runs were replicated three times for each K and set a different random seed for each. The three replicated runs for each K were averaged, and the cross-validation values which describe the approximate error of each ADMIXTURE run were inspected (Alexander and Lange 2011). The run with the lowest cross-validation error was selected for presentation, *K*=13.

The complete dataset used was converted into TreeMix format with the plink2treemix.py utility (Pickrell and Pritchard 2012). TreeMix (Pickrell and Pritchard 2012) was used to construct a model of population splits with no migration events. The tree was rooted with Asiatic mouflon and blocks of 500 SNPs were selected (-k 500). No sample size correction was used (-noss).

4.3.9. Trait genotyping

To investigate the woolly locus, two modified OviAri3.1 genome assemblies were produced. To exactly reproduce the locus detailed in (Demars et al. 2017), the sequence: ‘ACTTCTGTAAAATGCAAGATAAATTAAAGTTATTATAACAGTGATTCTTTCAAAAAAATAA

AAACACCATGA' was inserted at position chr25:7,452,231/7,452,232. This modified genome represents the woolly locus. To construct the ancestral locus (i.e. with no asEIFS insertion in chromosome 25), this assembly was edited to replace sequence chr25:7,450,861-7,452,231 with 'TAAAAACACCATGA', as per (Demars et al. 2017).

All UDG-treated sequencing data were aligned to these modified assemblies in an identical manner as above to produce two separate bam files. The coverage at the woolly locus was visualised using Integrative Genomics Viewer (IGV) (Robinson et al. 2011), before and after a MAQ filter of 30 was applied. Since there are three copies of EIF2S2 genes annotated on the OviAri3.1 genome assembly (One genuine *EIF2S2* gene located on chromosome 13 and two *EIF2S2* pseudogenes located on chromosome 7 and chromosome 25), the mapping quality of reads aligned to the insertion are 0 (Demars et al. 2017).

The 'breakpoints' of the woolly locus were defined at positions chr25:7,450,874/5 and chr25:7,452,217/8 of the modified oviAri3.1 assembly. Any reads which straddled either of these sites were deemed evidence of a woolly allele. The 'breakpoint' of the ancestral locus was chr25:7,450,861/2. Similarly any read that straddled this point was deemed evidence of the hairy/ancestral allele. Any read attributed to one genotype must only align to one modified assembly and not the other.

51 SNPs in the *PDGFD* gene associated with the derived fat-tail phenotype (Dong et al. 2020), were assessed using the genotype calls of modern fat and thin-tailed breeds to define the derived allele (Taylor et al. 2021; Li et al. 2020a). As the average genome coverage was too low for accurate diploid genotype calls, we report base calls for both alleles. The pileup tool in GATK (McKenna et al. 2010) v3.7 was used to report sample base calls at 51 SNP sites defined that are highly differentiated between fat-tailed and thin-tailed breeds, and thought to underlie the fat-tail phenotype (Dong et al. 2020). The derived fat-tail alleles were determined using genotypes of the 29 diverse breeds of fat and thin-tailed breeds. These breeds were genotyped as part of the International Sheep Genomics Consortium (ISGC). Genotypic data were retrieved from (Taylor et al. 2021), with methods detailed therein. The average genome coverage of MUM2 (~4X) was too low for accurate diploid genotype calling at these sites. Instead, a simple presence/absence of alleles based on base calls was reported (Appendix Table 4.5).

4.3.10. Microscopy analysis

Hair fibres were examined using scanning electron and light microscopes. Light microscopy was done by G. Ruß-Popa at the Austrian Academy of Sciences, Vienna, Austria. Zeiss

SteREO Discovery.V20 (Magnification: 1–150 X) was used for Stereomicroscope (reflected light) analysis. Zeiss Axio Scope.A1 Pol. (Magnification: 1 X – 2.5 X – 5 X – 10 X – 20 X – 40 X) was used for the polarised light microscope analysis for the thin section. The microscope camera used was Axiocam 305 colour.

SEM images were generated by G. Ruß-Popa and Andreas Steiger-Thirsfeld at USTEM, TU Vienna. FEI Quanta 250 FEG-SEM with AFM for Structural investigations (EBSD) was used for SEM imagery.

4.4. Results

4.4.1. DNA preservation patterns

The Chehrābād mummy sample (MUM2) was directly dated to the 5-6th century CE (2 sigma 1621-1481 cal BP, uncalibrated 1600 ± 30 BP). This aligns with the Sasanian Empire period of Iran, a time when the mine was in active use (Aali et al. 2012). Initial DNA screening indicated high endogenous DNA for MUM2 and the comparative Iranian sheep samples from relatively close time periods (Table 4.2). Sequencing of the Chehrābād mummy produced a 3.94X genome after quality filtering in addition to the low coverage comparative genomes, Khor1 and Azer2 (0.04X and 0.07X).

Table 4.2. Summary information of samples sequenced in this study. * = directly dated.

Name	Tissue	Origin	Period	Age	Endogenous DNA %	Coverage (X)
MUM2	Mummified skin	Chehrābād, Iran	Sasanian Empire period	*399-539 cal CE	31.01	3.94
Khor1	Petrous bone	Nishapur, Iran	Sasanian - Islamic periods	600 - 1200 CE	58.44	0.04
Azer2	Petrous bone	Tepe Hasanlu, Iran	Iron Age III	800-600 BCE	31.32	0.07

MUM2 differed from the two comparative sheep samples in displaying longer fragment lengths; the median fragments length of in MUM2 was 107bp vs 52bp and 56bp in Azer2 and Khor1 (collapsed reads-only: 90bp vs 50bp and 55bp). MUM2 also fell outside the ranges of both median fragment length when contrasting previously published ancient ovicaprid data from Southwest Asia and Europe (Figure 4.2). Related to the fragmentation, the depurination rates

of the three ancient sheep were similar; 6.3%, 5.6% and 5.3% for MUM2, Khor1, Azer2, respectively.

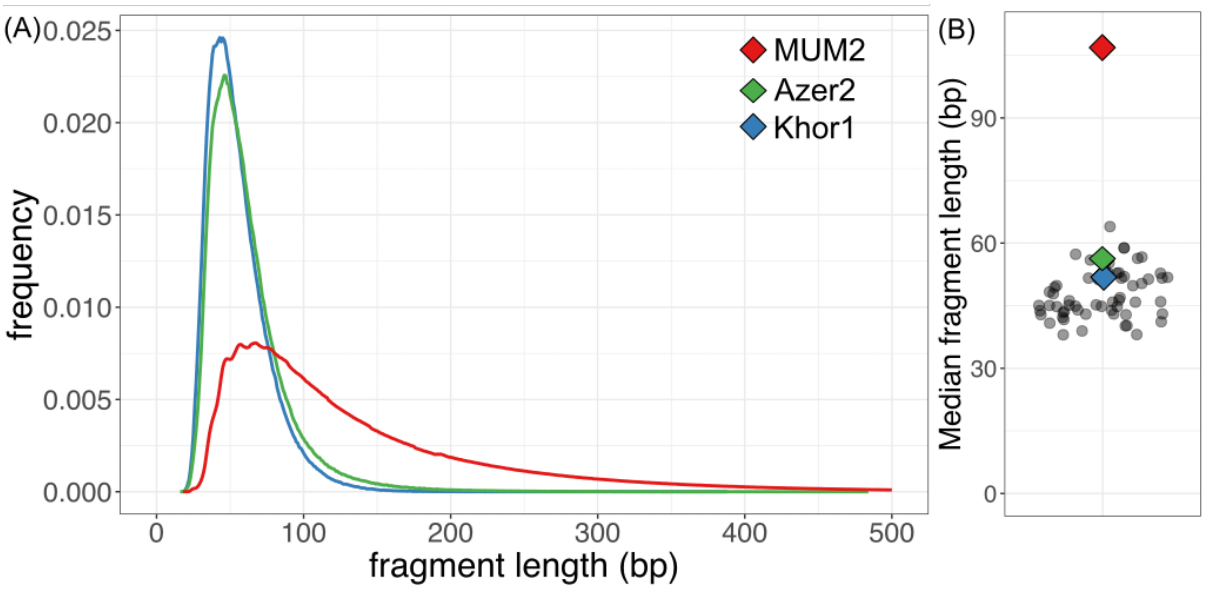


Figure 4.2 - (A) Read length distributions of MUM2, Khor1, and Azer2, calculated from PE data. MUM2 displayed a substantially reduced rate of fragmentation. The median read length of MUM2 (107 bp) exceeded the median read length of Khor1 and Azer2 (52 bp & 56 bp, respectively). **(B) Median read lengths of 61 published ancient Ovicaprid samples** (Daly et al. 2018)**.** The median read length of MUM2 (107 bp; 90 bp among collapsed reads only) exceeded all ovicaprid genomes (64 bp).

MUM2 also displayed substantially lower rates of deamination compared to Khor1 and Azer2 (Figure 4.3A). The mean δS (single strand cytosine deamination probability) was 0.012 in MUM2 vs. 0.382 and 0.334 for Khor1 and Azer2, respectively. The mean δS of MUM2 (0.012) was singular in its low levels of deamination compared to the mean δS of 182 published ancient bone samples (Figure 4.3B).

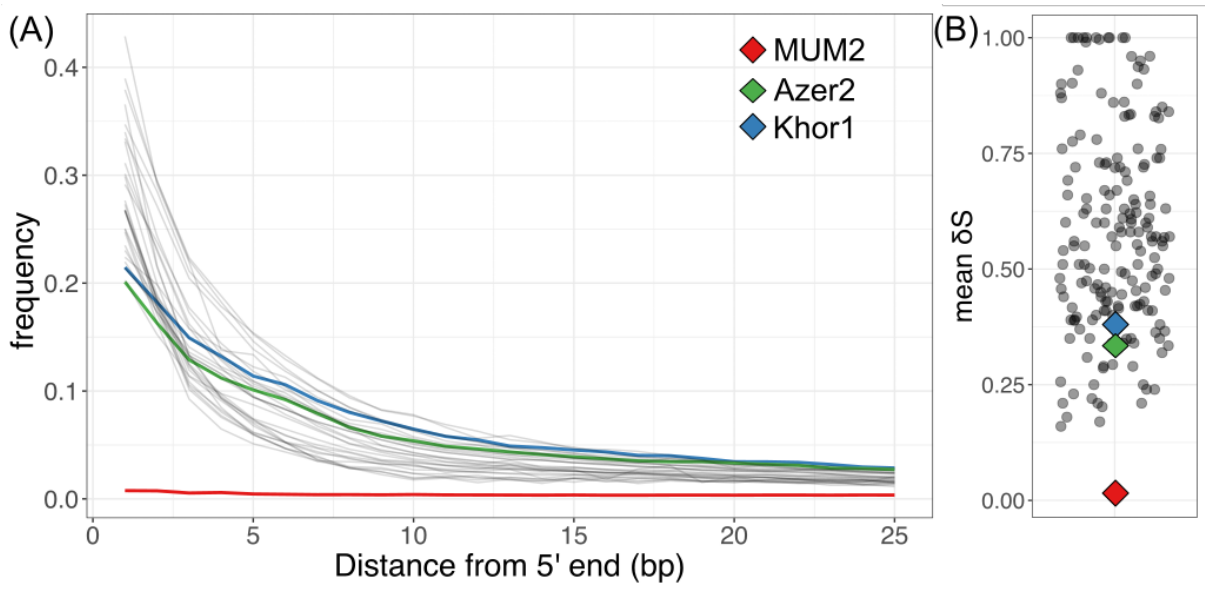


Figure 4.3 - (A) Deamination patterns of MUM2, Khor1, Azer2, and other ancient ovicaprids for non UDG-treated libraries. Low levels of base misincorporation at the 5' ends of reads were observed for MUM2 compared to Khor1 and Azer2. **(B) Mean δS of 182 ancient bone samples** (Daly et al. 2018; Kistler et al. 2017)**.**

To investigate if tissue providence was responsible for the lowered deamination rate, the mean δS of published ancient skins was compared to MUM2 (Figure 4.4). MUM2 displayed lower damage rates compared to all samples, including some that are just ~50 years old.

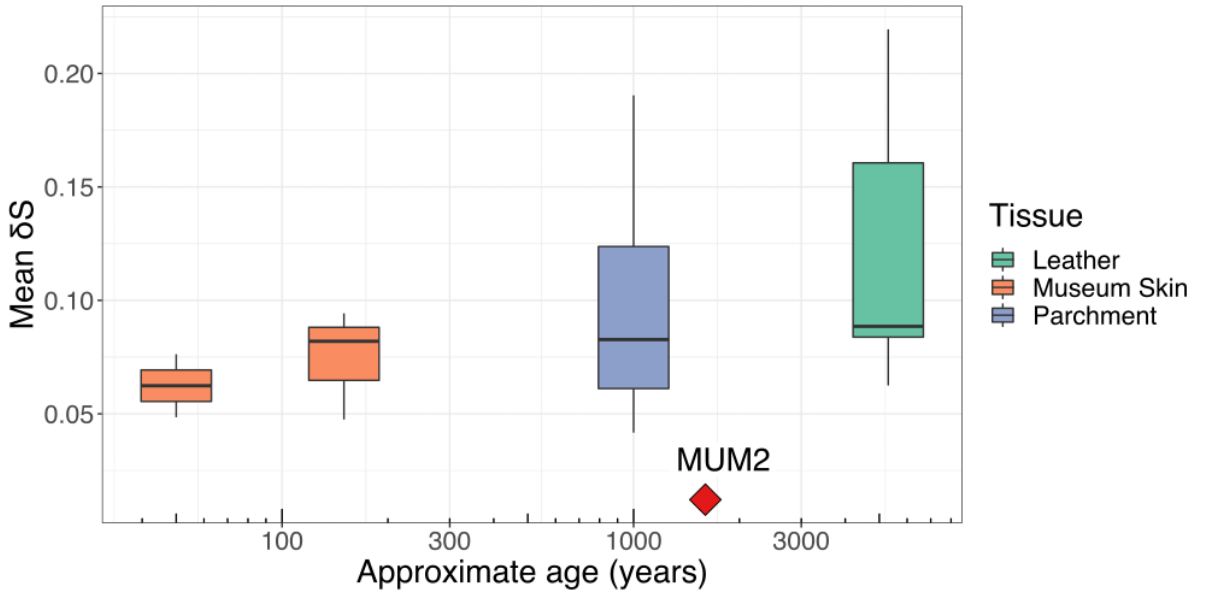


Figure 4.4 - Comparison of mean δS of published ancient skins.

4.4.2. Microbiome

In total 202,815,624 metagenomic reads were retained for MUM2 post-filtering (176,858,319 collapsed/single end reads and 25,957,305 paired end reads). Kraken2 classified 11.96% (24,247,293 reads) of the unmapped, deduplicated reads using the default database. 57.13% (13,499,165) of the classified reads belonged to the salt-adapted *Halobacteria* Class. Using the same methods, *Halobacteria* were found in very low abundance in our comparative bone samples, Khor1 and Azer2 (0.7% and 0.67% of assigned reads, respectively with a threshold of 0.1%). As a comparison with other ancient skins, metagenomic analysis of the Tyrolean Iceman's preserved clothes (O'Sullivan et al. 2016b) and ancient parchment (Teasdale et al. 2017) also revealed very low levels of *Halobacteria* present (0-0.02% and 0-0.3% of assigned reads, respectively).

Table 4.3: Taxonomically assigned reads of MUM2 by Kraken2, recalculated by Bracken, at Class level. The salt-adapted Class *Halobacteria* predominates with 57.13% of assigned reads.

Class	Total Reads Assigned	% of Assigned Reads	Domain
Halobacteria	13499165	57.13	Archaea
Actinobacteria	3879524	16.42	Bacteria
Clostridia	1243905	5.26	Bacteria
Alphaproteobacteria	914412	3.87	Bacteria
Gammaproteobacteria	823160	3.48	Bacteria
Betaproteobacteria	659427	2.79	Bacteria
Bacilli	295377	1.25	Bacteria
Other and Host	155083	8.54	N/A

Microbial source prediction of the metagenome of MUM2, Azer2 and Khor1 was done using SourceTracker2 and a custom metagenomic database at both species and genus level. The "Laboratory reagent" source was not usable at genus level due to an insufficient number of taxonomically assigned genuses (27) for the rarefaction level (1,000). At species level, SourceTracker2 estimated 0.4725 of the microbial community of MUM2 originated from a salt-rich environment, while only 0.003-0.007 of the comparative bone samples were estimated to be from this environment (Table 4.4; Figure 4.4A). This origin estimation of a salt-rich environment increased to 0.746 when estimated at genus level (Table 4.4; Figure 4.4B).

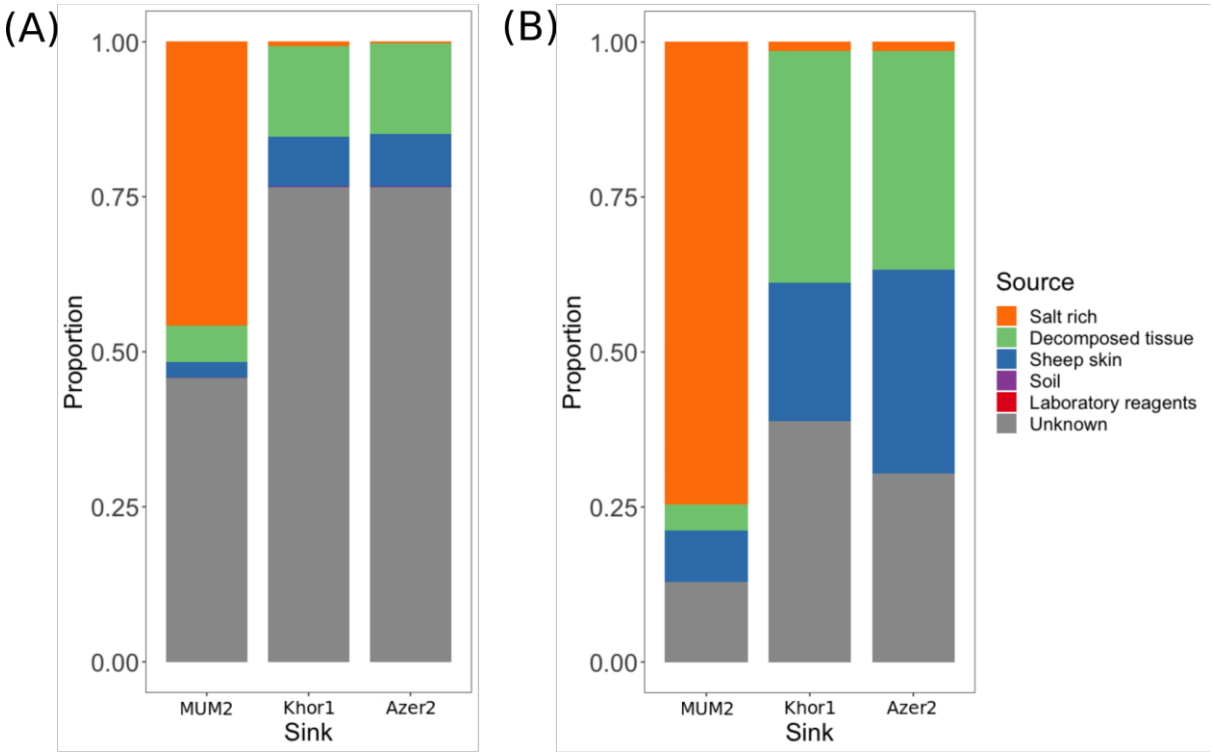


Figure 4.4 - Prediction of metagenomic source proportions by SourceTracker2 at (A) Species level and (B) Genus level.

Table 4.4: Predicted source proportion of metagenomic reads by SourceTracker2.

	MUM2 (Species)	Azer2 (Species)	Khor1 (Species)	MUM2 (Genus)	Azer2 (Genus)	Khor1 (Genus)
Tissue decomposers	0.0212	0.1458	0.0386	0.0426	0.3524	0.3727
Salt-rich	0.4725	0.0026	0.0036	0.7458	0.0145	0.0151
Laboratory reagents	0.0003	0.0002	0.0002	N/A	N/A	N/A
Sheep skin	0.0285	0.085	0.0463	0.0829	0.3294	0.2244
Soil	0.0009	0.0007	0.0046	0	0	0
Unknown	0.4766	0.7657	0.9067	0.1287	0.3037	0.3878

Using MIDAS, a large proportion of detected microbial species in MUM2 (8 of the 9 most abundant species) were observed as being salt-adapted or salt-tolerant. Of these microbes, just one species (*Streptomyces sp 59965*) is identified in ancient sample Azer2, which is a generalist. None of the defined salt-adapted bacteria in MUM2 were detected in shotgun data of leather from the Tyrolean Iceman while salt-tolerant species were found in low abundance (<2%) of several ancient parchment shotgun data except for *Saccharomonospora glauca 62535* which was found at 4% abundance in sample SAMEA104143134. To validate the

authenticity of the most abundant bacterium detected in MUM2, all filtered metagenomic reads were aligned to GCF_00371785.1, an *Actinopolyspora halophila* reference genome (Actinopolyspora halophila DSM 43834, retrieved from NCBI GenBank). A 0.33X genome was generated to test the validity of this sample (Warinner et al. 2017). An even coverage of this bam file was visualised (std per bp = 1.13X) and the median read length of the aligned reads was 74 bp.

Controls that were done throughout the sample preparation, extraction and library preparation were also screened for halophilic signatures. The number of reads taxonomically assigned by Kraken2 was low (643 - 4648 assignments) and in all the presence of the *Halobacteria* clade was low (0.009-0.01). Using SourceTracker2, the salt-rich environment was not predicted to contribute highly to any of the controls (0.003-0.024). After "Unknown", "Laboratory reagents" were predicted to be the highest contribution to all controls except the Library Control. Here, SourceTracker2 predicted 0.66 of its metagenome to belong to "Sheep skin", suggesting possible skin contamination during library preparation.

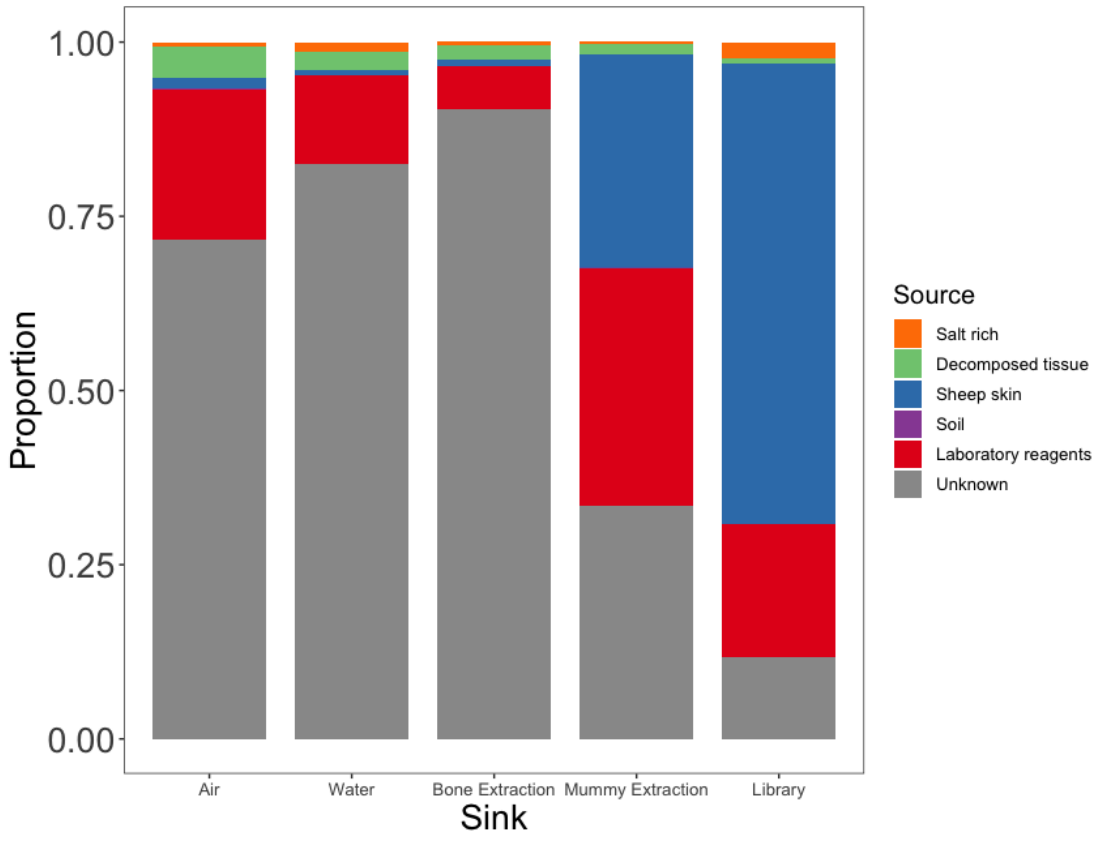


Figure 4.5: Prediction of metagenomic source proportions of controls by SourceTracker2 at Species level.

4.4.3. Population Genetic Analysis

The ancestral profile of the Chehrābād sheep MUM2 was investigated in the context of modern mitochondrial and autosomal variation. A 664X mitochondrial genome of MUM2 was generated. The consensus fasta of MUM2 falls within the C haplotype cluster in a ML phylogeny of modern sheep mitochondria.

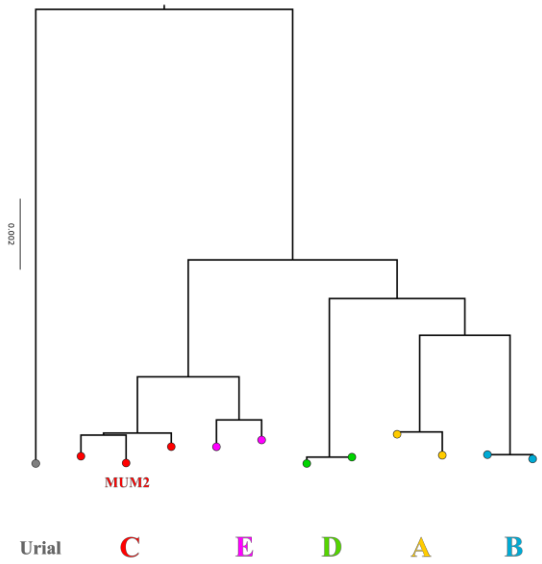


Figure 4.6: Whole mitochondrion maximum-likelihood phylogeny of domestic sheep (100 bootstraps). MUM2 falls within Clade C. All clades of the phylogenetic tree are well-supported by bootstrap values (>0.97).

The LASER PCA revealed that PC1 differentiated modern European breeds from breeds of the rest of the world (Figure 4.7). PC2 differentiated non-European breeds, with a cline from African, Asian and South West Asian appearing. These breeds did not create discrete clusters, however, and some individuals from different locations can be seen to cluster together in geographically-heterogeneous groups. MUM2 fell within the main cluster of Southwest Asian breeds. Khor1 and Azer2 both project within the diversity of SW Asian breeds but fall away slightly from this main cluster on PC1. However, the uncertainty of PC1 of Khor1 and Azer2 was larger than MUM2, likely due to the lower SNPs called.

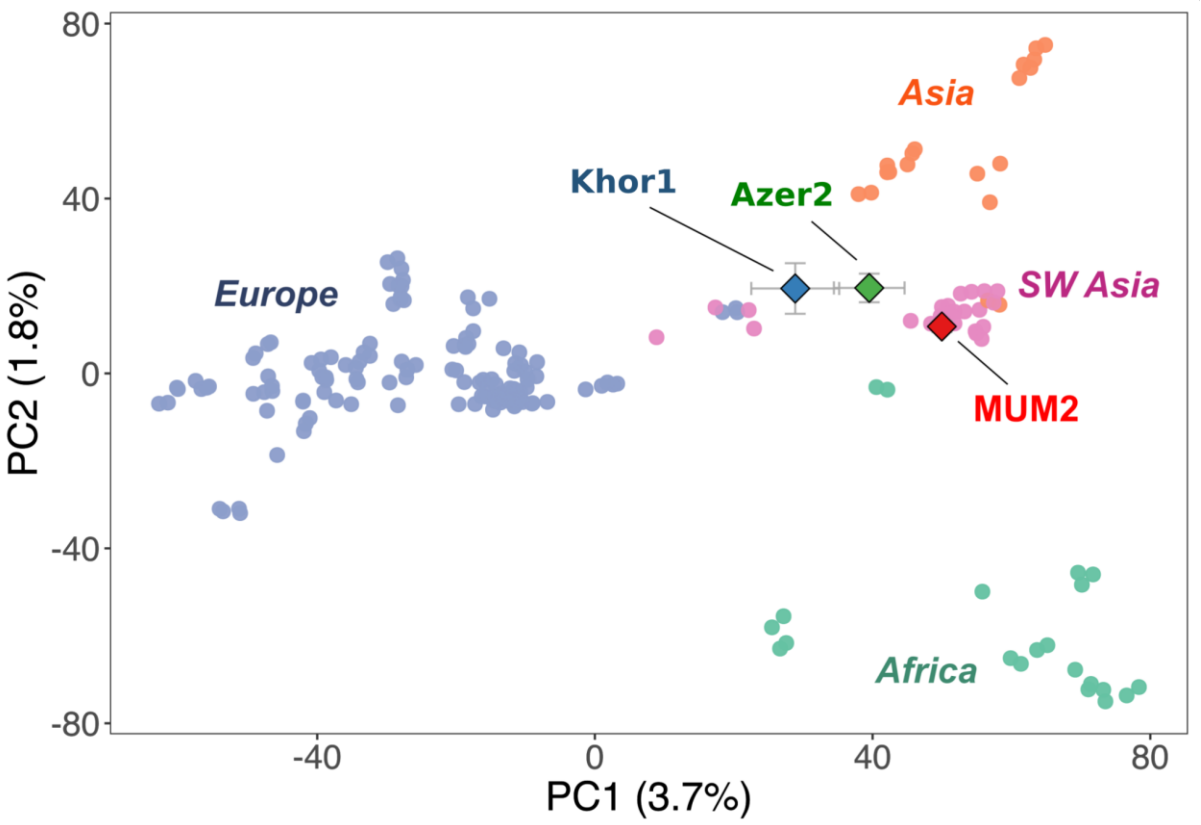


Figure 4.7 - PCA of global sheep affinity and Procrustes projection of MUM2, Khor1 and Azer2. Error bars represent the standard deviation of each PC coordinate between independent runs * 2. PC1 separated breeds of European origin and the rest of the world, while PC2 separated Asian and African breeds into an approximate cline of Central/East Asian, Southwest Asian, African (Kijas et al. 2012).

To investigate local affinity, a LASER PCA was produced using only breeds of SW Asia and Asia using LASER v2.03. Here, PC1 differentiated Asian and Southwest Asian breeds. PC2 then separated one Southwest Asian (Turkish) breed from all other Southwest Asian breeds. MUM2 and Khor1 clustered with Southwest Asian breeds rather than those from Asia (Figure 4.8). Azer2 falls away from this cluster of Southwest Asian breeds, however, there is a larger amount of uncertainty.

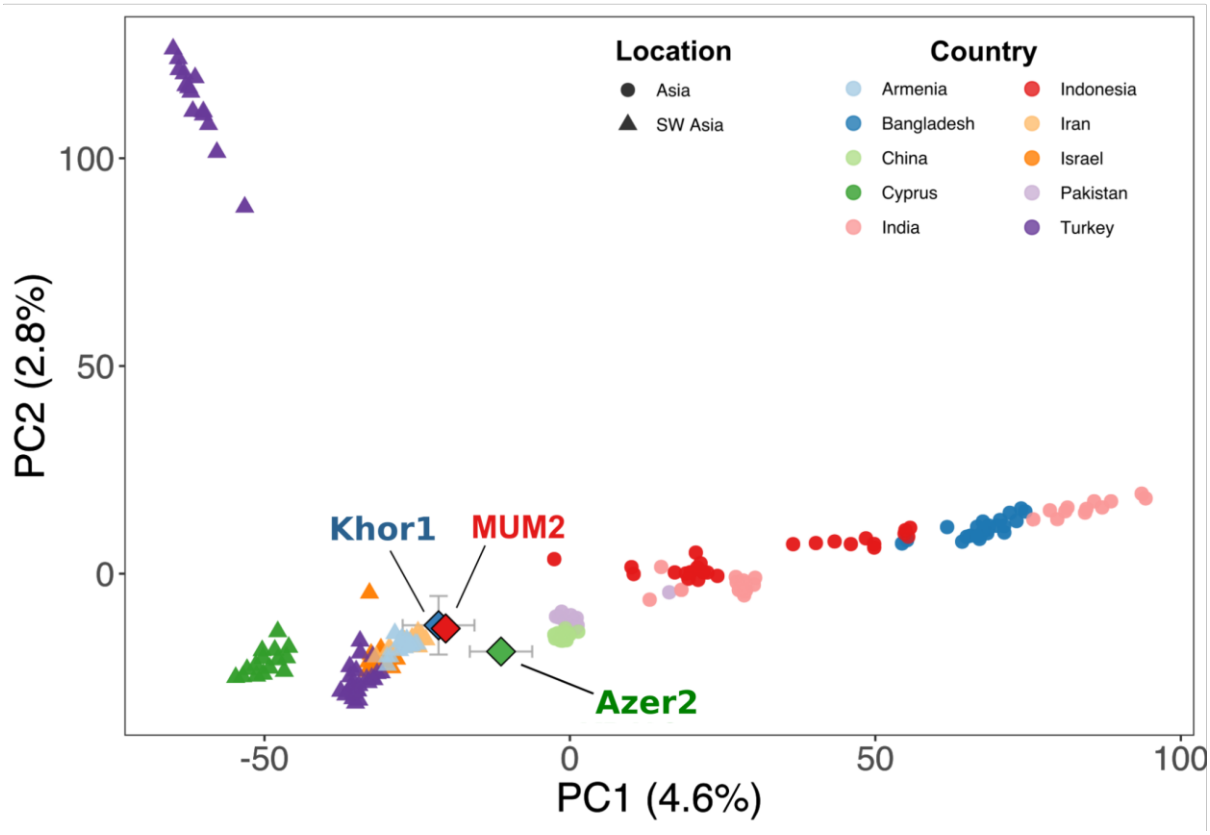


Figure 4.8 - PCA containing only Asian breeds. Error bars represent the standard deviation of each PC coordinate between independent runs * 2.

Using f_3 outgroup statistics, MUM2 displayed the most genetic drift sharing with modern breeds of Southwest Asia (Figure 4.9). A similar trend was observed in the K=13 ADMIXTURE analysis; the predicted ancestral profile of MUM2 most resembled Qezel, a sheep breed local to Iran (Figure 4.10). In TreeMix analysis, MUM2 formed a clade with modern Southwest Asian, reiterating this genetic affinity (4.11).

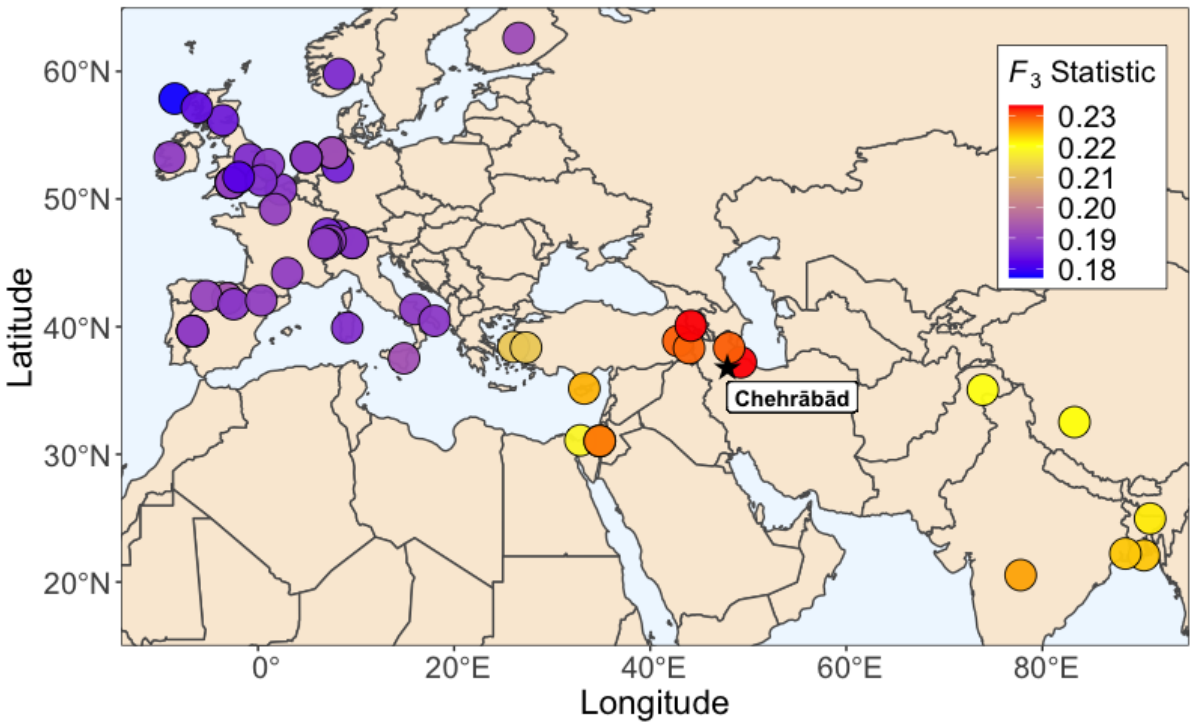


Figure 4.9 - Shared genetic drift between MUM2 and modern sheep populations. Higher f_3 values, in red, indicate higher shared drift, relative to the outgroup Asiatic Mouflon.

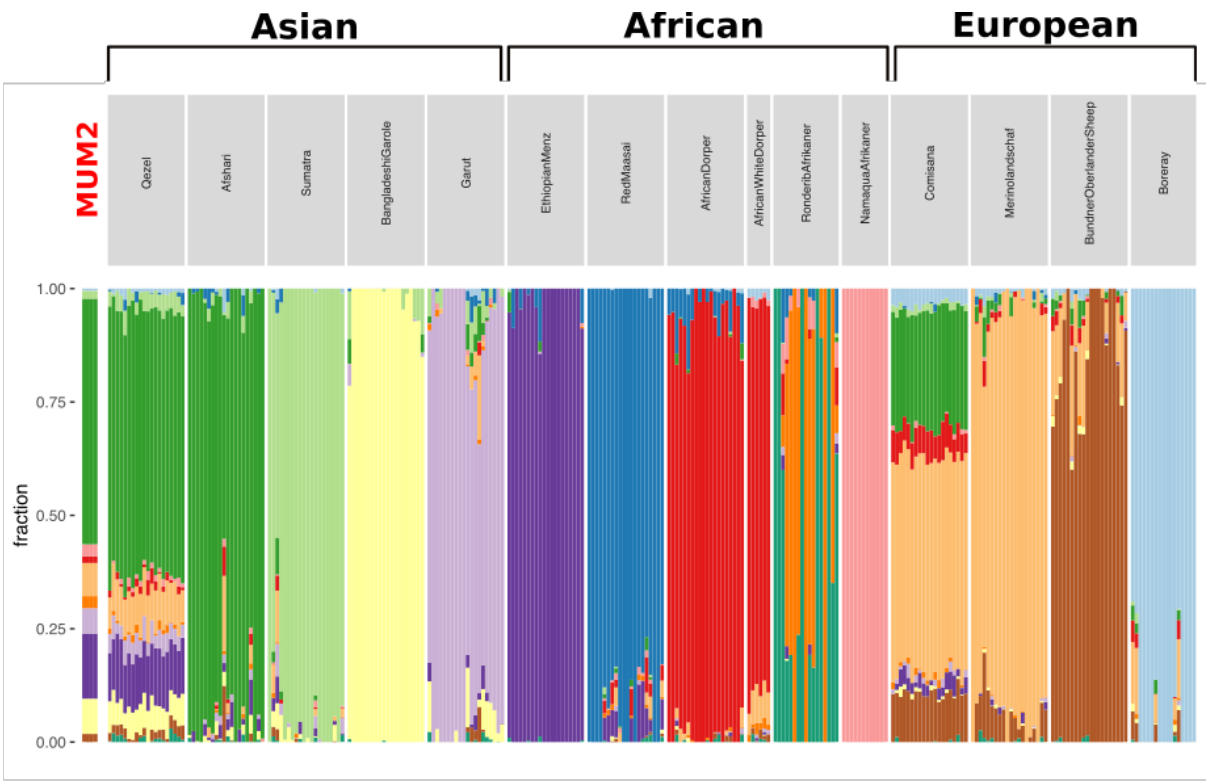


Figure 4.10 - ADMIXTURE plot with K=13 ancestral components. MUM2 displayed a similar ancestral profile to modern Iranian sheep, represented by the Qezel breed.

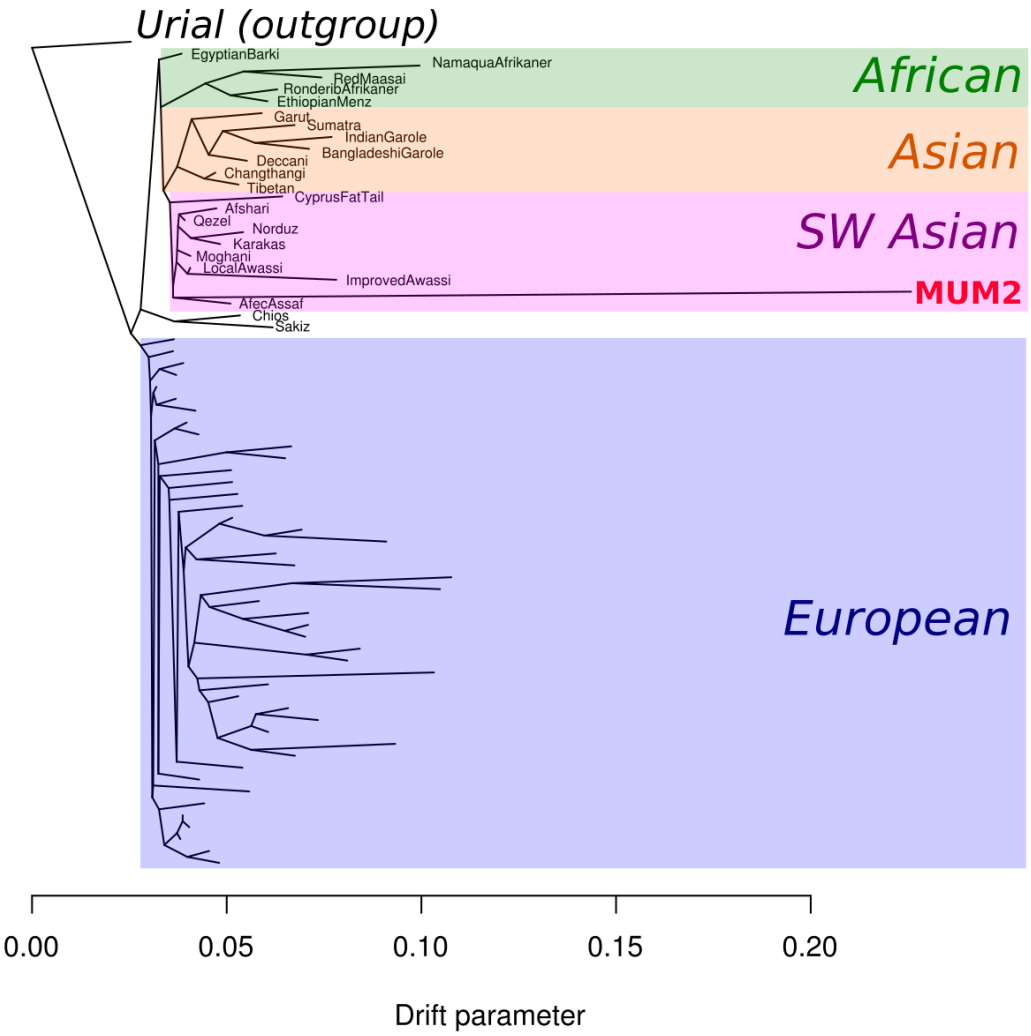


Figure 4.11 - TreeMix plot with 0 migration edges. MUM2 fell into the Southwest Asian breed clade. European breed labels were removed for clarity.

4.4.4. Genotyping traits

The length of the MUM2 aDNA fragments were exploited to investigate the “woolly” locus. The bam files were inspected of the two modified genome assemblies, searching for reads that overlapped with the diagnostic insertion breakpoint of either the “hairy” or “woolly” allele. No reads were found to overlap the diagnostic insertion breakpoints of the “woolly” insert (Figure 4.12A). Five reads were found to uniquely map to the ‘hairy’ allele diagnostic position, with a further two read pairs inferred to overlap this breakpoint (Figure 4.12B). This strongly suggests that MUM2 possessed the ancestral hairy allele.

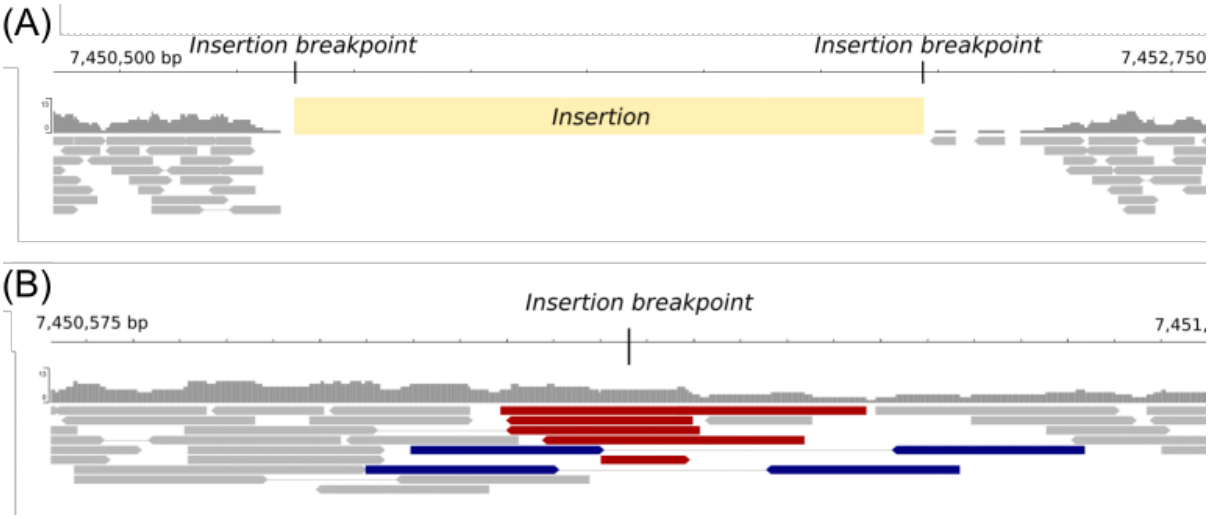
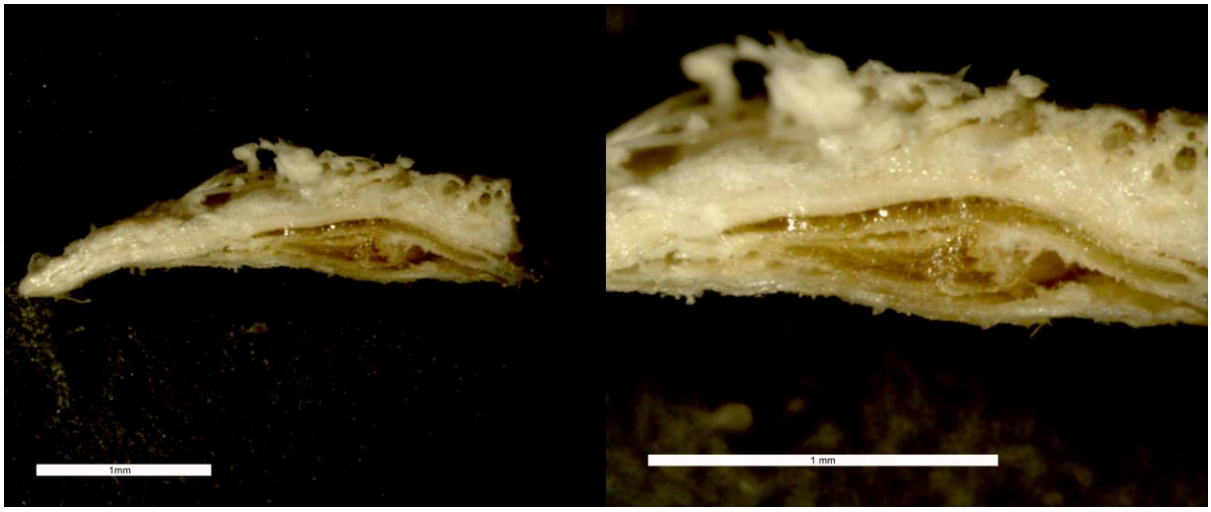


Figure 4.12 - Visualisation of read coverage of filtered bam files at woolly locus, in assemblies with (A) and without (B) the insertion. Reads highlighted in red overlapped the insertion breakpoint, blue indicated an inferred overlap of the insertion point by the straddling read pair. Highlighted reads mapped only to one assembly and did not align to the other.

Evidence of a fat-tail associated allele in MUM2 was also found at *PDGFD*, a gene likely controlling tail phenotype (Dong et al. 2020; Li et al. 2020b). Of 51 SNPs identified, 48 recorded the presence of the derived fat-tail allele. Of these 48, 14 sites were heterozygous; containing both the ancestral and derived allele base calls. While 11 of these sites are susceptible to misdiagnosis due to deamination-induced base transitions, it is also noted that the very low *post-mortem* deamination levels of MUM2 make spurious calls unlikely. One site (ss1139270994,ss1215832245,ss1198779144) contained one base call of a third allele (G) but was otherwise homozygous for the derived fat-tail allele (T) with four base calls (Appendix Table 4.5).

4.4.5. Microscopy analysis

Images produced by reflected light and transmitted light microscopy revealed a fat layer between the grain and corium, which is characteristic of domestic sheep skin (Michel 2014; Ruß-Popa 2018) (Figure 4.13). The leg of the sheep is dominated by coarse fibres (primary fibres), which appear mostly unpigmented (Figure 4.14A).



Supplementary Figure 4.13 - Cross section of the sheep skin. Fat is deposited between the corium layer and the grain. Images by G. Ruß-Popa.

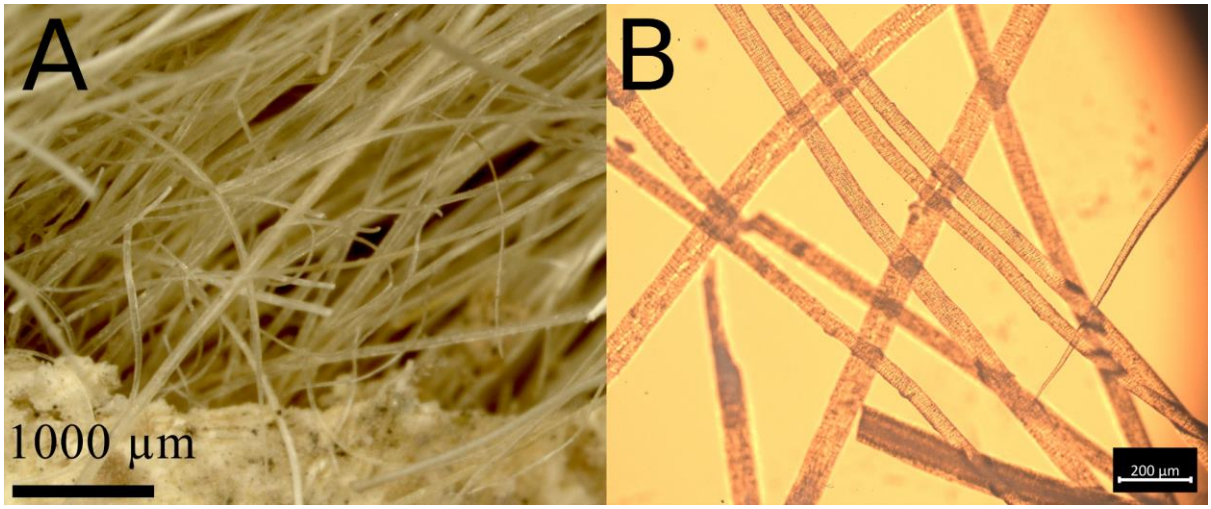


Figure 4.14 - (A) Microscopy displaying the thin fibres in front. (B) Light microscopy revealing the unmedullated underwool. Images by G. Ruß-Popa.

Scanning Electron Microscope (SEM) imagery of these hair fibres showed typical structure of sheep hair fibres, with a mosaic type scale and the fine lines on the scale surface (Figure 4.15A). This is characteristic of sheep hair fibres, and particularly for mouflon and medium-wool breeds. Light microscopy of these coarse fibres revealed an amorphous medulla (Figure 4.15B), with thin unmedullated fibres (underwool) were also detected (Figure 4.14B).

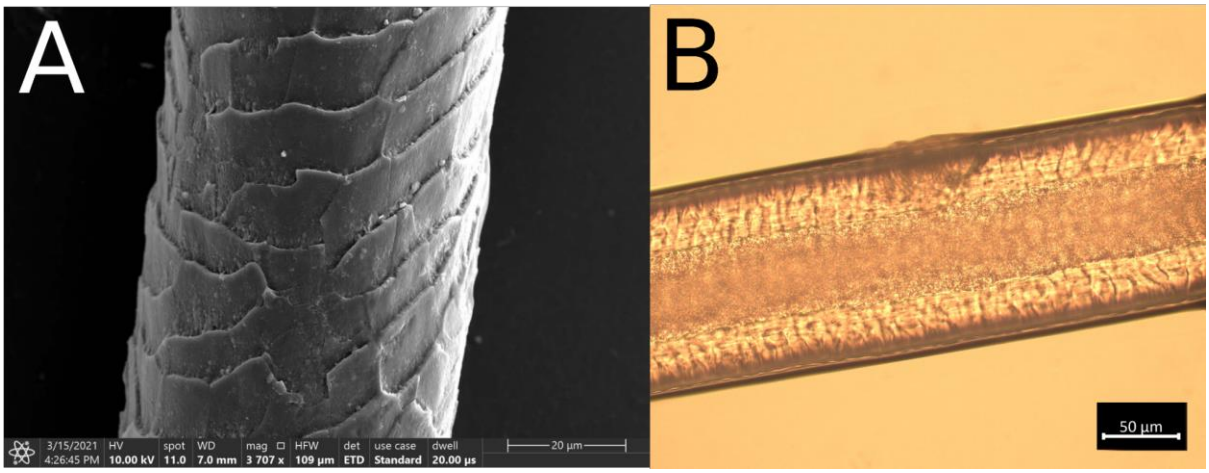


Figure 14.15 - SEM image of MUM2 hair fibre, displaying the typical mosaic scales of a sheep hair shaft and details of the fine lines on the scale surface. Image by A. Steiger-Thirsfeld and G. Ruß-Popa. **16B Light microscopy of the coarse hair fibre displaying the amorphous medulla.** Image by G. Ruß-Popa.

4.5. Discussion

4.5.1. DNA Preservation

Comparative analyses of published ancient sequences revealed the remarkable DNA integrity of MUM2, with hallmarks of post-mortem damage, fragmentation and hydrolytic deamination substantially reduced. It has been suggested that post-mortem DNA fragmentation is driven by rate-constant hydrolytic depurination over time (Allentoft et al. 2012), However, the depurination rates of MUM2 were similar to the more-fragmented comparative samples. Other models suggest that aDNA fragmentation is age-independent, driven by environment-dependent biotic and abiotic factors (Kistler et al. 2017). For example, the highly alkaline, cool and anhydrous conditions of the Chehrābād environment may have contributed to inhibition of cellular nucleases which would otherwise degrade and fragment endogenous DNA (David 2008). In particular, low levels of thermal fluctuation may have been key to aiding in DNA preservation as comparable fragment length distributions have been reported primarily from cave and permafrost environments (Broushaki et al. 2016; Miller et al. 2012; Orlando et al. 2013; Librado et al. 2015; Sikora et al. 2019). DNA preservation may also be influenced by its tissue-of-origin as bone hydroxyapatite rather than keratin fractions are associated with smaller fragment size (Schwarz et al. 2009).

Post-mortem DNA deamination via cytosine hydrolysis is thought to be strongly correlated with age (Sawyer et al. 2012) or thermal age (Kistler et al. 2017). Despite its age, the deamination

rates of MUM2 were substantially lower than any published ovicaprid genome, which implicates the unique taphonomic environment. As deamination is a hydrolytic reaction, the scarcity of environmental free water in the highly saline Chehrābād may have been conducive to lowered deamination rates. As hydrophobic keratinised tissue may provide resistance to environmental water (Gilbert et al. 2004), MUM2 was also compared to published ancient skins genomes to determine if tissue providence was responsible for DNA preservation. The mean δS of MUM2 falls outside the range of other ancient skin genomes, including 20th century CE goat skins (Cassidy et al. 2017) and leather recovered from the Tyrolean Iceman (O'Sullivan et al. 2016). While this does not discount keratinized tissue being specifically enriched with longer DNA fragments, the Chehrābād sheep mummy appears to be singular in its DNA integrity related to deamination among published ancient skin samples.

4.5.2. Microbiome

The distinctive salt-rich geochemical composition of Chehrābād was reflected in the metagenomic profile of MUM2. Taxonomic assignment and abundance estimation assigned 57.13% of classified reads to the halophilic Class of Archaea *Halobacteria* (Table 4.3). Similarly, SourceTracker2 predicted that 0.4725 - 0.7458 of the microbial community originated from a salt-rich environment (Table 4.4, Figure 4.4). A complementary analysis using MIDAS identified 76 unique bacterial species in the mummified sheep, the most abundant species of which being the halophilic bacterium *Actinopolyspora halophila 58532,* accounting for ~29% of identified reads. This signal of a dominant halophilic microbial community was not replicated in comparison samples or controls. Taken together, these results supported the observation that the unique high salinity taphonomic context of Chehrābād is reflected in the microbial profile of MUM2. Rapid colonisation by saprophytic microbial communities, with key decomposers being ubiquitous across soil types, is typical for mammalian corpses post-mortem (Metcalf et al. 2016). The halophilic metagenome profile observed in the Chehrābād sheep mummy skin indicates that the typical decomposers may be less abundant in this alkaline, salt-rich setting, which may have contributed to soft tissue and molecular preservation.

4.5.3. Population genomics

Population genomic analysis consistently depicts high genetic affinity of MUM2 to modern Southwest Asian breeds. The C haplotype harboured by MUM2 is also at its highest frequency in modern Southwest and East Asia breeds (Pereira et al. 2006; Chen et al. 2006), and has been reported in ancient samples from Bronze Age Türkiye (Demirci et al. 2013). f_3 outgroup statistics show MUM2 shares the most genetic drift with southwest Asian breeds, particularly those from Iran while ADMIXTURE and TreeMix analysis also confirmed the affinity of MUM2

with modern sheep breeds from Southwest Asia (Figure 4.9-11). Overall, there is genetic continuity between west Iranian sheep populations in Sassanid and modern time periods, although PCA using Ovine SNP50 genotypes of Asian breeds place MUM2 slightly apart from sampled breeds, suggesting a degree of genetic flux during the past 1500-1600 years in Iranian sheep (Figure 4.7-8). Additionally, the comparatively older sample Azer2 falls further away from the main cluster of Southwest Asian breeds toward East and South Asian breeds (Figure 4.8). While based on low coverage data, this is consistent with evidence for genetic exchange across Asia prior to the development of modern breeds (Chessa et al. 2009; Zhao et al. 2017; Lv et al. 2015).

4.5.4. Fibre Genotype and Phenotype and Fat-tail Genotype

Genotyping the allele at the “woolly” locus revealed that MUM2 was either homozygous or heterozygous for the dominant “hairy” allele. In addition, SEM imaging of the mostly-unpigmented mummified hair fibres revealed mosaic scales typical of sheep (Rast-Eicher 2016) with fine lines on the scale surface (Figure 4.15A), a characteristic of sheep hair fibres and particularly for mouflon and medium-wool breeds (Rast-Eicher 2010). Overall, these results are consistent with MUM2 having a hairy/medium-wool phenotype, although results are not conclusive as the fibre types of the lower leg do not reflect the fleece type from the sheep body. There was also strong evidence of a fat-tail associated allele (48/51 SNPs) at *PDGFD*, a gene likely controlling tail phenotype (Dong et al. 2020; Li et al. 2020b). This observation, along with MUM2 sharing mitochondrial haplotype C with the majority of modern fat-tailed breeds (Lv et al. 2015), and the genomic affinity of MUM2 to modern fat-tailed breeds, although based on SNP-chip data, is intriguing. While the MUM2 tail phenotype cannot be determined directly, its genotype is similar to a medium wool or hairy-coated fat-tail breed (Aali et al. 2012). Hairy-coated sheep may have lower mortality rates, have higher birth weights, and be more robust than woolly-coated (Allain et al. 2014), while fat-tail breeds are thought to be better adapted to arid environments (Dong et al. 2020). If phenotypically-similar to these sheep breeds, the flock represented by MUM2 could have provided a reliable meat and fat source for Chehrābād’s miners. The faunal assemblage of the mine and paleoparasitological studies, although not very abundant, support the fact that sheep/goat were the most consumed animals by the miners (Nezamabadi et al. 2013; Mashkour 2014; Mashkour 2013; Mashkour, Fathi, and Davoudi 2020). Taken together, MUM2 may have come from a herd maintained for meat or milk production rather than wool, consistent with suggestions that ovicaprids were used as food for workers, and that sections of the mine were used as stables (Aali et al. 2012).

Both woolly and fat-tailed sheep are depicted in the Early Bronze Age Mesopotamia but the spread of these phenotypes may have been uncoupled, and occurred via distinct processes (Pöllath, Schafberg, and Peters 2019; Vila et al. 2021). Fat-tailed breeds were likely introduced from Southwest to East Asia in a period (700 BCE-1000 CE) broadly coinciding with the age of MUM2 (Zhao et al. 2017); the observed *PDGFD* genotype supports an ancient origin of this economically important trait. Wider aDNA analysis may elucidate when wool and fat-tailed associated genotypes arose and how they may have influenced sheep breed development, which have their origins in 4th millennium BCE Mesopotamia (Pöllath, Schafberg, and Peters 2019; Vila et al. 2021). Although the archaeozoological assemblages in the Iranian Plateau from the Antiquity and later Mediaeval periods are still limited, the diversity of the size of sheep bones is already an indication of the diversification of breeds in these periods (Khazaeli 2013; M. Mashkour 2013). These results are consistent with MUM2 deriving from a herd used for meat and/or milk rather than wool production, and reflect sophisticated Sasanian-period husbandry practices and specialised sheep breeding.

5. Preservation, metagenomics and population genomic analysis of ancient Levantine goat coprolites

5.1. Introduction

Mineralised tissues such as teeth and bone are the preferred material for ancient genomic studies. Bones and teeth tend to be better identified and preserved in archaeological assemblages, and are well-documented as rich sources for endogenous aDNA. However, as the scope of aDNA studies widens, other underutilised ancient materials are increasingly being considered as a substrate. One such material is fossilised dung or coprolite (Green and Speller 2017). Coprolite studies generally involve interdisciplinary analysis to reconstruct information about the diet and health of the depositing animal (Shillito et al. 2020). While still relatively uncommon, coprolites have also been analysed for aDNA, especially in archaeological assemblages where bone and teeth are not present (Linseele et al. 2013). For example, there is a scarcity of human bone remains in prehistoric North American sites which has necessitated the use of coprolites in past aDNA studies (Gilbert et al. 2008).

While coprolite aDNA analysis offers an opportunity to increase the understanding of some under-studied populations, a disadvantage is that endogenous aDNA retrieval has been relatively poor. As discussed in Chapter 3, bones of morphologically similar animals are often difficult to distinguish e.g. *Capra* and *Ovis.* Difficulty in identification also extends to their faeces, where coprolites of ovicaprids and other ungulates may be indistinguishable (Linseele et al. 2013). This is further exacerbated by morphological changes through taphonomic processes. While aDNA analysis of coprolites can aid in species identification, it is also complicated by coprolites left by omnivores or carnivores carrying trace DNA of their last meals *e.g.* human faeces also containing *Canis* DNA (Gilbert et al. 2008). This reflects a wider issue of coprolites analysis: aDNA retrieved from coprolites and soil samples are acutely vulnerable to contamination. The effect of soil microbes on coprolite metagenomics varies between individual samples, but can obscure detection of ancient microbiomes (Tito et al. 2012). False genomic signals can also occur due to leaching of DNA from layers above the dated stratum. This is exemplified in a study that detected ovine aDNA in strata that predate the introduction of sheep to New Zealand by several hundred years (Haile et al. 2007). As the number of

published coprolite sequences is sparse, little research has investigated the preservation of endogenous DNA in coprolites.

While coprolites can act as a source of endogenous DNA for population genomic analysis, an attractive advantage is that ancient microbiomes may also be inferred from the analysis of the associated exogenous aDNA found in these samples (Christina Warinner et al. 2015). DNA from microbes in coprolites may, in some part, represent the ancient gut microbiome. Coprolite analysis may therefore offer additional information about the health and diet of ancient and historic populations (Shillito et al. 2020). In recent years, the multi-disciplinary field of paleopathology has increasingly utilised aDNA to expand the understanding of past epidemics, and provide insight into pathogen evolution and host immune reaction (*e.g.* (Bos et al. 2011), (Kay et al. 2014), (Schuenemann et al. 2018)). For ancient gastrointestinal diseases, coprolites may provide the best opportunity to capture and sequence causative pathogens to shed light on zoonotic disease evolution and herd health. In addition to insights into the health of ancient populations, coprolites offer the unique ability to investigate past diets, as faeces contains remnants of the depositing individual's last meal. Reconstructing ancient diets has revealed ancient animal husbandry practices and uncovered behaviours in ecological niches of extinct populations (Wood et al. 2016); Boast et al. 2018).

Ancient Levantine goat management has been documented in the Central Negev, Israel by the presence of ovicaprid coprolites dating to as early as the Late Neolithic (~6200 BCE) (Rosen et al. 2005; Rosen 2013). While few intact mineralised remains have been recovered, ancient dung layers suggest the presence of domestic goat management from the late Neolithic, Bronze Age and recent (300 BP) eras in the locality of Ramon Crater. Analysis of phytolith content of coprolites has suggested that these goat herds were seasonally fed on lush young barley (Hordeum) grasses (Landau et al. 2020). Ancient goat genomes in western (Anatolian), eastern (Iran) and southern (Levant) regions of the Near East were highly structured in Pre-Neolithic and Neolithic periods (Daly et al. 2018). Early managed goat populations were derived from what was likely the wild local population. The population structure breaks down in the post-Neolithic period, when a collapse in mitochondrial structure and admixture of domestic eastern goats into Levantine and Anatolian nuclear diversity is observed. The turnover in goat ancestries is associated with human migrations from eastern regions of the Near East into the Levant. Indeed, aDNA analysis of ancient Levantine humans revealed a mirrored eastern influence by the Bronze Age (Agranat-Tamir et al. 2020). These coprolite samples provide an opportunity to add to the codex of ancient Near Eastern goat genomes as well as offering a local view of domestic ancestries.

5.2. Aims

1. To sequence in ancient DNA extracted from goat coprolites and characterise damage patterns.
2. To interrogate the population genomics of these goats in the context of ancient and modern variation.
3. To investigate the metagenome of the coprolites and explore signatures of disease and diet.

5.3. Materials and methods

5.3.1. Material

Coprolite samples were recovered from Ramon I Rockshelter in Ramon Makhtesh, Negev Desert, Israel. 16 boxes of samples were received, each containing 1-3 coprolite pellets. Samples were dated by stratigraphy to Late Neolithic, Bronze Age and Recent (Rosen et al. 2005; Rosen 2013).

5.3.2. DNA extraction, library construction, capture and sequencing

Sample preparation is detailed in Section 2.1.5. DNA extraction and sequencing library construction are detailed in Section 2.2.4. and 2.3, respectively. DNA concentration was quantified using an Agilent 2200 Tapestation (Agilent Technologies). Libraries were amplified with 14 PCR cycles due to low concentration of amplified DNA.

Based on the screening results, 9 samples were selected for mitochondrial capture. Mitochondrial sequence enrichment was performed using MYcroarray MYbaits® In-Solution Sequence Capture for Targeted High-Throughput Sequencing tool as described in Section 2.3.5. The captured DNA was then reamplified for 16 cycles prior to sequencing.

Three rounds of sequencing were performed on Illumina on the Illumina MiSeq platform (single-end 1 x 63 bp) for screening and mitochondrial capture. One round of sequencing was performed on the Illumina NovaSeq platform (paired-end 2 x 50 bp). All sequencing was done at TrinSeq, the Trinity Genome Sequencing Laboratory, Trinity College Dublin.

5.3.3. Read-processing

After taxonomic identification (Chapter 3), trimmed sequencing reads were aligned to ARS1 using bwa 0.7.5 with relaxed parameters and post-alignment read processing was done

according to Section 2.4.3. Damage assessment and sexing were computed as detailed in Section 2..4.4. and 2.4.5., respectively.

5.3.4. Mitochondrial analysis

Mitochondrial haplotyping was performed by aligning trimmed reads (shotgun and capture) against a goat mitochondrial reference (NC_005044.2). The final consensus fasta sequence was generated with a minimum depth of 3 using ANGSD (angsd -doFasta 2 -doCounts 1 -setMinDepth 3 -minQ 20 -minMapQ 30), with trimming of 5 bp at read ends due to more extensive DNA damage. The resulting genomes were concatenated with genomes of representative haplotypes (Appendix Table 5.1) and aligned using MUSCLE in seaview. A Maximum-Likelihood tree with 100 bootstraps was then generated to determine the mitochondrial haplotype.

When the sample haplotype was determined, trimmed sequencing reads were then aligned to a custom haplotype-specific circularised mitochondrial genome produced by Dr Kevin Daly, Trinity College Dublin. Reads were processed identically to above and consensus fasta files were produced according to Section 2.4.3. with trimming of 5 bp at read ends due to more extensive DNA damage. Consensus fastas were generated with a minimum depth of 3X and 5X using ANGSD for separate BEAST analyses (angsd -doFasta 2 -doCounts 1 -setMinDepth 3/5 -minQ 20 -minMapQ 30).

A dataset of published modern and ancient goat mitochondrial genomes was provided by Dr Kevin Daly and used for BEAST analyses (Appendix Table 5.1). Two datasets were curated: the first was with coprolite samples with a mean coverage above 5X called at a minimum depth of 3X per site. The second was with coprolite samples with a mean coverage above 15X called at a minimum depth of 5X per site.

The dataset was aligned with MUSCLE in seaview and partitions were chosen based on NCBI annotations of NC_005044.2. The partitions were tRNAs, rRNAs, the first and second position of codons (C1+C2), the third position of codons (C3), the D-Loop and rest of the molecule. The tree model for all partitions was linked, while the site and clock models were unlinked. The clock model was set to Relaxed Clock Log Normal for all partitions. The site model was chosen using bModelTest. For ancient samples without direct C14 dating, priors were set based on contextual dating. All coprolite samples were given an age prior between 0 and 15,000 years. For each dataset, two BEAST analyses were run with 250 million chains each.

5.3.5. Autosomal analysis

A SNP dataset and dataset of ancient and modern goat nuclear genomes were provided by Dr Kevin Daly, as reported in (Daly et al. 2021). In total, 213,161 sites were used to call SNPs in coprolites. The ancient genomes used are summarised in Table 5.2.

An Identity-By-State matrix was generated with ANGSD (version: 0.930), as detailed in Section 2.X. Sites were specified using the -sites flag and each site was called by randomly sampling reads. A Neighbour-Joining tree was computed resulting matrix using the R package ape (Paradis, Claude, and Strimmer 2004). The tree was rooted by a sheep sample aligned to the goat reference genome.

Projection PCA using Procrustes analysis was performed using LASER v2.03. The PCA reference space and projection transformation were constructed using the modern breeds SNP dataset. Using the SNP dataset, pileup files were generated for higher coverage ancient Levantine samples using SAMtools mpileup (-q 30 -Q 20). Each sample was projected onto this PCA space and averaged across ten independent runs. Position uncertainty was represented by the standard deviation * 2 for each sample.

5.3.6. Contamination estimates

Contamination was estimated using an edited in-house Python script developed by Dr Lara Cassidy, Trinity College Dublin ("contamito.py"). Given a list of sites, this script produces two metrics: the average site heterozygosity and the average heteroplasmic site heterogeneity. The average heteroplasmic site heterogeneity is an upper estimate of contamination, as it measures the heterozygosity at heteroplasmic sites only.

The mitochondrial genomes were tested for contamination using all sites and haplotypic goat sites as defined by (Colli et al. 2015). The results were compared with contamination estimates of published goat mitochondrial bams published in (Daly et al. 2018).

5.3.7. Microbiome analysis

For all metagenomic analysis, reads originating from the host organism and human contamination were first removed. To do this, all trimmed reads were aligned to the goat genome (ARS1) and then the human genome (hg38), using methods detailed above. Unmapped reads were retained using SAMtools view -f4 and converted to fastq files using SAMtools fastq. The retained reads were then deduplicated using PrinSeq. Unmapped,

deduplicated reads were searched against the Kraken2 standard database using default settings. This database includes sequences of bacterial, archaeal and viral genomes. This process was repeated for the laboratory controls and comparative tissues.

DNA sequences of various relevant metagenomes (foregut digesters, hindgut digester, ruminant digester, soil from Negev desert and laboratory reagents) were downloaded from NCBI (Table 5.1). Ruminant digester microbiomes were defined by approximately 100 modern goat dung samples. The sequences were trimmed using AdapterRemoval v2.1.1 and taxonomic assignment for these reference sequences was done using Kraken2, as detailed above. All resulting kreport files (coprolites, ancient comparative tissues and modern databases) were searched for common disease-causing microbes.

SourceTracker2 was used to investigate if a ruminant gut microbiome was detectable in the metagenomic composition of the coprolites. The sources for SourceTracker2 were the various metagenomes. The kraken reports were merged and converted to a BIOM table using kraken-biom (https://github.com/smdabdoub/kraken-biom) and the human taxid (9606) was removed. Source predictions were done at both Species and Genus level using default settings. The "Laboratory reagent" source at Genus level was unusable as there was an insufficient number of taxonomically assigned Genera (27) for our rarefaction level (1,000). In addition, the Laboratory control sinks had insufficient number of taxonomically assigned reads at Genus level.

Table 5.1 - Metagenomic sources used for SourceTracker2.

Name	N	Reference
Soil	20	(Moreira-Grez et al. 2019)
Hindgut digester	16	(Costa et al. 2012)
Ruminant digester	41	(Haworth et al. 2019)
Monogastic digester	10	(Clayton et al. 2016)
Laboratory reagents	35	(Salter et al. 2014)

5.3.5. Plant matter

Several different methods were used to detect plant matter in the coprolites. The Kraken2 RefSeq database only includes 27 whole-genome Plant sequences so three custom plant databases were made with Kraken2-build: a "genome-skimming" database containing all filtered sequences, a "chloroplast" database containing only full chloroplast sequences and

finally a “trnL” database. Each database was built using the default Kraken2 RefSeq databases available for Bacteria, Virus, Archaea, UniVec_Core and a human genome.

All uploaded Plant sequences between the sizes of 200 and 5,000 bp were downloaded from GenBank using esearch as per (Harbert 2018). In total, 29,145,264 Plant sequences of length 200-5000 bp were downloaded from GenBank. Genome assembly scaffolds, genome assembly contigs and genome survey sequences (gss) were removed leaving 4,064,273 Plant sequences. Of these, 18,389 were *Hordeum*. These sequences were used to build the Kraken2 database for “genome-skimming” along with the RefSeq databases above. A second dataset comprised only full chloroplast genomes with the default RefSeq databases above. In total 19,048 chloroplast genomes were added to the database, 39 of which were *Hordeum*. Finally, a barcoding database was built using 281,211 trnL genes. 277 of these sequences were *Hordeum*. These sequences were used to build the trnL database along with the RefSeq databases above. Unmapped, deduplicated reads were searched against the genome-skimming database using Kraken2 with default settings as above.

5.4. Results

5.4.1. DNA Preservation

PCR amplification and sequencing results are displayed in Table 5.2. Despite relatively high PCR cycling (14 cycles per library), the concentration of DNA was consistently low. Of the 29 libraries produced, 18 (62%) libraries yielded concentrations of DNA that were not detectable by Agilent TapeStation 2100. For this reason, 10 ul of each amplified sample library were added to the screening pool regardless of concentration and the samples were screened on the Illumina MiSeq platform.

For most samples, relatively few reads were returned after sequencing demultiplexing (Table 5.1). The concentration of DNA was positively correlated to the number of reads produced (R^2 = 0.5755, p-value = 0.00109) (Figure 5.1A). The endogenous DNA content of the samples was also low for all samples (0.08-6.65%). Furthermore, there was a risk of inflation of endogenous content associated with low total number of sequencing reads. For example, CR074 had a high endogenous concentration (4.88%) relative to other coprolites but this was calculated based on just 80 aligned reads.

Table 5.2 - Screening sequencing results of 29 samples of coprolites

Sample ID	Era	DNA concentration (ng/ul)	Total trimmed reads	Total reads aligned post QC	Endogenous %
CR073	Late Neolithic	0.178	366919	14470	3.94
CR074	Late Neolithic	0	1641	80	4.88
CR075	Bronze Age	0	955	10	1.05
CR076	Bronze Age	0.623	757302	8825	1.17
CR077	Recent	0.296	447165	5710	1.28
CR078	Recent	0.0872	94246	497	0.53
CR079	Late Neolithic	0.0766	169354	4076	2.41
CR080	Late Neolithic	0	1282	11	0.86
CR081	Late Neolithic	0.114	174106	142	0.08
CR081B	Late Neolithic	0	16906	31	0.18
CR082	Late Neolithic	0	2714	10	0.37
CR083	Late Neolithic	0	7792	33	0.42
CR083B	Late Neolithic	0	1243	2	0.16
CR084	Bronze Age	0.376	74204	907	1.22
CR085	Bronze Age	0.071	11628	24	0.21
CR086	Bronze Age	0.267	10875	47	0.43
CR086B	Bronze Age	0	1178	3	0.25
CR086C	Bronze Age	0	2116	2	0.09
CR087	Bronze Age	0.16	26638	98	0.37
CR087B	Bronze Age	0	33615	812	2.42
CR087C	Bronze Age	0	999	3	0.3
CR088	Bronze Age	0	1815	9	0.5
CR088B	Bronze Age	0	466	1	0.21
CR089	Bronze Age	0	9094	192	2.11
CR090	Recent	0	3580	49	1.37
CR091	Recent	0	6793	27	0.4
CR092	Recent	0.407	88302	5876	6.65
CR093	Recent	0	6547	68	1.04
CR094	Recent	0	2346	17	0.72

The endogenous DNA content of samples with at least 10,000 sequenced reads are displayed in Figure 5.2B. The recent coprolites had the highest average endogenous DNA content, with the Bronze Age coprolites yielding the lowest average endogenous DNA content. The highest DNA content was a recent coprolite, CR092 (6.65%).

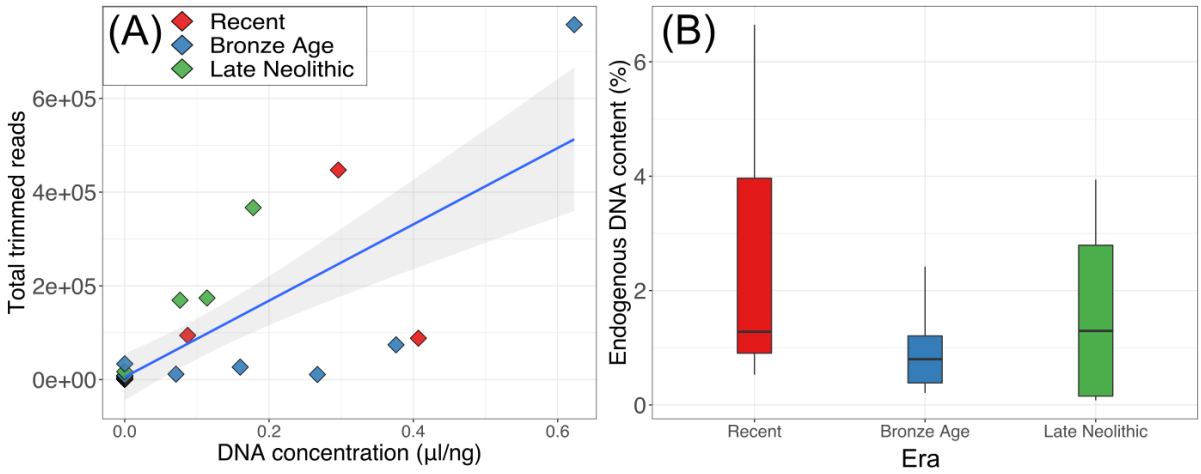


Figure 5.1 - (A) Total number of trimmed reads versus DNA concentration. The concentration of DNA was positively correlated to the number of reads generated. **(B) Endogenous DNA content of coprolites with at least 10,000 reads sequenced**.

Three samples were selected for higher depth sequencing based on age, endogenous DNA content and DNA concentration (Table 5.3). 24 PCR libraries were amplified in total. Clonality in these samples was high across samples, reaching 73% duplicates in one PCR library (Appendix Table 5.2). This is despite each PCR library only producing ~14M raw reads, which falls far below the average aimed yield of 50M reads per library (Chapter 3). Additionally, a high number of raw reads were lost after adapter trimming. On average, 34.56% reads were discarded (Appendix Table 5.2).

Table 5.3 - Coverages of coprolite and published ancient samples post-mapping quality filters.

Sample ID	Site	Location	Region	Era	Coverage
ainghazal4	Ain'Ghazal	Amman	Levant	Neolithic (ca. 8,000 BCE)	0.01
ainghazal1	Ain'Ghazal	Amman	Levant	Neolithic (ca. 8,000 BCE)	0.03
ainghazal2	Ain'Ghazal	Amman	Levant	Neolithic (ca. 8,000 BCE)	0.06
CR079	Ramon Crator	Northern Negev	Levant	Neolithic (6,000 - 6,200 BCE)	0.01
CR073	Ramon Crator	Northern Negev	Levant	Neolithic (6,000 - 6,200 BCE)	0.02
gilat8	Gilat	Northern Negev	Levant	Chalcolithic (4,500 - 4,200 BCE)	0.02
CR076	Ramon Crator	Northern Negev	Levant	Bronze Age (3,000 BCE)	0.01

yarmut7	Tel Yarmuth	Bet Shemesh	Levant	Bronze Age (2,650 - 2,200 BCE)	0.01
yoqneam2	Tel Yoqne'am	Haifa	Levant	Bronze Age (1,650 - 1,540 BCE)	2.13
abdul1	Tepe Abdul Hosein	Luristan	Iran	Neolithic	0.16
abdul2	Tepe Abdul Hosein	Luristan	Iran	Neolithic	0.24
abdul4	Tepe Abdul Hosein	Luristan	Iran	Neolithic	2.31
abdul6	Tepe Abdul Hosein	Luristan	Iran	Neolithic	0.08
lur12	Tepe Abdul Hosein	Luristan	Iran	Neolithic (9977 cal yrBP)	1.01
semnan1	Sang-e Chakmaq	Semnan	Iran	Neolithic (ca 6,000 BCE)	6.61
semnan3	Sang-e Chakmaq	Semnan	Iran	Neolithic (ca 6,000 BCE)	14.44
semnan10	Sang-e Chakmaq	Semnan	Iran	Neolithic (ca 7,000 BCE)	1.42
qazvin1	Tepe Chizar	Qazvin	Iran	Bronze Age (2,400 - 1,900 BCE)	3.13

Mismatch frequency plots generated by mapDamage were uninformative for most samples due to a lack of coverage. However, for samples with at least 500 aligned reads, an unusual pattern of mismatching was observed. For almost all samples, mismatches associated with post *mortem* damage (i.e. C>T at 5' end and G>A at 3' ends) were observed. However, these mismatches were accompanied by all other possible mismatch types at varying frequencies. These mismatch frequencies peaked just prior to the 5' and 3' ends of the fragments, then decreased steadily until they levelled off at a low frequency after around 4 bases. Notably, a single recent coprolite (CR092) did not display these damage patterns. There are some trends observed across samples i.e. T>C transitions are consistently the lowest frequency mismatch.

To test the effectiveness of USER treatment, deamination patterns were plotted at CpG sites only. Here, elevated mismatch frequency for C>T transitions at the 5' end and G>A transitions at the 3' end of DNA fragments was observed, which suggests that USER treatment was efficient at removing age-related deamination. The observed damage patterns are higher in the Neolithic samples (CR073, CR079) compared to the Bronze Age sample (CR076).

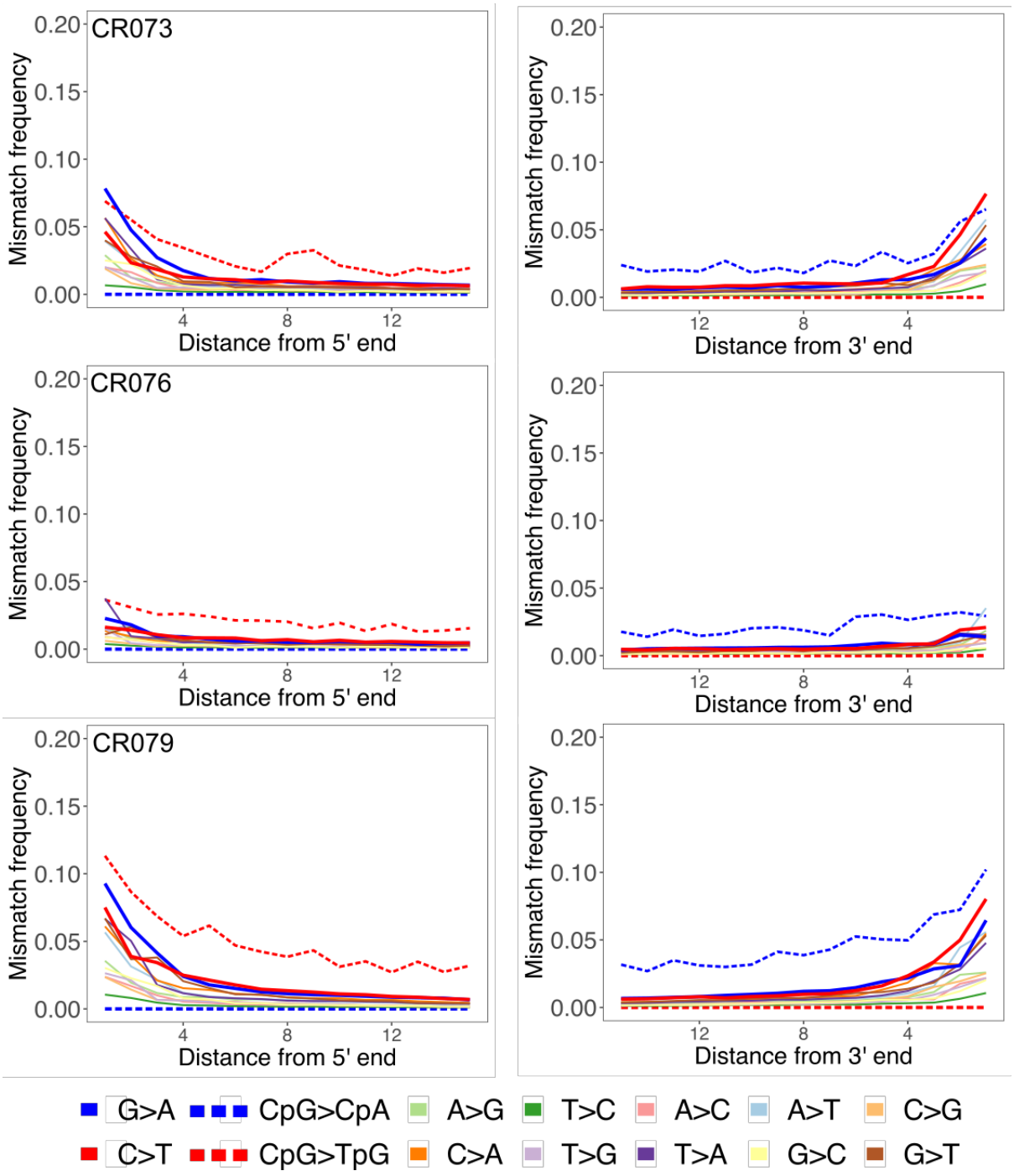


Figure 5.2 - Mismatch frequencies of USER-treated coprolite libraries. Mismatches are observed for every possible transition and transversion. The mismatches peak the base just prior to the strand break and decrease steadily. Deamination patterns of C>T and G>A transitions at CpG sites only are also plotted. The mismatch frequencies are elevated compared to the frequency of transitions at all sites.

5.4.2. Mitochondrial enrichment and analysis

Mitochondrial enrichment was successful for 8 samples with an average enrichment of ~247X (Appendix Table 5.3). All samples returned domestic haplotype A, which is the most common haplotype in modern goat populations.

The BEAST runs of the “minimum 3X coprolite depth” dataset converged after 250 million chains with all parameters reaching Effective Sampling Size (ESS) of >300 except for the rate of C1+C2 which displayed poorer mixing in both runs (ESS of ~100). The BEAST runs of the “minimum 5X coprolite depth” dataset also ran successfully after 250 million chains but also displayed poor mixing for C1+C2 (ESS of ~100). In addition, the tip date CR092 did not converge well (ESS of ~100). The two runs were combined and 95% HDP tip dates were recorded (Table 5.4). A phylogeny was constructed with the most likely tree in the “minimum 3X coprolite depth” dataset (Figure 5.3).

The BEAST tip dating of coprolite samples returned mixed results. The 95% HPD of the tip age for the Bronze Age CR076 and recent coprolite CR092 were consistent with stratigraphic dating while the Neolithic samples estimated were not consistent. The Neolithic CR079 produced a wide 95% HPD age estimate while the tip age of CR073 was estimated to be much more recent than the Neolithic. This suggests that CR073 is incorrectly dated. However, it may have also been due to incorporation of extra mutations by contamination or damage deflating the age estimate.

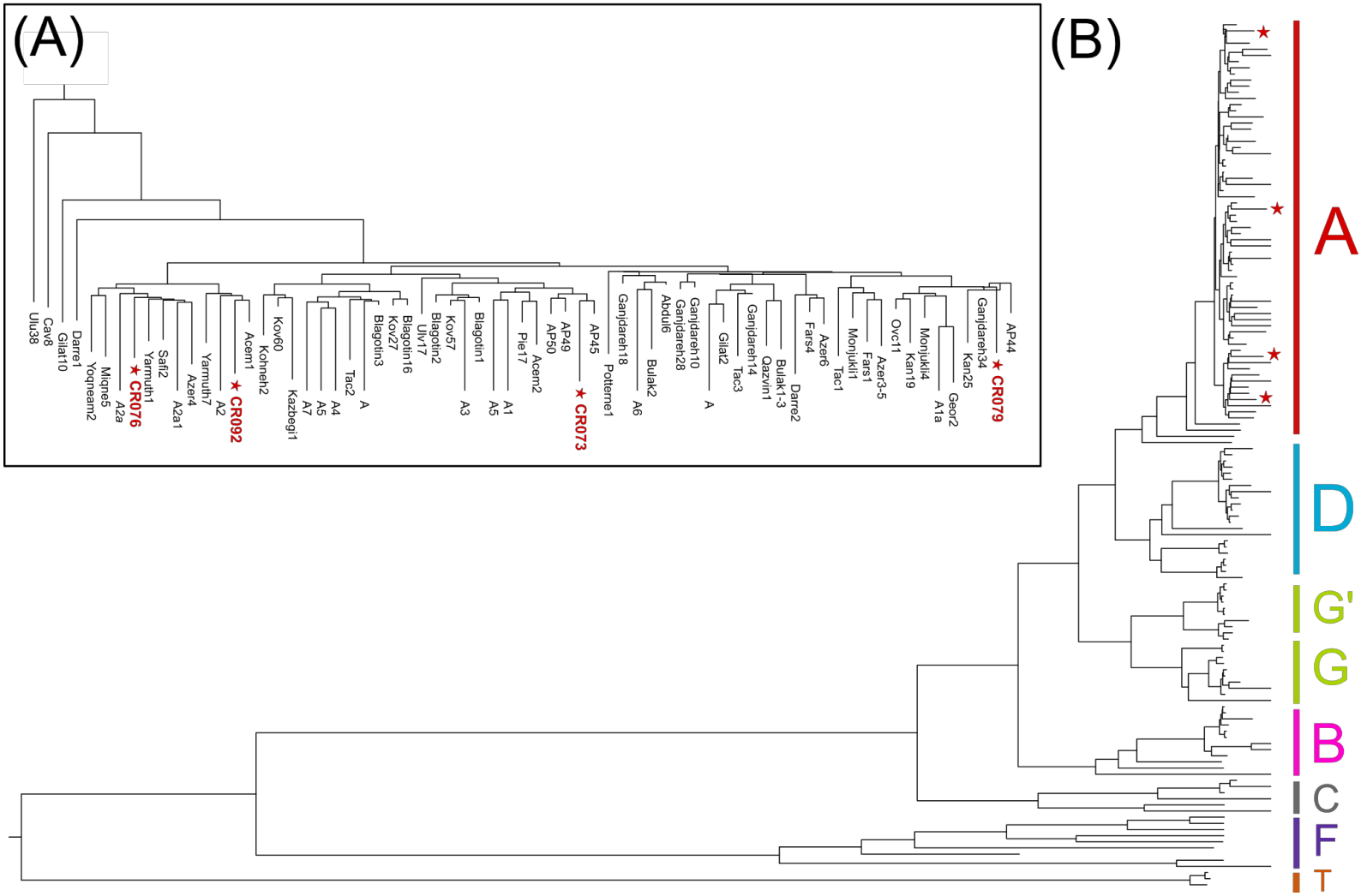


Figure 5.3 - A Bayesian phylogeny placing the coprolites in the context of ancient and modern mitochondrial diversity. The tree topology reflects previous goat phylogenies reported in Daly et al. 2018 and Daly et al. 2021. While all genotyped coprolite returned an A haplotype, the four samples above 5X displayed variation within the macro-A clade, forming separate clades with diverse populations.

Table 5.4 - BEAST tip dating of coprolite samples.

Sample	Age by stratigraphy	95% HPD interval (median) minimum 3X	95% HPD interval (median) minimum 5X
CR073	6000-6200 BCE	0-2967 (886)	0-4240 (1438)
CR076	3000 BCE	2-5833 (3069)	307-7025 (4182)
CR079	6000-6200 BCE	3-7440 (3561)	N/A
CR092	300 yrBP	0-4229 (1644)	0-3769 (1411)

5.4.3. Autosomal analysis

The low coverage of coprolite nuclear genomes generated restricted the number of feasible autosomal analyses (Table 5.3). The NJ IBS tree places ancient samples from Iran and the Levant cluster in groups based on their age and location. Nestled within the modern African cluster are three Levantine samples from Neolithic ‘Ain Ghazel in Jordan and one sample from a Chalcolithic site in Northern Negev (gilat8). Clustering just outside the African/European clusters are the three coprolite samples and one Levantine Bronze Age (safi2). Clustering closer to the modern Southwest Asian goats are two more Bronze Age Levantine goats and a Bronze Age Iranian goat. Neolithic Iranian goats cluster together further along the “Eastern” gradient.

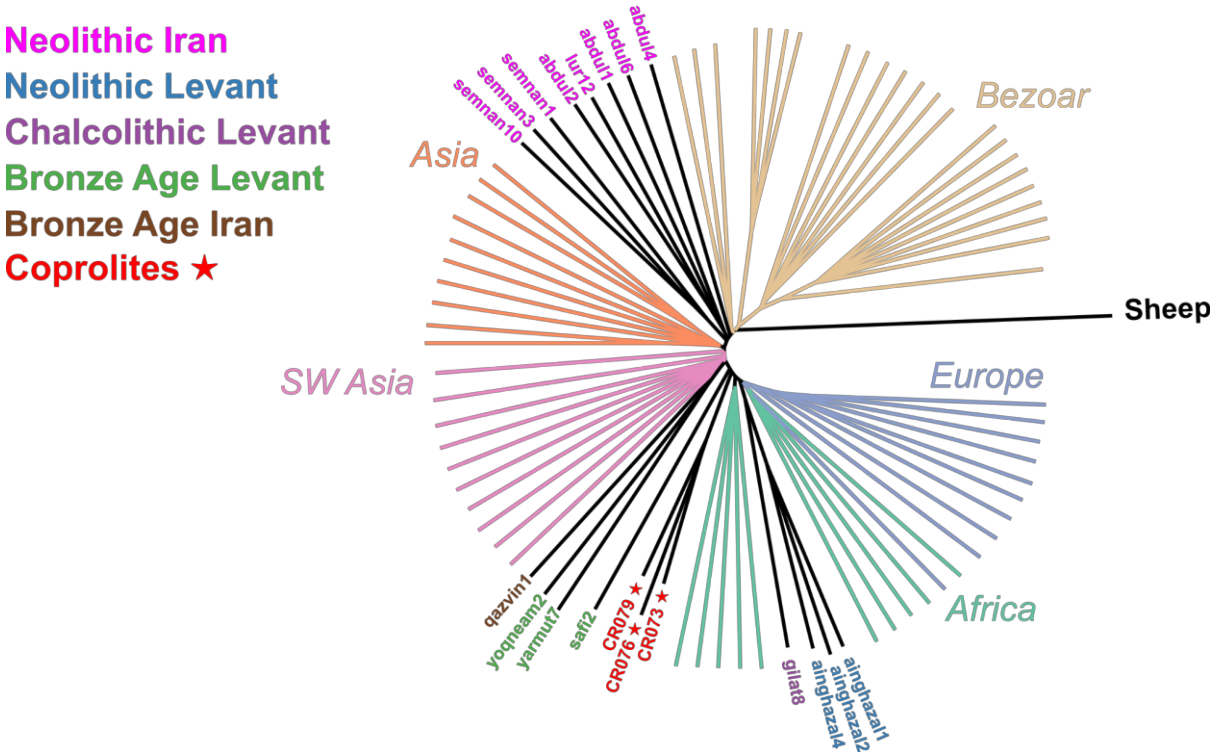


Figure 5.4 - A rooted Neighbour-Joining IBS tree was constructed using ancient and modern goat genomes of at least 0.01X average genome coverage.

A LASER projection PCA depicts a similar trend of genomic affinities (Figure 5.5). Here, modern goat populations cluster by geography, with a clear Western to Eastern gradient emerging in PC1. Ancient Levantine goat genomes are projected onto this space revealing the previously discussed eastern influence on post-Neolithic genomes. The Neolithic coprolite samples occupy a space along this gradient represented by post-Neolithic populations of goats, suggesting that they have been influenced by eastern Fertile Crescent populations.

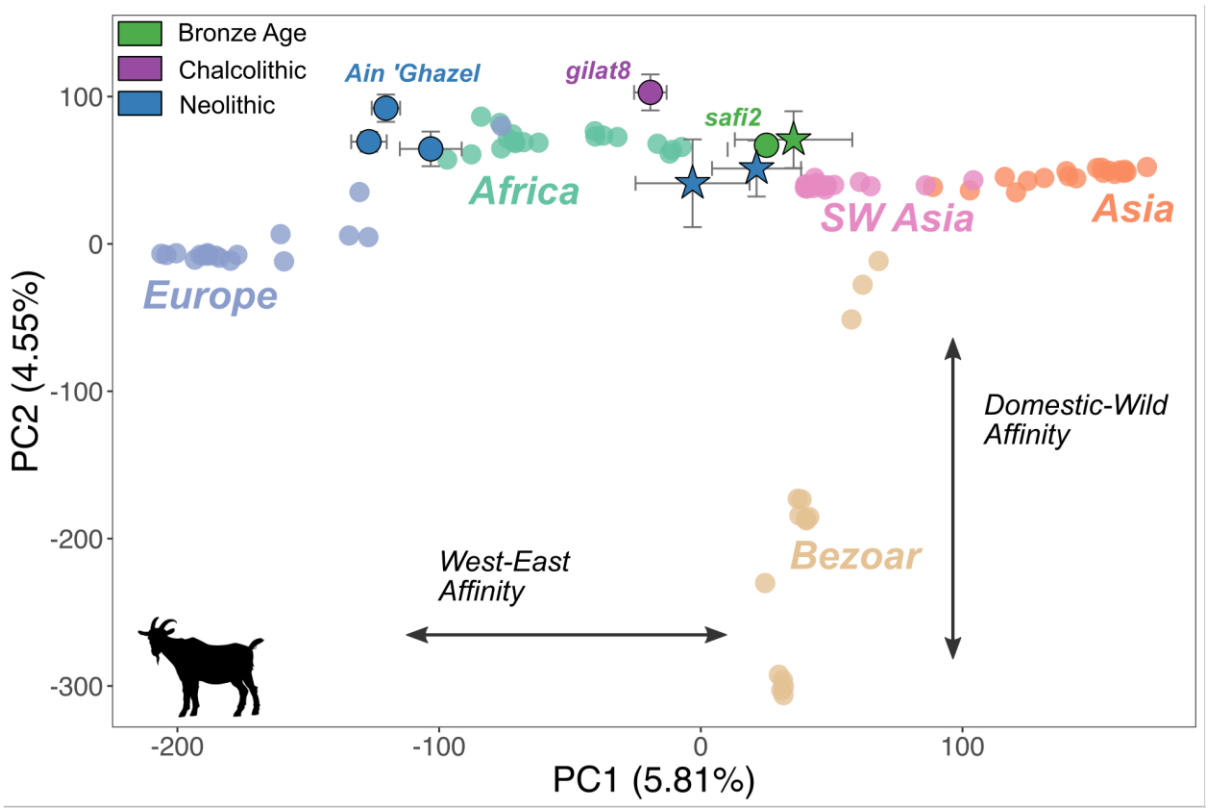


Figure 5.5 - PCA of global goat affinity and Procrustes projection of ancient Levantine samples. Error bars represent the standard deviation of each PC coordinate between 10 independent runs * 2. PC1 separates breeds along a West-East gradient, while PC2 separates domestic and Wild animals. Coprolite samples (CR073, CR076, CR079) are depicted using a star shape.

5.4.4. Contamination

The heterozygosity of the mitochondrial genomes was contextualised using published ancient bams for goats also of haplotype A (Appendix Table 5.4). Four outliers were identified when considering heterozygosity at all sites (Figure 5.6A). Of these three outliers, one was the Neolithic coprolite sample CR073 and two were recent coprolites CR077, and CR078. Five

outliers were identified when considering haplotypic sites, including three coprolite samples (CR073, CR077, and CR078) (Figure 5.6B).

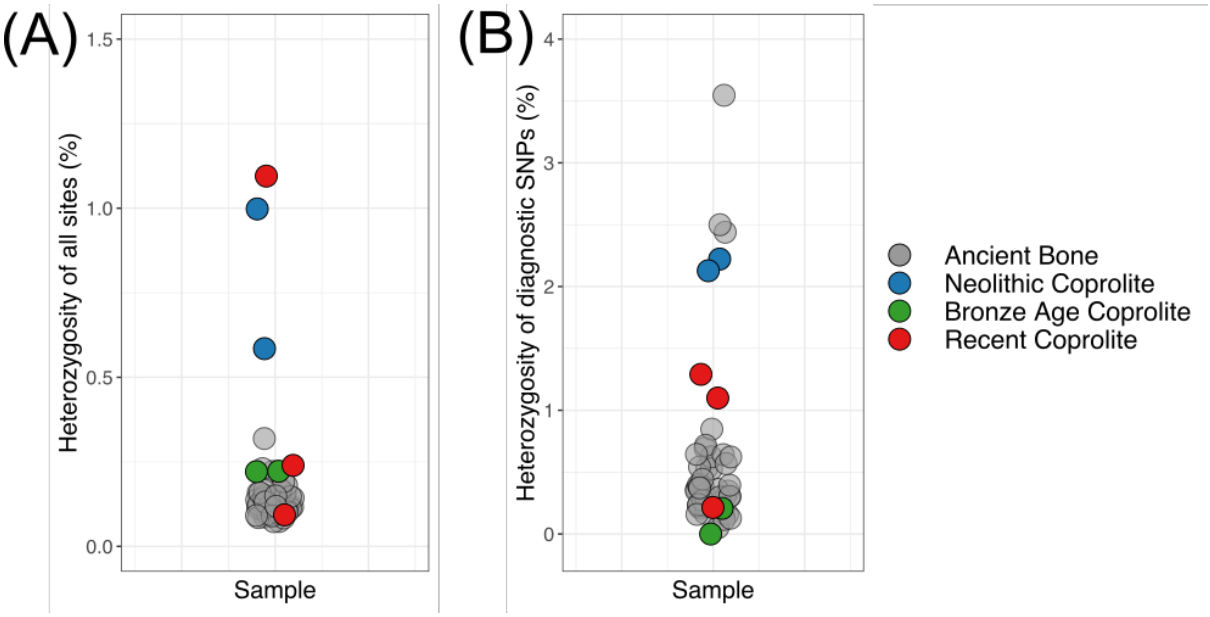


Figure 5.6 - Contamination estimates based on mitochondrial heterozygosity rates at (A) all sites and (B) haplotypic diagnostic sites only.

To provide an upper estimate of contamination, heterozygosity of heteroplasmic sites only was computed (Figure 5.7A). 5 outliers were identified based on this metric, 2 of which were coprolites (CR078 and CR081). However, it was observed that this heteroplasmic estimate may be correlated to the total number of reads, although the correlation was non-significant (R^2 = 0.1991, p-value = 0.1529).

Contamination of the higher coverage samples was also investigated using sexing plots generated for the three higher coverage samples. It was noted that while CR073 was denoted male, the ratio of X chromosomal reads was not proportional to the autosomal reads unlike the female CR076 and CR079 samples (Figure 5.8). This suggests CR073 may be derived mainly from a male goat(s), with a lesser contribution from a female goat(s).

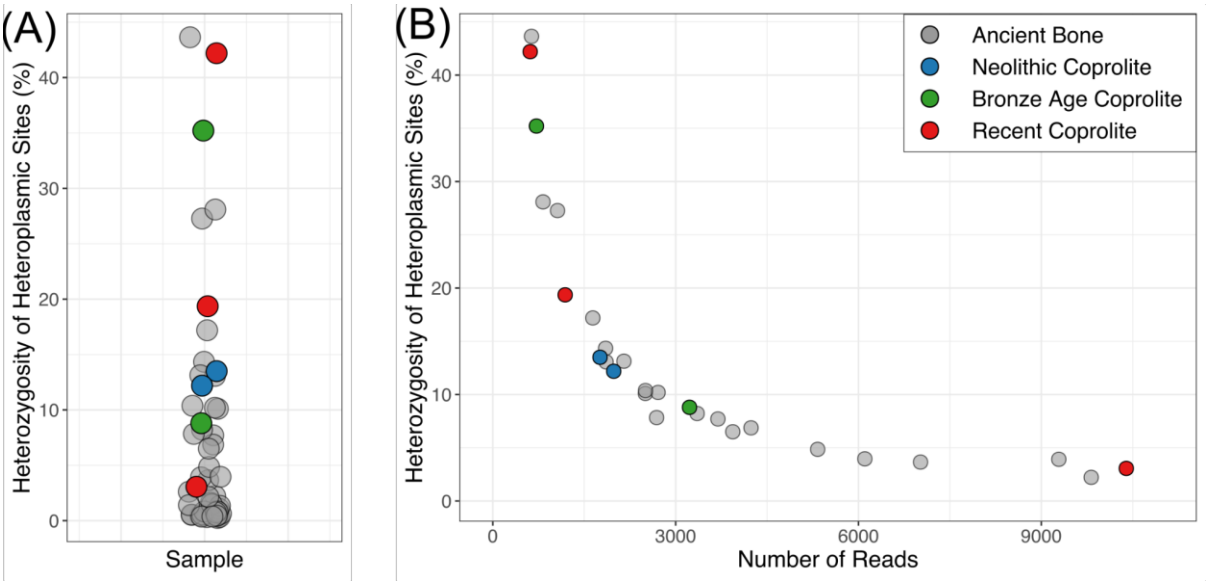


Figure 5.7 - Upper estimates of contamination in ancient goat mitogenomes. (A) Upper limits of contamination based on heterozygosity of heteroplasmic sites only. (B) Heterozygosity of heteroplasmic sites plotted against the total number of reads in each bam.

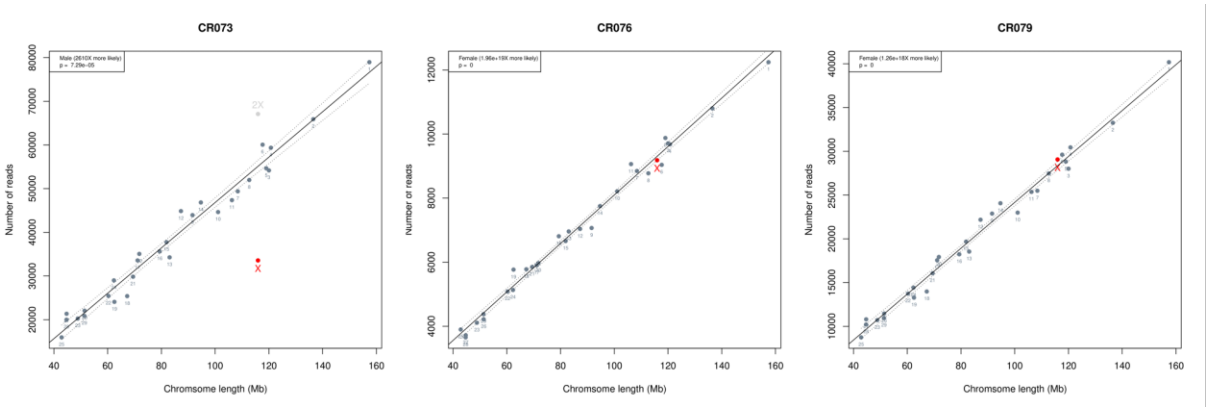


Figure 5.8 - Sexing plots of the three high coverage coprolite samples.

5.4.5. Microbiome

Using the full Kraken2 database, 2.93%, 1.58% and 1.67% of filtered metagenomic reads were taxonomically assigned in CR073, CR076 and CR079. The most commonly represented phyla are displayed in Figure 5.9.

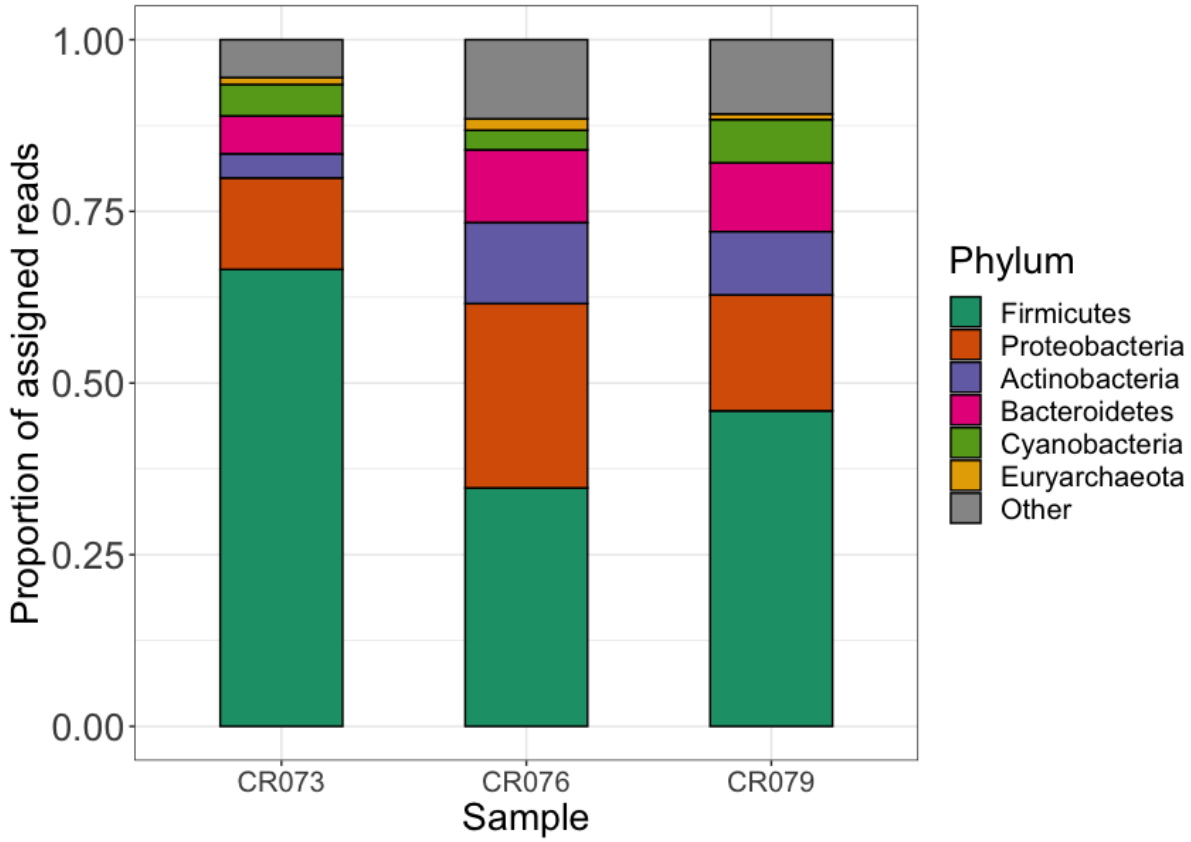


Figure 5.9 - The most commonly represented phyla in coprolite samples. The ratio of Firmicutes:Bacteroidetes was 11.94:1, 3.27:1, 4.58:1 for CR073, CR076 and CR079, respectively.

Using SourceTracker2, the metagenomic sources of the coprolites, laboratory controls and ancient bone/tissue control samples were predicted. At species levels, SourceTracker2 predicted between 12-19% of the coprolite metagenome was derived from a Ruminant gut (Figure 5.10). However, the most commonly predicted source was “unknown”. None of the control samples had a prediction of Ruminant gut above 5%.

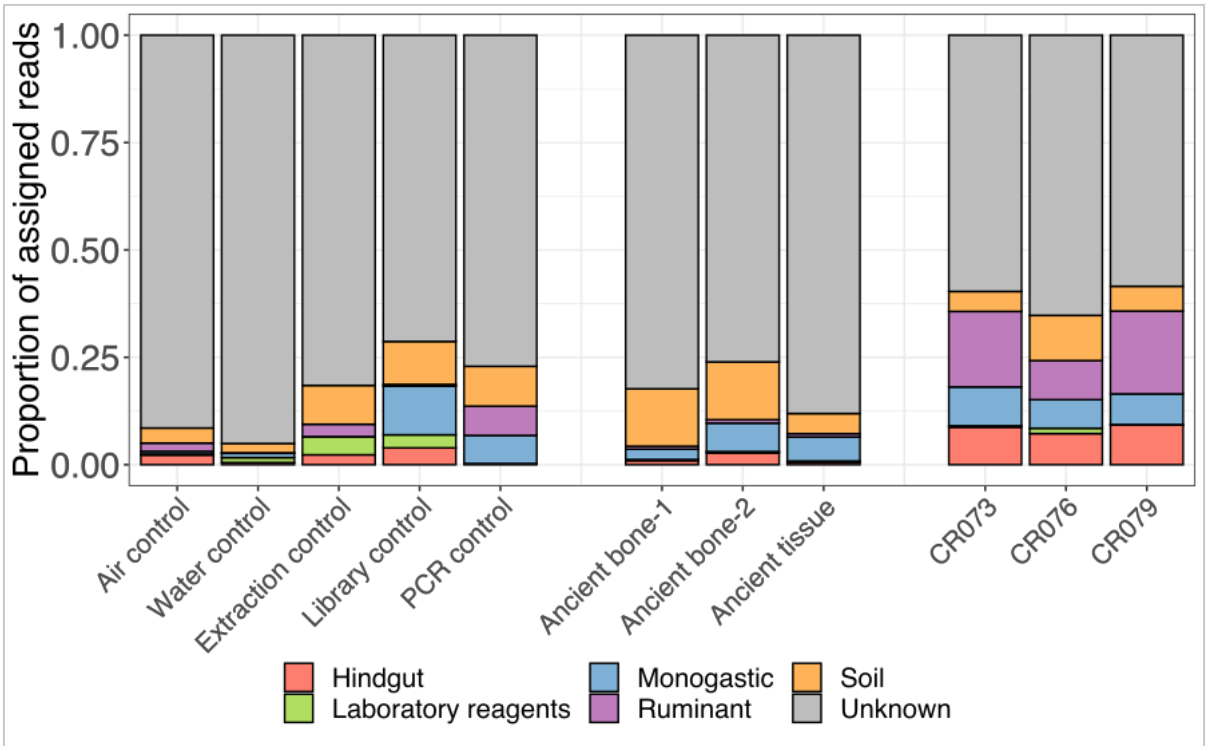


Figure 5.10 - The source prediction of the coprolites and controls using SourceTracker2 at Species level.

When samples were predicted at Genus level, Ruminant source prediction jumped to 0.367, 0.5473, and 0.562 in CR073, CR076 and CR079 respectively. Ruminant source prediction remained low in the control samples (0.0002-0.0018).

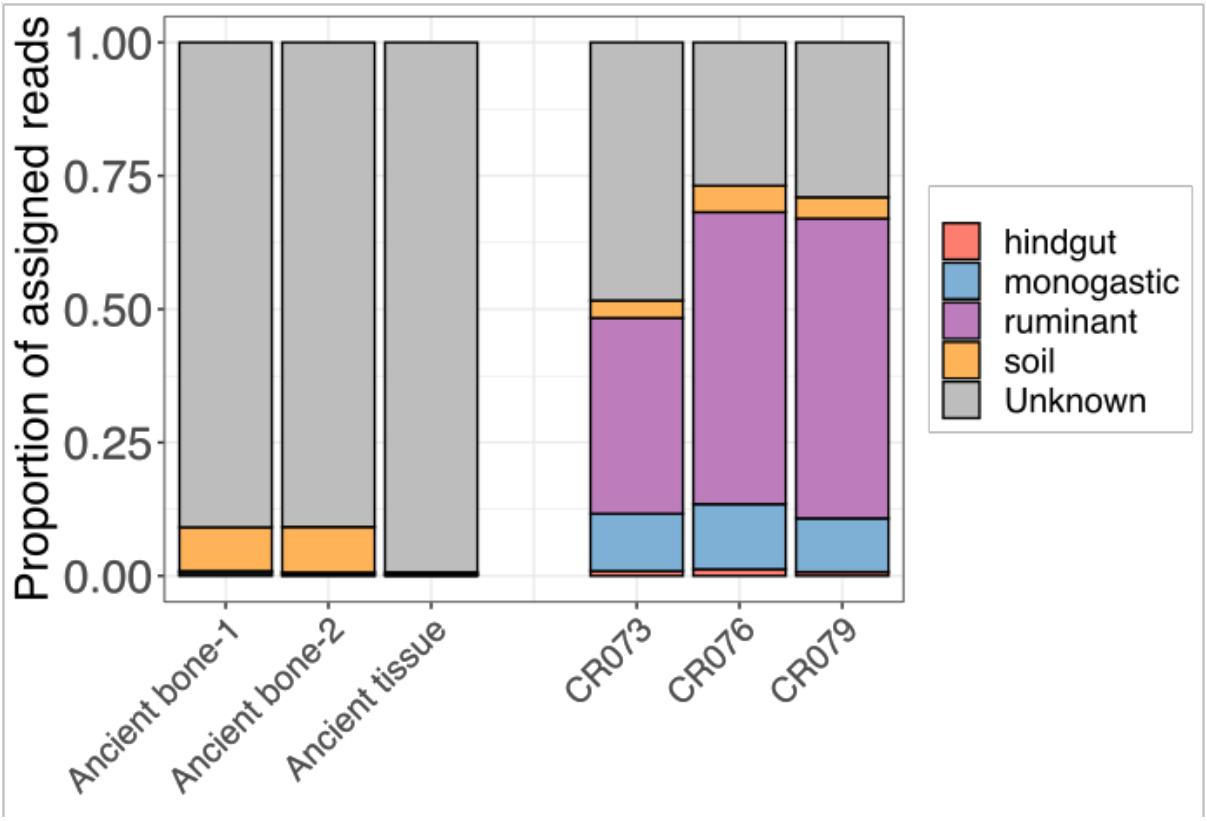


Figure 5.11 - The source prediction of the coprolites and controls using SourceTracker2 at Genus level.

5.4.6. Disease detection

Using Kraken2, four CR076 reads matched uniquely to *Mycobacterium avium paratuberculosis,* the cause of Johne's disease in goats. A further 27 reads matched the *Mycobacterium avium complex* (MAC). Of modern goat dung samples, only one sample had reads matched to *Mycobacterium avium paratuberculosis.*

The metagenomic reads of CR076 were subsequently aligned to the *Mycobacterium avium paratuberculosis* reference genome. 27,068 raw reads aligned successfully resulting in an approximately 0.2X genome. After duplicate removal and mapping quality filters (mq30), 1,136 reads were retained. Edit distances were explored in the resulting genome. The edit distances displayed characteristics of a poorly chosen reference (Figure 5.12A). Additionally, a very uneven genome coverage was observed across the genome (Figure 5.12B).

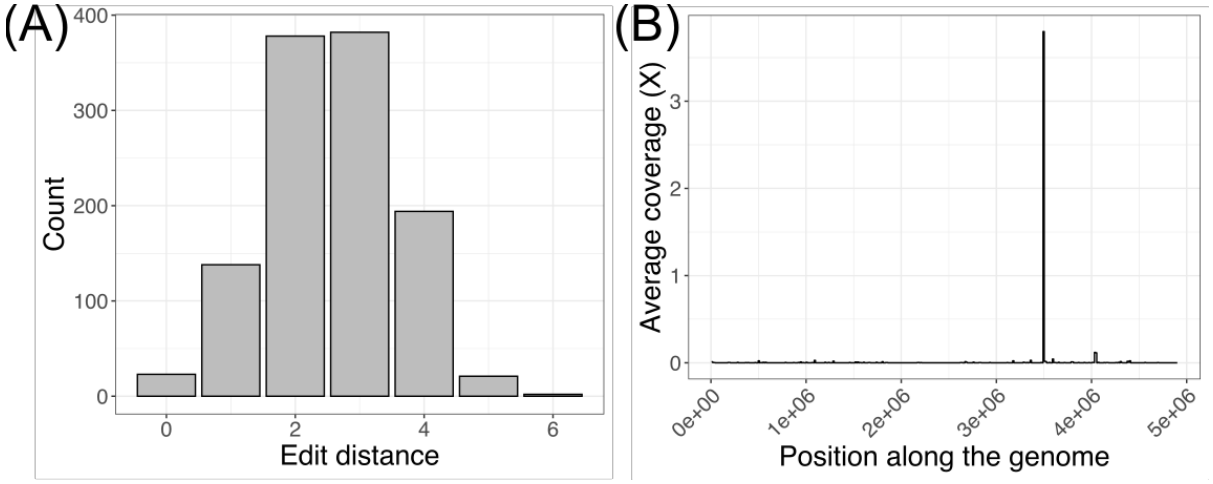


Figure 5.12 - Putative ancient pathogen authentication measures. (A) The edit distance of aligned reads. (B) The average coverage along the genome.

5.4.7. Plant detection

Using the Kraken2 genome skimming database, between 1.06-4.76% of reads were assigned to Viridiplantae. Of the assigned reads, the coprolite samples were the only samples to have above 1% assigned to Kingdom Plantae, except for the water control (Figure 5.13).

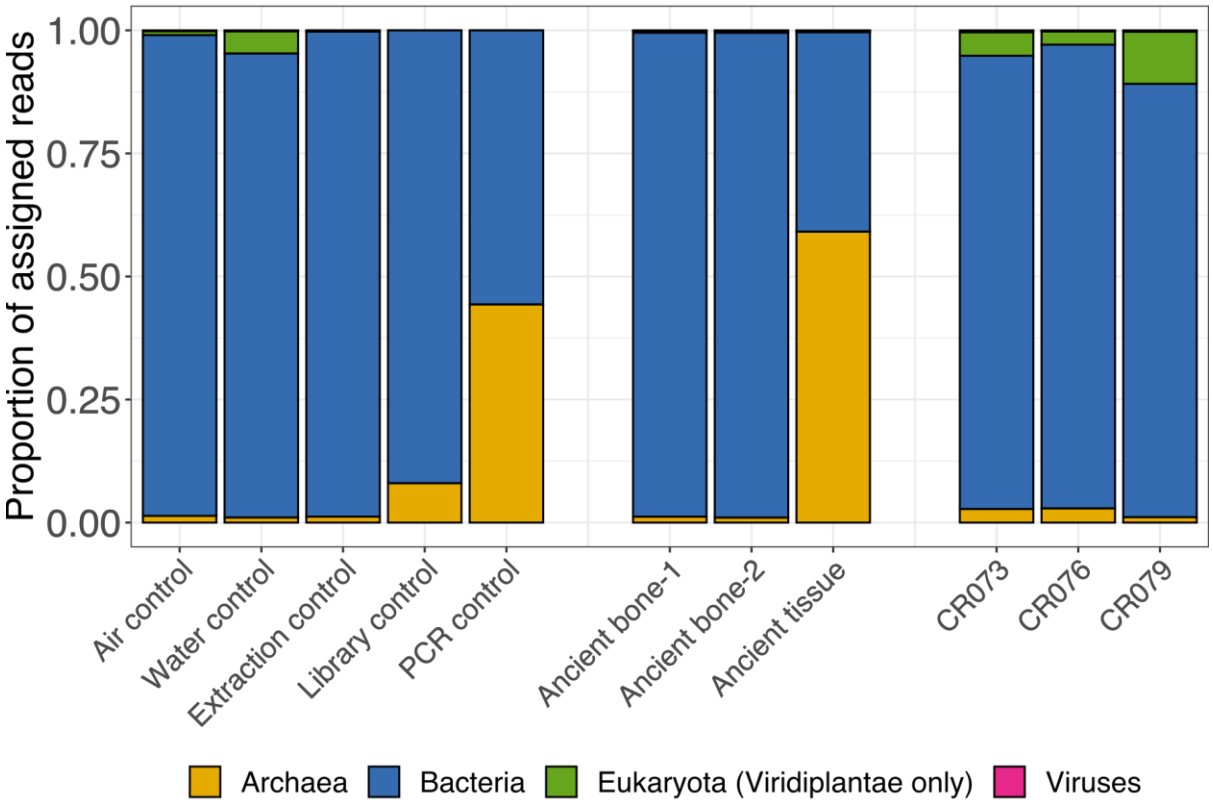


Figure 5.13 - Proportion of assigned reads using a custom genome-skimming database.

The presence of reads matching *Hordeum* was investigated in coprolites and controls using the three custom Kraken2 Plant databases (Table 5.5). *Hordeum* was detected using all databases in CR073 and CR076, but not CR079 using the trnL database. *Hordeum* was also widely detected in control databases. Notably, it was detected in each database for Ancient tissue and Ancient bone-2 sample controls. *Hordeum* was not detected in any laboratory controls except in the Genome-skimming database.

Table 5.5 - Presence or absence of *Hordeum* assigned reads in coprolites and controls using three custom Kraken2 databases.

Sample	Genome-skimming	Chloroplast	trnL
Air control	No	No	No
Ancient bone-1	Yes	No	No
Ancient bone-2	Yes	Yes	Yes
Ancient tissue	Yes	Yes	Yes
CR073	Yes	Yes	Yes
CR076	Yes	Yes	Yes
CR079	Yes	Yes	No
Extraction control	No	No	No
Library control	No	No	No
PCR control	No	No	No
Water control	Yes	No	No

5.5. Discussion

5.5.1. DNA preservation

This study presents the first NGS sequencing data of ungulate coprolite aDNA and the first instance of shotgun sequencing of autosomal coprolite aDNA. The endogenous content of the coprolites was low when compared to endogenous DNA content of sequenced bone and teeth in this thesis (Chapter 3). However, the expected endogenous content of coprolites should be scaled to that of modern faeces. Most host DNA in faeces is derived from gut epithelial cells that form a thin layer on the faeces as it passes through the digestive tract (Bach et al. 2022). This endogenous DNA content is generally low in modern faeces - as little as 1.8% in primates (Perry et al. 2010). Endogenous DNA may be then leached into the environment and fragment

to unamplifiable levels due to age-related decay of aDNA. Therefore, the level of endogenous content was relatively successful.

DNA preservation is generally poor in bones and teeth in the Levant due to the high mean annual temperatures (Kistler et al. 2017). As a result, ancient goat bones sampled from this region have rarely yielded more than 1% endogenous DNA. However, for aDNA preservation in these coprolites, the high temperature may have conferred a paradoxical benefit to preservation. A hot and arid environment rapidly desicates faeces, encrusting and perhaps protecting the endogenous colonocytes from environmental leaching and invasion of saprophytic microorganisms. Further protection to the endogenous DNA was afforded by the rockshelter where these coprolites were less exposed to precipitation (Landau et al. 2020), as it has previously been reported that damper coprolites yielded poorer aDNA (Wood et al. 2016). In sum, coprolites may be a desirable substrate to target for future aDNA studies in arid environments, contrary to ancient mineralised tissue.

While the endogenous content of the samples was encouraging, the coprolites displayed substantial and unusual DNA damage. There is evidence of typical age-related DNA damage, with elevated deamination patterns at CpG sites in USER treated samples. Furthermore, the damage patterns were highest in the older Neolithic samples. However, all coprolites displayed substantial indiscriminate mismatching prior to read ends except for CR092 (a recent coprolite). While this could be due to laboratory methods that were specific to coprolites, this cannot explain why CR092 was seemingly unaffected. Research of aDNA decay dynamics has been mostly focused on mineralised tissue deposited in soil or permafrost (Kistler et al. 2017). As discussed in Chapter 4, it may be incorrect to expect aDNA to decay in a similar fashion for different tissue types (*i.e.* epithelial cells) and taphonomic contexts (*i.e.* in an encrusted dung pellet). While outside the scope of this project, the unique taphonomic context of these coprolites may have exposed the endogenous DNA to oxidative agents leading to this unusual damage to the DNA structure (Lindahl 1993). Such damage has not been reported in other coprolite studies (e.g. (Rampelli et al. 2021)), however very few aDNA coprolite studies have been reported and none with autosomal DNA. In addition to the indiscriminate base mismatching, high fragmentation of coprolites was observed. This likely contributed to the low concentration of aDNA even after high PCR cycling.

5.5.2. Population genomic analysis

The domestic goat haplotype A was genotyped in all goats in this study. The earliest detection of the A haplotype is in Neolithic Anatolian and Iranian herds of domestic goats (Daly et al.

2021). Conversely, the three published ancient mitogenomes in the Neolithic Levant ('Ain Ghazel, Jordan) are exclusively F haplogroup (Daly et al. 2018). In the Levant, a major turnover in haplotypes was observed at the Neolithic to Chalcolithic transition where D and A mitochondrial haplogroups emerge. As two of the haplotype A genotypes here were stratigraphically dated to the Neolithic, this would indicate this turnover occurred earlier than previously thought or that A had co-existed with F during the Neolithic Levant. From the Bronze Age, the A haplotype became dominant in the region, and has persisted to the present (Daly et al. 2018). In this dataset, haplotype A is also found through the Bronze Age until the modern era. The F haplogroup, which existed in Neolithic Levant, is now scarcely found in modern domestic goat herds; only encountered in Sicilian goat populations (Sardina et al. 2006).

Mitochondrial in post-Neolithic Levantine goats turnover was accompanied by a nuclear transition (Daly et al. 2018). In the post-Neolithic, an eastern influence is observed in the nuclear genomes of domestic Levantine goats from as early as the Chalcolithic Era (4,500 BCE) in Northern Negev. In this study, PCA and NJ IBS trees reveal that the Neolithic higher-coverage coprolite samples also display an eastern affinity, clustering with Bronze Age Levantine genomes from Northern Israel. While the Neolithic Levantine genomes are very low coverage, their clearly different nuclear affinities corroborates the mitochondrial turnover observed. Specifically for Ramon Crater, it also suggests a genetic continuity of goat herds from the Late Neolithic through the Bronze Age and into the recent age. Taken together, this would suggest that the eastern influence on Levantine goat genomes happened in the Late Neolithic, where two of these samples are dated to (CR073 and CR079). Explicitly, this intrusion of eastern goat ancestry into the Levant predates the previous earliest detection by over 1,000 years (Daly et al. 2018).

Interpretations of this arrival of eastern ancestry to post-Neolithic Levantine goat herds has been attributed to human-mediated movement of animals (Daly et al. 2018). The findings of this study indicating an earlier arrival of eastern ancestry in late Neolithic goats does not directly conflict with current evidence of human migration in the Near East using aDNA analysis. However, there is no direct evidence to support an earlier migration either. Low levels of human migration from the eastern or northern Near Eastern Neolithic populations has been detected in the Neolithic northern Levant populations (Skourtanioti et al. 2020). By the arrival of the Bronze Age, more than half of the Northern Levantine gene pool had been replaced by these eastern or northern ancestries, indicating widespread migration. In Bronze Age Southern Levantine populations, both local and eastern Fertile Crescent ancestries are detected in ancient human genomes, indicating human mobility had influenced the southern Levant by this

time (Agranat-Tamir et al. 2020). No goat genomic data from the Neolithic Southern Levant has been published. The late Neolithic introduction of eastern goat ancestry into the Southern Levant (Negev desert) would be the earliest evidence of eastern migration into the region.

While these results are intriguing, they should be interpreted with caution. There are a number of aspects to the analysis that are prone to error or contamination. Firstly, these coprolites were not directly dated but stratigraphically dated by the rock layer they were recovered from at Ramon Crater. Each layer was assigned a date based on C14 dating on plant and charcoal material in the horizon. Using molecular dating of the mitochondrion, the stratigraphic dating broadly tracks for all samples except CR073, which predicts a much more recent date. However, there is evidence that CR073 may consist of DNA from multiple individuals based on molecular sexing and mitochondrial contamination estimates. If accrued mitochondrial mutations from multiple individuals were incorporated into the consensus fasta, the inflated branch length could result in a more recent tip date estimation. To resolve this ambiguity it may be possible to C14 date the coprolites with the remaining material (Welker et al. 2014).

Secondly, coprolite aDNA is vulnerable to contamination by leaching of genetic material from upper layers. It is possible that urination of more recent goat herds contaminated lower levels of sediment, resulting in false signals of more recent ancestry in our Neolithic goat herd. While leaching is of less concern in arid environments where there is less precipitation to flush DNA molecules to lower strata (Haile et al. 2007; Hebsgaard et al. 2009), heterozygosity estimates in coprolites were generally higher than that of ancient bone samples (0.22-2.21%). However, contamination calculated by mitochondrial heterozygosity reaches as high as 5% in other ancient samples (Teasdale et al. 2015). This suggests low levels of contamination, or contamination from goats of the same haplotype. In light of the uncertainty associated with coprolite aDNA analysis, these results should be reviewed with scepticism.

5.5.3. Microbiome analysis

A ruminant microbiome was detected in all of the higher coverage coprolite samples. While other studies have already detected ancient gut microbiomes of humans and dogs, this is the first attempted characterisation of a ruminant microbiome (Tito et al. 2008; Borry et al. 2020). Microbiome analysis can reveal important aspects of gastrointestinal health of herd animals. For example, previous studies have demonstrated that healthy kids and goats not only displayed higher species richness, but also exhibited higher bacterial diversity than diarrheic kids (Wang et al. 2018). While it would be informative to compare the microbiome composition of ancient herds to modern herds, drawing conclusions of species richness and diversity in

coprolites may be problematic as the metagenomic composition of coprolites likely consists of a mixture of microbiomes that would inflate species richness and diversity (Raul Y. Tito et al. 2012).

Ancient herd health is perhaps better ascertained by scanning for disease-causing pathogens. Approximately 60% of human diseases are zoonotic, and many likely arose as a direct result of animal domestication (Woolhouse and Gowtage-Sequeria 2005; Diamond 2002). The Neolithic brought increased contact with live animals and a denser population for humans which accelerated zoonotic disease transmission. Coprolites offer a rare opportunity to capture genetic information of ancient domestic animal gastrointestinal pathogens. However, a scan of common disease-causing microbes affecting goat gastrointestinal health did not yield any convincing results. While there was a potential hit of the pathogenic *Mycobacterium avium paratuberculosis,* this hit ultimately failed basic ancient pathogen authenticity procedures (Hübler et al. 2019). The matching *Mycobacteria* hits may have been due to invasion of soil microbes where *Mycobacteria* are particularly prevalent (Walsh et al. 2019).

The detection of ancient pathogens has a solid framework to scrutinise and validate potential hits. In contrast, investigating evidence of diet in the coprolites was challenging as the detection of ancient metagenomic plant DNA is still relatively *ad hoc* (Harbert 2018). Phytolith analysis suggested these herds fed on barley (*Hordeum*) grasses (Landau et al. 2020), and therefore detection of barley sequence in the coprolites was attempted. Exploration of diet detection in coprolites was conducted in this study with databases that were filtered to account for overrepresentation of barley. In the coprolite samples, *Hordeum* was detected across several curated datasets, further supporting evidence that these goats fed on barley grasses in the Negev. However, barley DNA reads were also detected in several control samples which undermines this result.

A central issue regarding metagenomic Plant detection is the curation of a reliable database. Some studies have attempted to use chloroplast or universal gene markers to detect ancient plant matter in a barcoding method (Wood et al. 2016; Boast et al. 2018). However, such barcoding of plant chloroplast genomes may be problematic due to the lower variation in plasmid DNA and also the widespread hybridisation in plants (Coissac et al. 2016). Other studies have opted to use all available plant sequences but this can also result in controversial and perhaps spurious results. For example, purported detection of wheat aDNA (*Triticum*) in Mesolithic Britain 2,000 years earlier than previously thought was based upon 152 sedaDNA reads assigned to *Triticum* using BLAST (Smith et al. 2015). A later rebuttal challenged this

conclusion as modern contamination due a lack of *post mortem* aDNA damage patterns (Weiß et al. 2015). This controversial result may have also been due to a database bias. Approximately 2.3% of the ~27 million Plant sequences downloaded from GenBank in the database used in this Chapter were *Triticum* prior to removal of contigs and gss. This further emphasises the potentially problematic nature of detecting a model Plant organism using aDNA. Due to the risk of false-positive hits, it has been recommended that any aDNA detection is complemented with plant macrofossils found in coprolites (Fuks and Dunseth 2020).

6. Next generation sequencing of modern Chromium (III) tanned leather using ancient DNA techniques

6.1. Introduction

DNA analysis of leather has wide-reaching applications; from ancient DNA (aDNA) analysis of livestock animals, to detection of counterfeit modern leather products (Matar and Merheb 2016). Despite a wide interest in leather DNA, very few studies have reported successful DNA amplification from modern or ancient sources. Furthermore, those who have successfully amplified DNA from leather (*e.g.* (Vuissoz et al. 2007), (Merheb et al. 2015), (Schröder et al. 2016), (O'Sullivan et al. 2016)) have only reported mitochondrial sequences, with some suggesting that nuclear DNA (nucDNA) may not be recoverable from leather. The difficulty in amplifying DNA from leather has been attributed to the tanning process, which is an ancient human practice used to preserve animal skins and hides as a durable and versatile material (Urlings 1995). The tanning process features several steps that may hinder DNA accessibility and survivability.

While industrial leather tanning differs at several points in the process from ancient and traditional practices, the key steps thought to affect nucleotide integrity remain the same (Vuissoz et al. 2007). Post-slaughter, moist animal hides are preserved in an acidic brine or stored on rock salt to slow decomposition. The salting process can be done over many weeks and prevents putrefaction of the hide. The epidermal and fatty adipose tissues are then removed from the salted hides through alkaline washing in a process known as liming. Modern liming causes a change in the pH of the animal hide to pH 13-14, using caustic chemicals such as calcium hydroxide, sodium carbonate, sodium hydroxide and sodium sulphide to remove keratinous tissue (Urlings 1995). During these steps, hydrolytic and oxidative reactions are facilitated by the acidic salting and subsequent alkaline conditions, during which the endogenous DNA may be damaged (Matar and Merheb 2016). In an experiment that attempted to amplify leather DNA at each stage in the tanning process, (Vuissoz et al. 2007) reported a loss of amplifiable nucDNA after the alkaline liming stage.

Following this stage, the hides are pickled to achieve pH levels as low as 1.5 to prepare for the incorporation of the tanning agents (tannins) into the skin. This is a critical stage as tannins crosslink with the collagen proteins to produce a resistant, hydrolytically stable leather (Urlings

1995). Traditionally, polyphenolic tannins derived from vegetal materials were used to tan hides. In most modern industrial settings, mineral tanning agents like trivalent chromium (Cr(III)) compounds are used. The tanning stage may negatively affect DNA survivability in a number of ways. Firstly, tanning agents have been suggested to crosslink with DNA accessory proteins in a similar manner to how they crosslink collagen, thus negatively affecting the amplification potential of the DNA (Vuissoz et al. 2007). Unlike nucDNA, mitochondrial DNA (mtDNA) is not bound to histone proteins, and so this may help to explain why only mtDNA has been successfully amplified in leather samples. Crosslinked collagen triple helices may also complex with DNA double helices, trapping DNA fragments and rendering them unamplifiable (Vuissoz et al. 2007). The use of Cr(III) compounds as tanning agents in modern contexts may also pose unique issues in DNA damage. Cr(III) ions and compounds have been shown to cause genotoxicity by direct binding with DNA and fragment cleavage (Fang et al. 2014). These Cr(III)-DNA adducts and short fragment sizes may limit the PCR potential of leather DNA.

In this chapter, ancient DNA laboratory and bioinformatic methods are used to sequence DNA from modern bovine leather and characterise the preservation patterns. NGS data has been generated for ancient bovine parchment DNA which revealed enrichment of repetitive sequences (Teasdale et al. 2017). The authors here suggest that tight DNA binding of heterochromatin has led to preferential survival of these repetitive sequences. Repetitive sequences, such as transposable elements and satellite DNA, are usually transcriptionally deactivated through tight binding to histone proteins in constitutive heterochromatin (Nishibuchi and Déjardin 2017). Conversely, transcriptionally active genic regions are generally loosely bound by histone proteins in euchromatin. Heterochromatin is also used in female X-inactivation, whereby it transcriptionally silences gene expression for gene dosage compensation. Early in embryogenesis in female mammals, one X chromosome (Xi) is randomly and permanently inactivated by becoming tightly packed into facultative heterochromatin (Wutz 2011). Highly degraded modern leather may serve as an analogous substrate to investigate impact of DNA architecture on DNA preservation, which is poorly characterised for NGS sequencing.

6.2. Aims

1. To develop laboratory methods to extract and amplify DNA from modern Chromium (III) tanned leather.
2. To investigate the preservation patterns of the leather DNA with a comparative dataset of modern and ancient tissues.

6.3. Materials and methods

6.3.1. Materials

Ten samples of modern bovine leather sourced from an industrial tannery were used for this study. Each sample was derived from a unique animal. All samples were tanned using Cr(III) compounds resulting in the signature blue colour (known as chrome-in-blue or wet-blue).

6.3.2. DNA extraction, library construction and sequencing

Approximately 0.5g of finely-grated leather was collected by shredding the leather in a blender and placed into a 2ml Eppendorf tube.

The extraction buffer (1X) was made using 387µl H_2O, 12.5µl 4M NaCl, 24.75µl 1M Tris, 1µl 0.5 EDTA, 24.75µl 10% SDS, 2.7µl 0.5M CaCl2 and 50µl fresh Proteinase K. 500µl of the extraction buffer was added to each 2ml tube containing the grated leather. The tubes were covered in parafilm and vortexed for 60 seconds. The samples were then incubated at 56°C for 24 hours at 350 rpm in an Eppendorf ThermoMixer. Each tube was spun down for 60 seconds at 17,000xg in a centrifuge and the supernatant was removed to a clean Eppendorf tube (EXT1, Extraction 1). A fresh 500µl of extraction buffer was prepared and the extraction was repeated twice more (EXT2, supernatant removed) on the pellet. The final EXT3 of each leather sample (supernatant and leather pellet) was then used for Phenol/Chloroform DNA purification (Section 2.2.3). The DNA was eluted with 40µl of EBT in the tube.

Library preparation was performed in a dedicated aDNA laboratory in the Smurfit Institute of Genetics, following standard aDNA protocols (Section 2.3). Two rounds of sequencing were performed on Illumina MiSeq (single-end 1x63bp) and Illumina NovaSeq platforms (paired-end 2x50bp) at TrinSeq, the Trinity Genome Sequencing Laboratory, Trinity College Dublin.

6.3.3. Read-processing

For paired-end libraries, adapters were trimmed using AdapterRemoval v2.1.1 (Schubert, Lindgreen, and Orlando 2016). All trimmed reads of samples were aligned to a modified bosTau8 reference assembly that included a Y chromosome using BWA (H. Li and Durbin 2009) 0.7.5a with relaxed parameters (-l 1024 -n 0.01 -o 2). Trimmed reads were also aligned to a mitochondrial reference assembly (V00654), as above.

6.3.4. Damage patterns

Age-related damage patterns were computed using mapDamage2.0 and PMDtools (Section 2.2.4.). Read length distributions and median read lengths were calculated using an in-house script. Depurination levels were calculated as described in Section 2.2.4.

6.3.5. Coverage statistics

To investigate enrichment of repetitive elements, RepeatMasker (open-4-0-5) was used. RepeatMasker screens DNA sequences for interspersed repeats and low complexity DNA sequences. RepeatMasker was run with the -s (sensitive) setting on the bosTau8 assembly. The RepeatMasker .out file was used as a bedfile containing 5,547,099 unique repeat regions. BEDTools (Version 2.19.1) (Quinlan and Hall 2010) coverageBed (using -mean and -sorted flags) was used to calculate the mean coverage of BAM files at each of the repeat regions. Using an in-house script, the mean coverage of all repeat regions was calculated. The relative coverage was then determined by dividing the mean coverage of repeat regions by the mean coverage of the whole genome, as calculated by GATK DepthOfCoverage.

To investigate enrichment of euchromatin, a GTF file containing gene predictions by refGene was downloaded from UCSC. This GTF file was converted to bed format using an in house script. Gene prediction was restricted to exons only to produce a bed file containing a total of 136,005 exon regions. Mean coverage was calculated, as above.

Coverage statistics for genic regions, repetitive regions and the whole genome were generated for three deep sequenced leather sample genomes (without a mapping quality filter). Ancient and modern genomes were aligned and processed as a comparison (Appendix Table 6.1). In total, four comparative datasets were generated: modern blood, modern skin, ancient bone (untreated) and ancient parchment (treated).

6.4. Results

6.4.1. DNA Preservation

DNA amplification and sequencing was successful for all modern leather samples (Table 6.1). The base quality of sequenced reads was high. Taxonomic identification using FastQ Screen was also successful for all samples, assigning each sample as *Bos*.

Raw endogenous DNA content indicated high survivability of bovine DNA and minimal exogenous contamination. However, a substantial loss of reads was observed after applying mapping quality filters. An average of 63.76% of reads had a mapping quality score of 0, signifying that these reads align equally well to multiple positions in a genome assembly. The average percentage of reads with a MAQ score of 0 in leather samples was more than double that of ancient bovine samples (Appendix Table 6.2), which suggests that there is an overrepresentation of repetitive sequences in leather DNA. Three samples (Leather-2, Leather-5, Leather-10) were selected to be deep sequenced for further analysis (Table 6.2).

Table 6.1 - Amplification and sequencing results of wet-in-blue leather samples. All samples were sequenced successfully. Furthermore, all samples except Leather-7 had a high raw level of endogenous content of cattle DNA. However, there is an unexpectedly large loss of reads post mapping quality filter (average loss of reads: 77.05%). n=10 unique samples.

Sample	Total reads	Total aligned reads post PCR duplicate removal	Clonality (%)	Raw Endogenous % after PCR duplicate removal	Aligned Reads post mapping quality filter applied	Final Endogenous %	Loss of reads due to mapping quality filter (%)	% of total reads that have a mq=0
Leather-1	48501	39307	0.08	80.98	8034	16.56	79.54	65.32
Leather-2	78796	62383	0.09	79.1	14685	18.64	76.44	62.84
Leather-3	41574	27703	1.39	65.71	7361	17.71	73.06	57.96
Leather-4	37287	25774	0.21	68.98	7126	19.11	72.29	57.7
Leather-5	47879	39842	0.16	83.08	6931	14.48	82.58	69.95
Leather-6	81942	65583	0.17	79.9	14231	17.37	78.26	65.63
Leather-7	8270	1882	0.16	22.72	220	2.66	88.29	79.67
Leather-8	38161	26108	0.12	68.33	7228	18.94	72.28	58.32
Leather-9	42022	32786	0.07	77.96	6791	16.16	79.27	65.77
Leather-10	75931	53033	0.11	69.77	16717	22.02	68.44	54.4

Table 6.2 - Alignment statistics of the deep-sequenced leather samples. A similar loss of reads is observed post mapping quality filters as seen in Table 6.1. Clonality became an issue at deeper sequencing, likely due to low levels of ng of DNA.

Sample	Total trimmed reads	Total aligned reads	Raw Endogenous %	Total aligned reads post PCR duplicate removal	Clonality (%)	Coverage pre-mapping quality filter (X)	Aligned Reads post mapping quality filter applied	Final Endogenous %	Loss of reads due to mapping quality filter (%)	Coverage post-mapping quality filter (X)
Leather-2	559560220	471205162	84.21	330017350	29.96	5.42	72612375	12.98	78	1.27

Leather-6	468340213	392585918	83.82	263220533	32.95	4.65	53746126	11.48	79.58	1.01
Leather-10	527698589	419405898	79.48	307214827	26.75	4.69	80520132	15.26	73.79	1.31

6.4.2. Genomic Preservation

The coverage of repetitive and genic regions relative to the mean whole genome coverage was calculated for each material group (Figure 6.1; Appendix Table 6.1). The coverage of repetitive regions is highly enriched in leather and ancient parchment. Correspondingly in these materials, the relative coverage of genic regions is low. The modern and ancient untreated tissues display a relative coverage much closer to 1X for both genic and repetitive regions although repetitive regions are overrepresented relatively to the average genome. Modern untreated skins display the most overrepresentation of repetitive regions of the three samples.

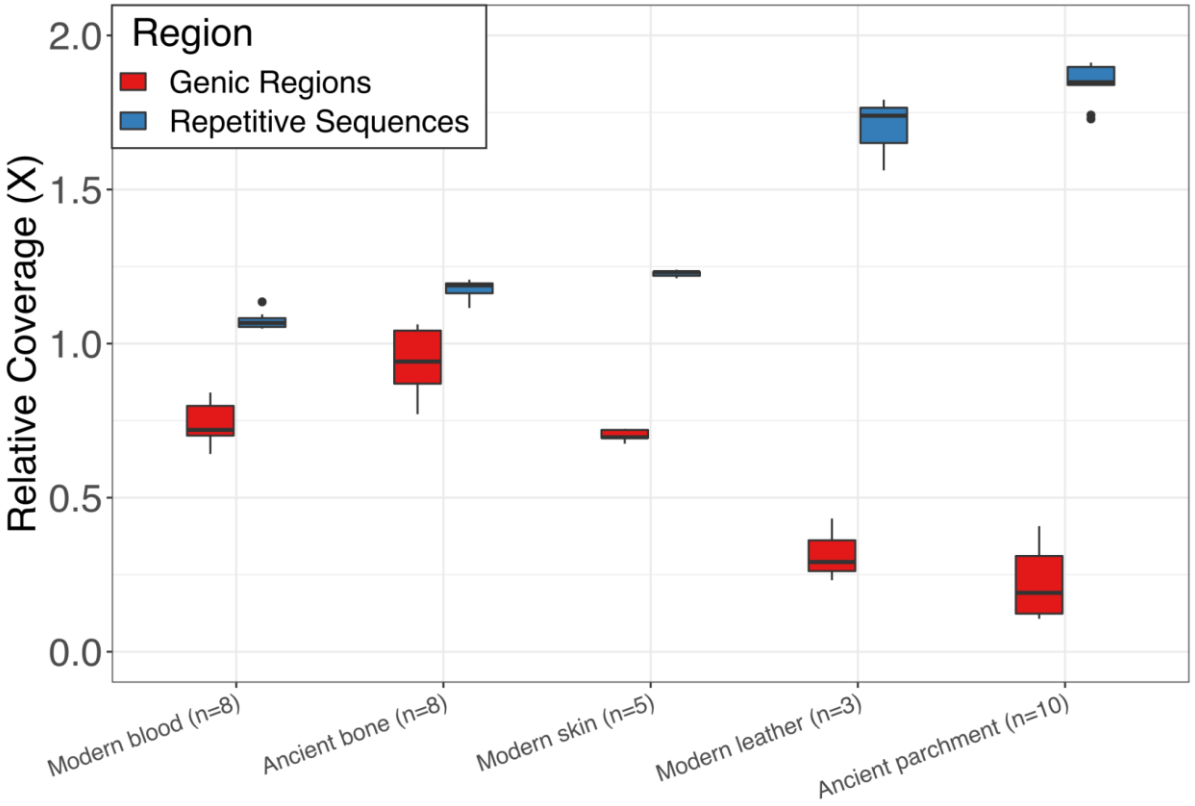


Figure 6.1 - Coverage of genic and repetitive regions relative to the average whole genome coverage in different bovine sequencing data. Untreated ancient and modern bovine tissues display a typical relative coverage of approximately 1X, although repetitive sequences are more represented. The modern leather samples and ancient parchment samples display a dramatic divide in relative coverage between genic regions and repetitive sequences.

The coverage of repetitive regions was directly compared to the coverage of genic regions (Figure 6.2). While repetitive sequences were overrepresented in every material type, the relative coverage did not surpass 2X in the untreated tissues. In contrast, modern leather and

ancient parchment saw an average relative repetitive region coverage of 5.77X and 10.53X, respectively. There is a clear disparity between the survival of repetitive sequences and genic sequences in leather and parchment DNA, suggesting that repetitive sequences are enriched through an analogous process in leather tanning and parchment production.

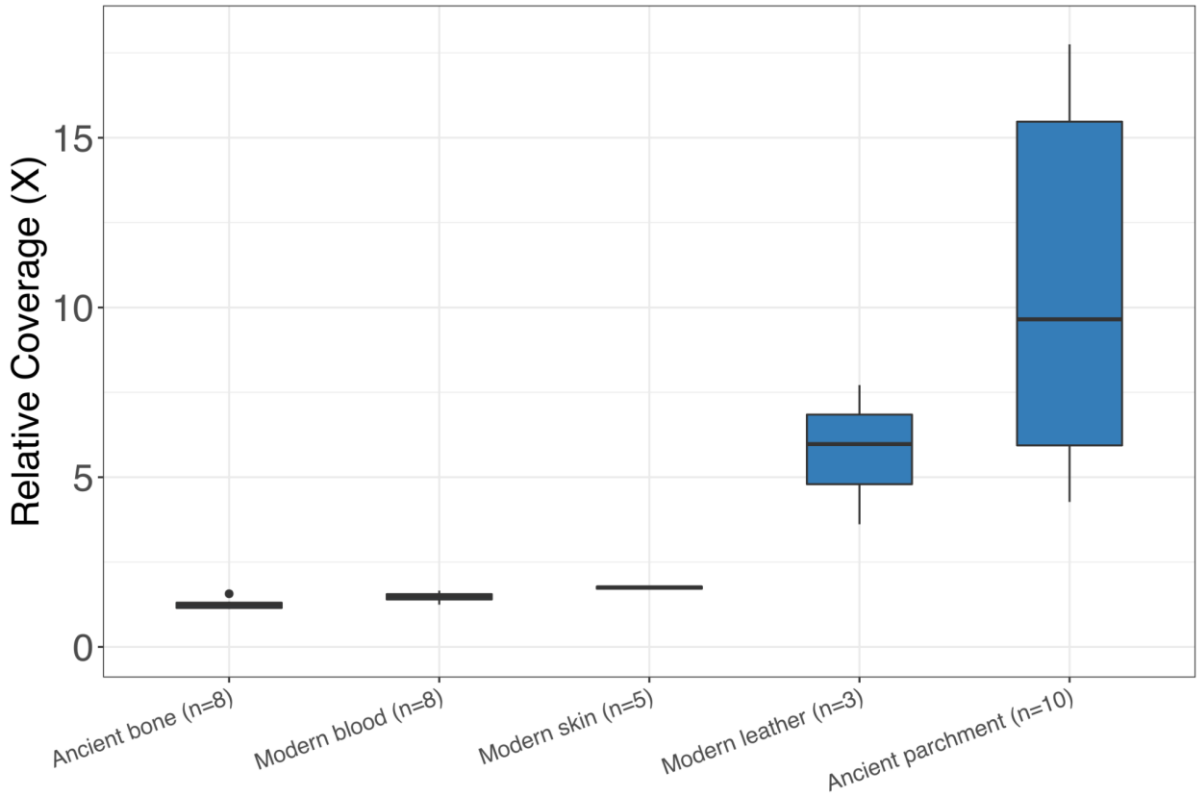


Figure 6.2 - Coverage of repetitive regions relative to genic regions in different bovine sequencing data. Untreated ancient and modern bovine tissues display a typical relative coverage between 1.2-1.8X. The modern leather samples and ancient parchment samples once again display a dramatic enrichment of repetitive sequences, with a relative coverage between 3.6-17.8X.

To further investigate elevated preservation of heterochromatin, survival of the X chromosome relative to the autosomes was investigated in male and female samples (Figure 6.3). The average coverage of X chromosomes in all leather samples was elevated compared to the average coverage of the control dataset. However, this is observed in both male and female samples, suggesting there is no extra protection afforded to facultative heterochromatin in the female Xi. An enrichment of 4-58% was observed in female X chromosomes compared to an 11-30% enrichment in male X chromosomes. All 5 of the male leather samples are outliers in the context of the comparative dataset compared to only 3 of the female leather samples.

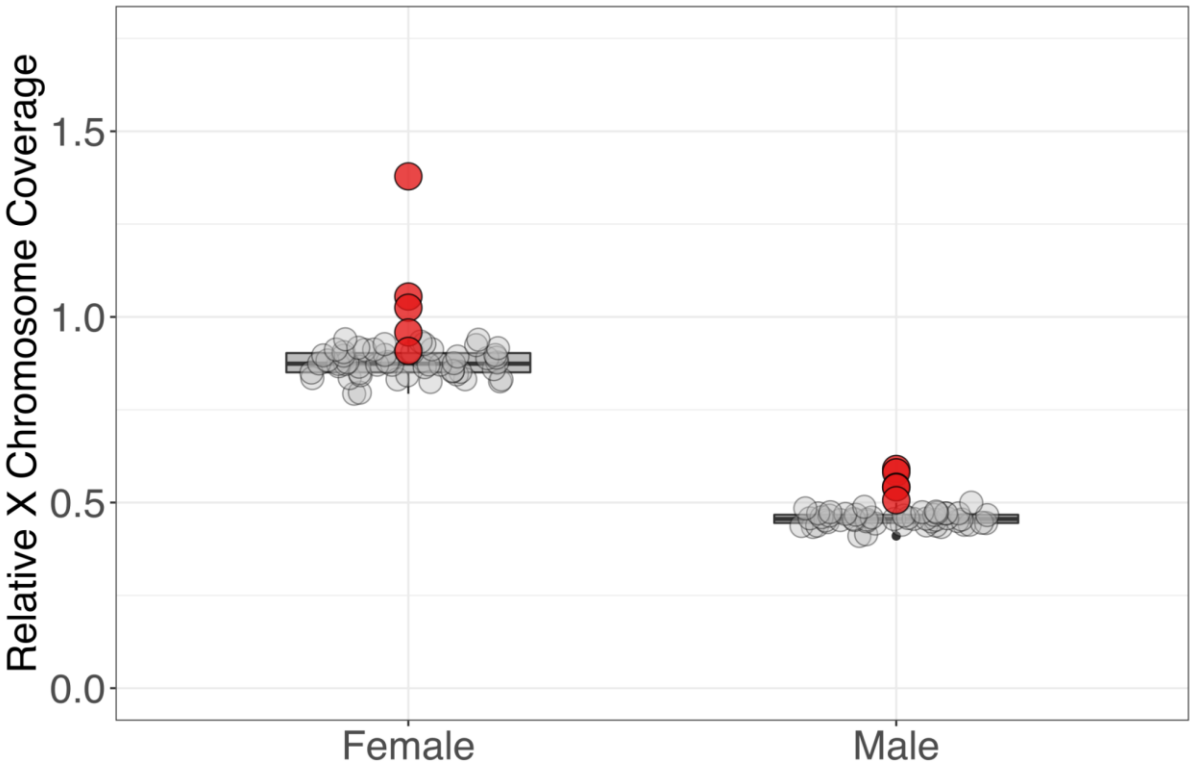


Figure 6.3 - Coverage of X chromosome relative to average autosome coverage in leather samples. There is an observed enrichment in X chromosome sequences in bovine leather samples (red) compared to non-leather bovine samples (grey).

The survival of mtDNA compared to nucDNA was also investigated. Here, the average ratio of mtDNA reads to nucDNA reads in modern leather samples is 1:3155, approximately similar to the ratio of ancient parchment (1:2969) (Appendix Table 6.3). This is a higher ratio of mtDNA reads compared to modern untanned bovine skin (1:7860), but lower than the average ratio of ancient bone (1:2346).

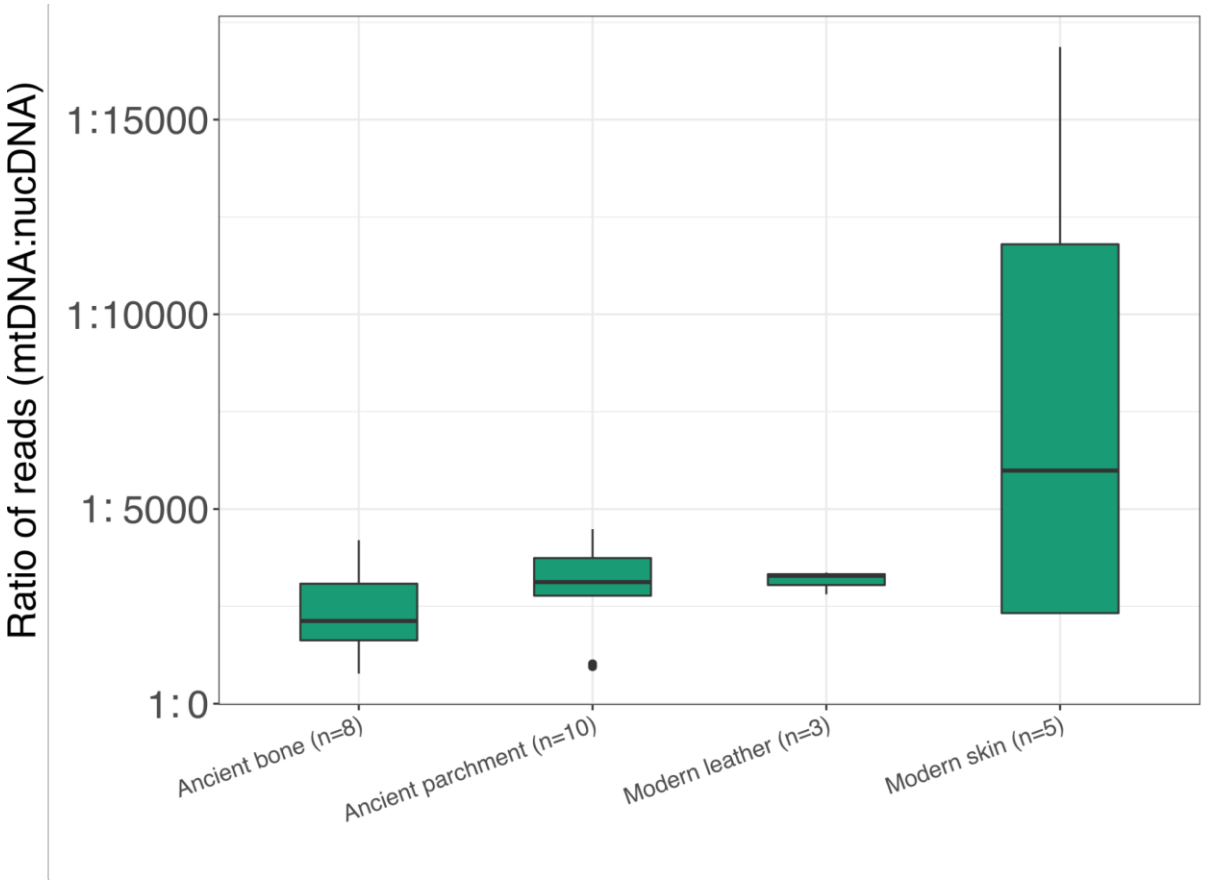


Figure 6.4 - Ratio of realigns aligned to mitochondrial genome and nuclear genome. although the ratio was notably more variable. The ratio was even more extreme in modern blood samples (1:760-1:861952), and for this reason is not displayed.

6.4.3. Damage patterns

Unusual damage patterns were observed within the NGS reads of modern leather DNA. Excess levels of base mismatches occurred near the ends of aligned reads (Figure 6.5). Similar patterns were observed for all possible base substitutions. Approximately 14% of the nucleotide bases directly before strand-breaks were mismatches with the genome assembly, which decreased steadily from the strand-breaks. While some of these base mismatches may be due to SNPs/sequencing error, the frequency of mismatches suggests that it is the result of DNA damage caused by the leather tanning process. Fragmentation of DNA was also high (Figure 6.6). The median read length ranged from 38-44 bp. Furthermore, >99% of reads were below 100 bp in all samples.

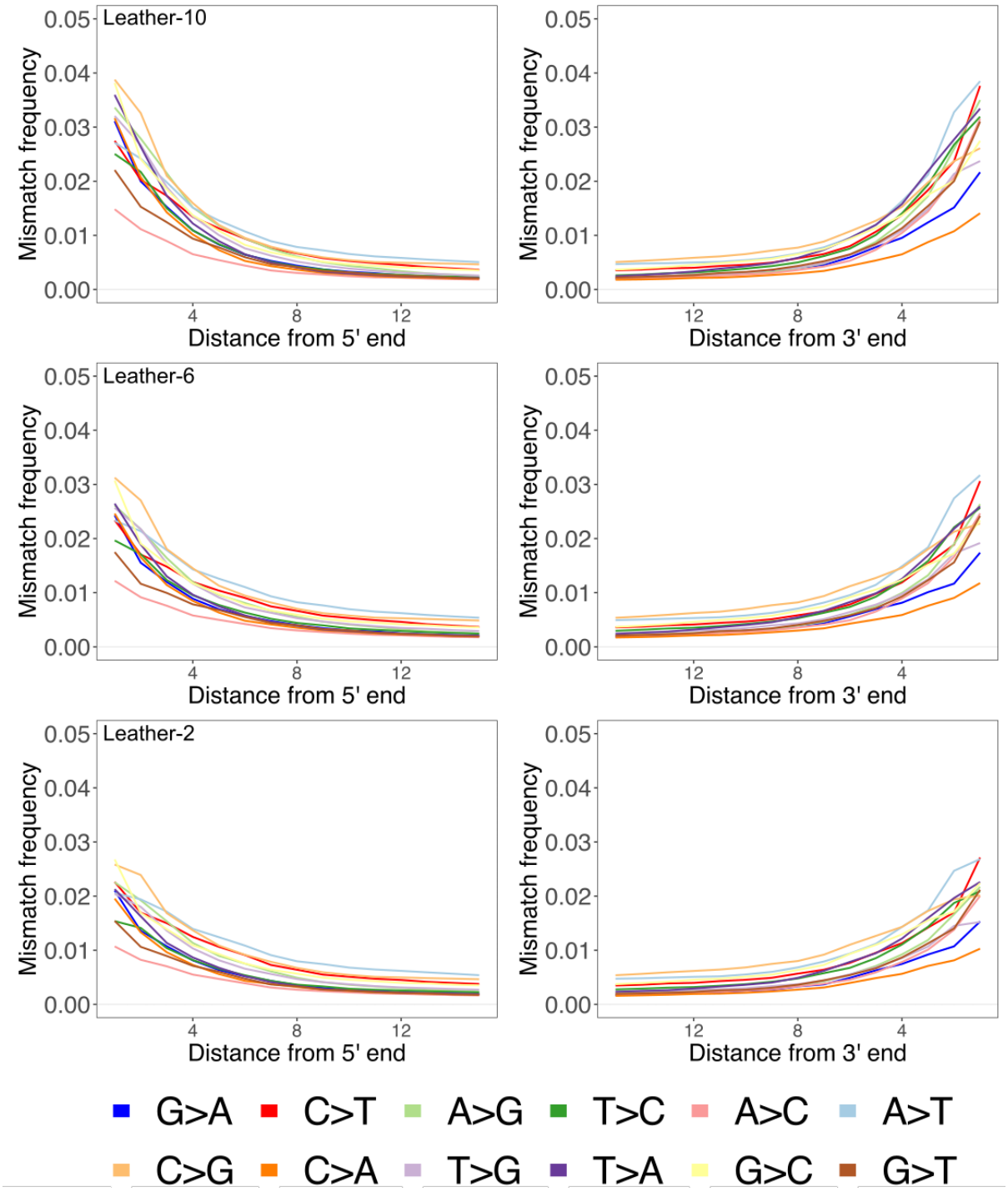


Figure 6.5 - DNA damage patterns in NGS reads, as assessed using mapDamage2.0. Every possible misincorporation is observed. Each misincorporation curve peaks at the strand-break point and decreases steadily. Some general patterns were observed across samples i.e. at the 5' end C>G and G>C had the highest frequency of mismatches while A>C had the lowest. Similarly, A>C and A>T had the highest mismatch frequency at the 3' end, with C>A having the lowest.

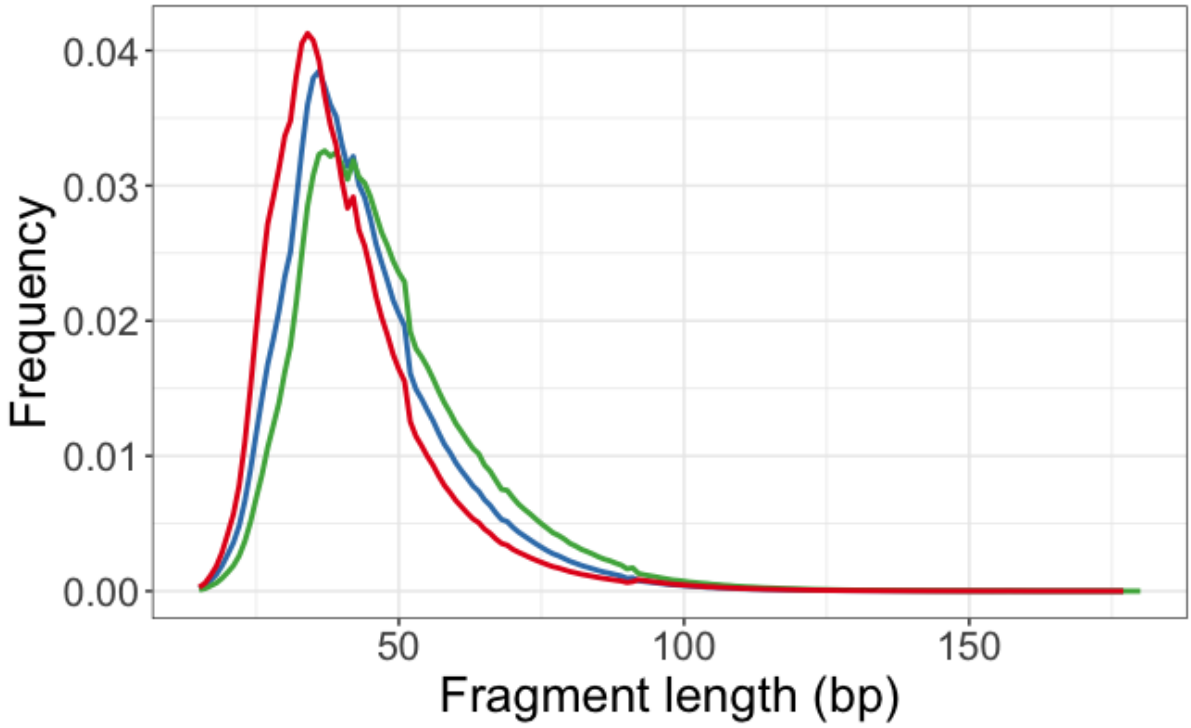


Figure 6.6 - Read length distribution of Leather-2, Leather-6 and Leather-10, as calculated by paired-end reads. Median read length was 41 bp, 44 bp and 38 bp, respectively.

6.5. Discussion

In this study DNA from modern Cr(III)-tanned bovine leather was successfully extracted, amplified and sequenced using techniques commonly used in aDNA analysis. The leather DNA is damaged, highly fragmented and displays an unusual genomic preservation pattern, likely due to the chemistry of the tanning process.

6.5.1. Differential DNA preservation

The unusual pattern of genomic preservation resulting in an overrepresentation of repetitive sequences may be explained by genome architecture. Tight binding of repetitive sequences to histones likely protects the DNA, while loosely bound euchromatin is more easily destroyed during the harsh leather tanning process. The preferential destruction of euchromatin may occur during the alkaline treatment of skin, which is present in both leather tanning and ancient parchment production (Teasdale et al. 2017). In highly alkaline conditions, deprotonation of double-stranded DNA results in denaturation as hydrogen bonds are broken between the two strands (Wang, Lim, and Son 2014). DNA that is tightly bound to basic histone proteins may be conferred some protection from degradation at this step. Notably, it has been suggested

that during leather tanning, the histone proteins may negatively affect nucDNA survival compared to mtDNA survival due to crosslinking with the histone proteins (Vuissoz et al. 2007). Instead, these histone proteins may actually provide protection through tight-binding to the DNA.

While this is a plausible explanation of the enrichment of autosomal repetitive sequences, such heterochromatin protection is not observed in the female Xi. As discussed, the second X chromosome is deactivated by facultative heterochromatin via tight histone-binding in female mammals (Wutz 2011). If all types of heterochromatin are expected to provide DNA protection during leather tanning, the female Xi should be better preserved during tanning and this should elevate X chromosome coverage relative to autosomal coverage. As males do not have a tightly-packed Xi, suchX chromosomal enrichment should only occur in females. However, both female and male leather samples see an enrichment of X chromosome sequences compared to the untreated samples, contrary to expectation. This suggests some complexity in heterochromic protection. Structure of the facultative heterochromatic Xi may be different to constitutive heterochromatin (Wutz 2011); also it occupies a perinucleolar location (Rego et al. 2008). The data suggests a nuanced preservation of heterochromatin with tight histone-packing having provided parts of the nuclear genome protection from the harsh leather tanning process.

While the mitochondrial genome does not contain any accessory proteins, the double-membrane of the mitochondrial genome could confer protection during leather tanning, as previously suggested (Vuissoz et al. 2007). The mitochondrial survival in tanned leather was approximately similar to that of ancient parchment. Although the sample size was small, both treated tissues had a higher ratio of mtDNA:nucDNA compared to modern untreated bovine skin. This suggests a preferential survival of mitochondria and emphasises that a number of factors relating to genomic architecture can contribute to the differential survivability of DNA through this process. Studies of mitochondrial survival suggest that energetic tissues like hair shafts may yield better mitochondrial survival compared to bone which generally have low copy numbers (Gilbert et al. 2004). For this reason, it is surprising to observe that mtDNA sequences had the highest ratio in sequencing data of ancient petrous bones rather than ancient, damaged or modern skin tissue. This suggests that the degree of protection of the double-membrane of mitochondria is poorly characterised in degraded tissues and warrants more attention.

6.5.2. DNA damage and fragmentation

Nucleotide misincorporation via DNA hydrolysis has been suggested to occur during the highly acidic pickling and salting stages of leather tanning (Matar and Merheb 2016). Post-mortem, acid-catalysed hydrolytic cytosine deamination accumulates over time on DNA strand overhangs resulting in C>T and G>A mismatches upon PCR amplification (Briggs et al. 2010). While these transitions are observed in the leather samples, hydrolytic deamination alone cannot account for this DNA damage since most of the observed misincorporations are not known to occur via hydrolytic deamination. Another possibility is that Cr(III) compounds, which are used as tanning agents, react with H_2O_2 in Fenton-like reactions to produce reactive hydroxyl radicals ($^{\bullet}OH$) that can cause base substitutions via DNA oxidation (Qi et al. 2000). However, H_2O_2 is not known to be widely used in leather tanning productions, so this is not likely to be a causative mechanism.

Instead, it is likely the observed base misincorporation is primarily caused by direct binding of Cr(III) compounds to DNA. Cr(III) compounds are generally considered to be non-carcinogenic due to their inability to enter cells (Eastmond, Macgregor, and Slesinski 2008). However, during the leather tanning process, DNA is likely released via cell lysis by osmotic shock during salting stages or alkaline lysis during liming stages (Islam, Aryasomayajula, and Selvaganapathy 2017). As a result, DNA would be exposed to the substantial amount of Cr(III) compounds used in leather tanning, culminating in DNA damage. Cr(III)-DNA adducts have previously been shown to cause single base substitutions in DNA (Voitkun, Zhitkovich, and Costa 1998; Quievryn et al. 2003; Zhitkovich 2005). It has been suggested that the Cr(III) binds to the phosphate backbone in DNA or to guanine residues, the latter previously supported by evidence that most single base substitutions occur at G/C pairs leading to resulting in G/C→T/A transversions and G/C→A/T transitions (Zhitkovich et al. 2001; Zhitkovich 2005; Quievryn et al. 2003). While these transitions are observed in the leather samples, they do not predominate other possible misincorporations.

Interactions between the Cr(III) compounds and DNA likely led to the highly fragmented DNA observed. Studies have shown that several Cr(III) compounds can cleave both supercoiled DNA (Vijayalakshmi et al. 2000)(Fang et al. 2014)(Vaidyanathan and Nair 2003), and linearised DNA (Fang et al. 2014). Several mechanisms of cleavage have been suggested, from direct interference with DNA structure (Fang et al. 2014), to oxidative stress (Vaidyanathan and Nair 2003) and DNA hydrolysis (Vijayalakshmi et al. 2000), and are likely to be specific to the Cr(III) compound in question. Acid-catalysed DNA hydrolysis can also cause strand cleavage via β-elimination by producing abasic sites (mostly apurinic) in the DNA

chain (Lindahl 1993). While DNA hydrolysis has been suggested to occur during the acidic stages of the tanning process, evidence of this occurring in the NGS reads was not observed. The short fragment length likely explains the difficulty in amplifying leather DNA using traditional PCR techniques, which rely on long fragments and high concentrations of DNA. This further emphasises the utility of using aDNA techniques to sequence degraded modern DNA as these techniques have been developed specifically for short DNA fragments.

7. Mitochondrial analysis of *Bos primigenius*

7.1. Introduction

Mitochondrial DNA was the marker of choice for many early phylogeographic studies due to several genetic and technical advantages for experimental design (Avise 2000). The high mutation rate of mtDNA is sufficiently rapid to create the required variability to test genetic distance in even recently diverged populations and as it is inherited through female lineages, no recombination can occur between generations, allowing for preserved haplotypes to be easily tracked (Ballard and Whitlock 2004). Specific primers to amplify conserved mitochondrial sequences allowed genotyping a range of species with relative ease. These were later replaced by enrichment strategies which exploited the high copy number and small haploid genome of mtDNA to make the capture and sequencing of complete mitochondrial genomes highly achievable on NGS platforms (Briggs et al. 2009). These attributes made mtDNA a popular target for ancient samples which typically have low levels of endogenous DNA (Gilbert et al. 2007); enrichment of mitochondrial sequences via capture hybridisation has been proven to be an efficient means of genotyping highly degraded ancient samples (O'Sullivan et al. 2016a; Briggs et al. 2009). Despite this utility, the use of the mitochondrion as a marker of evolution is not without its limitations. Mitochondrial trees do not always reflect true lineages and can result in poorly resolved species trees due to incomplete lineage sorting (ILS) and introgression. An example is found in the *Bison* lineage, where the mitochondria of European bison cluster with *Bos primigenius* despite its autosomal variation and morphology being much more similar to American bison (Verkaar et al. 2004). This has led to competing hypotheses of ILS or hybrid origins to explain the non-monophyly (Wang et al. 2018).

Initial genetic inferences of domestication and diversity of *Bos primigenius* were based in mtDNA analysis of modern cattle. Early studies using mitochondrial sequences uncovered a deep bifurcation between extant taurine and indicine cattle mitochondrial DNA (Loftus et al. 1994), which provided compelling genetic evidence of two separate domestication events of long-diverged populations of aurochs. Modern mitochondria of taurine cattle are dominated by the T mega-haplogroup which is characterised by geographically structured sub-haplotypes, with southwest Asia displaying the highest diversity (Troy et al. 2001; Achilli et al. 2008). Meanwhile, most indicine cattle fall in the I haplogroup, with the highest nucleotide diversity existing in the northern part of the Indian subcontinent (Chen et al. 2010). The publication of the first ancient *Bos* sequence from two Late Pleistocene British aurochs later supported southwest Asia as a domestication centre for taurine cattle, as this haplotype (later named

haplotype P) outgrouped all mitochondrial taurine diversity (Bailey et al. 1996). This suggested that aurochs mtDNA diversity likely did not survive their extinction and that the use of aDNA would be necessary to fully characterise this variation. As archaeological remains and cave paintings indicate that the aurochs were widely dispersed through most of Eurasia and North Africa during the Pleistocene, extant cattle may only represent a small proportion of the diversity that once existed (van Vuure 2005).

Extensive sequencing of ancient mtDNA from aurochs fossils revealed that the P haplotype predominated within Late Pleistocene and Holocene European aurochs (Edwards et al. 2007). Given that the defining European wild haplotype, P, was thought to have gone extinct along with European aurochs around 600 years ago it was surprisingly genotyped in modern east Asian cattle breeds (Mannen et al. 2020). This was despite the fact that early to mid Holocene east Asian aurochs had been shown to display a divergent haplotype (C) (Cai et al. 2018; Zhang et al. 2013). It was suggested that European P aurochs expanded their ranges eastward in the Holocene, and subsequently recruited into the domestic stock. This was plausible as after the intense cold of the Last Glacial Maximum (LGM), many warm-adapted animals expanded from southerly refugia to colonise the newly available habitats (Hewitt 1999). This is also observed in archaeological evidence of aurochs which suggests that European populations dwindled in numbers during the cold period before expanding again (Leonardi et al. 2022). While a star-like phylogeny of P haplotypes further suggests a population expansion in the post-LGM (Edwards et al. 2007), explicit evidence of migration across Central Asia is lacking. Similarly, mitochondrial introgressions from the wild were suggested by the sporadic presence of divergent haplotypes (Q and R) in some Italian breeds (Achilli et al. 2008; Bonfiglio et al. 2010). This was originally suggested to have been due to a third domestication centre in Italy, although ancient sampling of Neolithic southeast European cattle disproved this for Q (Scheu et al. 2015). Overall, this highlights the importance of sampling time-stamped ancients data in understanding the population structure of aurochs prior to domestication, how these wild populations interacted with domestic stock and their legacy in modern breeds today.

7.2. Aims

1. To characterise the relationship between different members of the *Bovini* Tribe using mitochondrial variation.
2. To construct Bayesian mitochondrial phylogenies of Bos lineages to detect phylogeographic structure.
3. To time the expansions of several mitochondrial lineages that are associated with *Bos* domestication events.

4. Define the European P haplotype and track its postglacial expansion.

7.3. Materials and methods

7.3.1. Materials

A total of 115 ancient whole mitochondrial genome FASTA files were constructed from NGS sequencing data (Table 7.1; Appendix Table 7.1). These genomes were supplemented by 27 ancient *Bos* and *Bison* samples that were downloaded from NCBI (Appendix Table 7.2). In addition, 107 modern *Bovini* samples were downloaded from NCBI (Appendix Table 7.3). In sum, 249 mitochondrial genomes were analysed in this study.

The oldest sample in this dataset is VEM212, an aurochs recovered in Norton Bottoms Quarry, Lincolnshire, England. The intact skull of VEM212 was recovered in a layer of interglacial sediment corresponding to Marine Oxygen Isotopic Stage (MIS) 7 (243-191 Ka).

7.3.2. Laboratory methods

Sample preparation is detailed in Section 2.1. DNA extraction and sequencing library construction are detailed in Section 2.2 and 2.3, respectively. Screening results for samples are described in Chapter 3.

For mitochondrial sequence enrichment, a bait and capture approach was employed using MYcroarray MYbaits® In-Solution Sequence Capture for Targeted High-Throughput Sequencing, as detailed in Section 2.3.5. The captured DNA was then reamplified for 16 cycles.

7.3.3. Reference fasta construction

Systemic bias and error may occur when a single reference genome is used in phylogenetic reconstruction of haploid genomes (Bertels et al. 2014). Phylogenetic reconstruction using mitochondrial sequences may be acutely vulnerable to such bias due to the high mitochondrial mutation rate, especially in the rapidly evolving D-loop. Several methods of ancient mitochondrial consensus constructions have been used to account for sequence divergence. For example, sequence data from an ancient hominin from Sima de los Huesos was aligned to the human mitochondrial reference, allowing five mismatches to account for sequence divergence and nucleotide misincorporation associated with aDNA damage (Meyer et al.

2013). Haplotype-specific genome references have also been used during alignment to improve coverage and consensus construction (Daly et al. 2021, 2018).

Mitochondrial reference genome assemblies for each major haplotype were constructed to produce higher quality consensus fasta files. To produce fasta references of newly described haplotypes (i.e. E, K, G, N) a “baiting and iterative mapping” approach was employed based on methods in (Hahn, Bachmann, and Chevreux 2013). Sequencing data from a single high-depth haplotype reference sample was first aligned to the “bait” T3 *Bos taurus* mitochondrial reference genome (V00654.1) and processed. A consensus fasta sequence was generated with a minimum depth of 1 using ANGSD (angsd -doFasta 2 -doCounts 1 -setMinDepth 1 -minQ 20 -minMapQ 30). This consensus fasta sequence was then manually edited to account for indels identified using IGV. This sequence was used as a primary reference genome for a subsequent re-alignment, consensus fasta generation and indel correction of the same high-depth haplotype reference sample. This process was iterated until an alignment with no indels or mismatches was achieved. The final consensus fasta sequence was generated with a minimum depth of 5 using ANGSD (angsd -doFasta 2 -doCounts 1 -setMinDepth 5 -minQ 20 -minMapQ 30). This consensus fasta sequence was used as a reference genome for its respective haplotype.

As mitochondrial genomes are circular, sequencing reads that straddle the artificial start or endpoint of the genome assembly may be lost. To this end, each reference fasta was circularised by concatenating the last 15 bp of the fasta reference to the start of the reference and concatenating the first 15 bp to the end of the reference fasta using an in-house script developed by Dr Kevin Daly.

7.3.4. FASTA construction

To first determine the haplotype of each sample, filtered sequencing reads were first aligned to the *Bos taurus* mitochondrial reference genome (V00654.1) and a consensus sequence was generated using ANGSD (angsd -doFasta 2 -doCounts 1 -setMinDepth 3 -minQ 20 -minMapQ 30). The fasta was combined with reference fastas representing each of the major haplogroups and loaded into SeaView Version 5.0.1. The sequences were aligned using MUSCLE and a Maximum Likelihood tree was generated using PhyML. The model chosen was GTR with 100 bootstraps. The resulting tree was inspected and a haplotype was assigned when possible. All sequencing data was then aligned to the reference genome of the assigned haplotype, as above.

Sequencing data of VEM212 was aligned against each available *Bos* haplotype reference and an outgroup *Bison* reference to determine the best reference. A Maximum Likelihood tree was constructed for each of the consensus calls.

7.3.5. Bayesian analysis

BEAST analyses were performed using BEAST v2.6.6. Four datasets were curated (Appendix Table 7.4):

-dataset01 - a *Bovini* tribe dataset containing modern and ancient sequences of *Bos* and *Bison*
-dataset02 - a dataset restricted to ancient and modern *Bos primigenius* samples only.
-dataset03 - a dataset restricted to ancient and modern *Bos primigenius* samples only with some undated Pleistocene samples excluded.
-dataset04 - a dataset restricted to ancient European *Bos primigenius* samples of the P haplotype.

Partitioning of data was also done for some BEAST runs of dataset02 and dataset03. The partitions were based on the NCBI partition annotations of the *Bos taurus* mitochondrial reference genome (V00654.1). The partitions were tRNA, rRNA, the first and second codon positions of gene regions (C1+2), the third codon position (C3), the D-loop, and the remainder of the molecule.

The sequence data were prepared by performing a multiple sequence alignment using MUSCLE, converted to a Nexus file using SeaView Version 5.0.1 and loaded into BEAUTi. For ancient samples, tip dates were set to the midpoint age of the radiocarbon 95% confidence interval. The site model was chosen by bModelTest while the clock model was set to Relaxed Clock Log Normal. A Coalescent Bayesian Skyline model was chosen as the tree model and a BEAST xml file was generated.

BEAST analyses of the three databases were performed using BEAST v2.6.6. For each dataset, two independent BEAST runs of 250 million chains were performed, partitioned and unpartitioned where applicable. The MCMC chain was sampled every 25,000 iterations, with 10% burn-in. Log files were assessed using Tracer to inspect parameter convergence. A Bayesian Skyline Plot was produced for dataset04 in Tracer with default settings. Successful runs were combined using LogCombiner and a maximum clade credibility tree was constructed with median node heights using TreeAnnotator. The resulting tree was visualised with FigTree.

7.3.6. Defining P sub-haplotypes

The P haplogroup is the most commonly found haplotype in ancient European aurochs (Edwards et al. 2007). Several P sub-haplotypes have been suggested although nomenclature has been inconsistent (*i.e.* Mannen et al. 2020 and Cubric-Curik et al. 2021). To track these maternal lineages in a meaningful way, sub-haplotypes were defined if the timing of the LGM (27,000-18,000 years ago) overlapped with the 95% HPD age interval of an internal node.

Table 7.1 - Summary of ancient mitochondrial consensus sequences generated in this Chapter. Context is reproduced as received from the curator unless the sample is radiocarbon dated. * = sample dated by context; ^ sample radiocarbon dated indirectly. Laboratory work is defined as any handling of the sample from sample drilling to sequencing of the screening data. AH=Ashleigh Haruda, AS = Amelie Scheu, BL = Bastien Llamas, CR = Conor Rossi, ER = Erika Rosengren, KD = Kevin Daly, KM = Kieren Mitchell, LF = Laurent Frantz, MS = Mikkel Sinding, MV = Marta Verdugo, VM = Valeria Mattiangeli, VEM = Victoria E. Mullin.

Sample ID	Laboratory ID	Species	Haplotype	Country	Region/Site	Average mtDNA coverage	Given Context/Age	Laboratory Work
Baikal1	MV189	*Bos primigenius*	C	Russia	Barguzin river	1697.96	12015 - 12529 cal yrBP	MV/CR
Bor2	CR101	*Bos primigenius*	C	Kazakhstan	Borly	80.36	4913 - 5099 cal yrBP	CR
ROS001	ROS001	*Bos primigenius*	C	Kazakhstan	Roschinskoe	3.68	5539 - 5372 cal yrBP	KD/CR
Tula1	MV194	*Bos primigenius*	C	Russia	Tula district	2443.07	40746 - 42460 cal yrBP	MV/CR
Ust1	CR111	*Bos primigenius*	C	Kazakhstan	Ust ' -Narym	1.97	Unknown	CR
Yen1	Yen1	*Bos primigenius*	C	Russia	Yenisei River	771.29	14193 - 14019 cal yrBP	KM/BL
UKR08	CR108	*Bos primigenius*	E	Ukraine	Starobels	10.71	9000 - 8000 yrBP*	CR
UKR11	CR109	*Bos primigenius*	E	Ukraine	Starobels	6.07	9000 - 8000 yrBP*	CR
Rhi1	CR180	*Bos primigenius*	G	Germany	Upper Rhine Valley	1222.79	47000+ cal yrBP	CR
Rhi2	CR181	*Bos primigenius*	G	Germany	Upper Rhine Valley	1582.56	47009 - 45312 cal yrBP	CR
Rhi4	CR183	*Bos primigenius*	G	Germany	Upper Rhine Valley	40.40	Pleistocene*	CR
Rhi6	CR185	*Bos primigenius*	G	Germany	Upper Rhine Valley	49.86	Pleistocene*	CR
Rhi7	CR186	*Bos primigenius*	G	Germany	Upper Rhine Valley	19.00	Pleistocene*	CR
VEM212	VEM212	*Bos primigenius*	G	England	Norton Bottoms	3.05	MIS 7 (by stratigraphy)	VEM
Bor1	CR100	*Bos primigenius*	K	Kazakhstan	Borly	221.10	6740 - 6521 cal yrBP	CR

Gyumri1	MV198	*Bos primigenius*	K	Armenia	Gyumri district	1393.30	14097 - 13863 cal yrBP	MV/CR
Ura1	CR115	*Bos primigenius*	K	Kazakhstan	Uralsk	5.21	Unknown	CR
Var2	CR116	*Bos primigenius*	K	Kazakhstan	Varfolomeevka	2469.01	7486 - 7238 cal yrBP	CR
Rhi5	CR184	*Bos primigenius*	N	Germany	Upper Rhine Valley	36.33	Pleistocene*	CR
Rhi8	CR187	*Bos primigenius*	N	Germany	Upper Rhine Valley	6.47	Pleistocene*	CR
Bar1	CR119	*Bos primigenius*	P	Italy	Barche di Solferino	69.26	13246 - 13005 cal yrBP	MS/CR
Bed3		*Bos primigenius*	P	Germany	Bedburg-Königshoven	5004.96	11875 - 11399 cal yrBP	AS
Bed4		*Bos primigenius*	P	Germany	Bedburg-Königshoven	702.44	11949 - 11476 cal yrBP	AS
Belarus1	MV197	*Bos primigenius*	P	Belarus	Brest district	333.68	4476 - 4173 cal yrBP	MV
Blec1	CR118	*Bos primigenius*	P	Bulgaria	Vratsa	87.00	6258 - 6006 cal yrBP^	CR
Bor3	CR102	*Bos primigenius*	P	Kazakhstan	Borly	931.30	5115 - 4925 cal yrBP	CR
Fal7		*Bos primigenius*	P	Germany	Falkenwalde	186.42	1000 yrBP*	AS
Fif1	CR166	*Bos primigenius*	P	Scotland	Newburgh, Fife	52.30	10634 - 10277 cal yrBP	MS/CR
Fre1	CR173	*Bos primigenius*	P	Denmark	Frederiksborg, Alsønderup	24.75	10825 - 10496 cal yrBP	MS/CR
Gal1	CR120	*Bos primigenius*	P	Spain	Galicia	939.13	9548 - 9113 cal yrBP	MS/CR
Gal2	CR170	*Bos primigenius*	P	Spain	Galicia	151.52	9481 - 9100 cal yrBP	MS/CR
Gal3	CR176	*Bos primigenius*	P	Spain	Galicia	1445.04	9538 - 9109 cal yrBP	MS/CR
Gallo1	CR165	*Bos primigenius*	P	Scotland	Galloway	144.91	3764 - 3479 cal yrBP	MS/CR
Hjo1	CR174	*Bos primigenius*	P	Denmark	Hjørring, Tofte Bæk	354.54	8245 - 8053 cal yrBP	MS/CR
Horn_10537		*Bos primigenius*	P	Norway	The royal Horn of Norway	411.02	400 yrBP*	MS
Horn_10543		*Bos primigenius*	P	Unknown	Unknown provenance	566.02	600 yrBP*	MS

Horn_CMLIX		*Bos primigenius*	P	Germany	Firemen's Guild, Büsum, Holstein	356.55	400 yrBP*	MS
Hxh2		*Bos primigenius*	P	Germany	Herxheim	2529.15	5000 yrBP*	AS
Kalisz1	MV192	*Bos primigenius*	P	Poland	Kalisz	370.45	4487 - 4229 cal yrBP	CR
Kalisz2	MV195	*Bos primigenius*	P	Poland	Kalisz	513.54	5813 - 5668 cal yrBP	(Verdugo et al. 2019)
Kok1	MV239	*Bos primigenius*	P	Uzbekistan	Kok Tepe	94.91	5000-3500 yrBP*	MV/CR
Lil1	CR171	*Bos primigenius*	P	Denmark	Randers, Bønnerup Strand	478.92	4484 - 4228 cal yrBP	MS/CR
Lun1	CR178	*Bos primigenius*	P	Sweden	Skåne	66.29	9744-9525 / 9045-8656 cal yrBP^	CR
Lun2	CR179	*Bos primigenius*	P	Sweden	Skåne	19.87	9744-9525 / 9045-8656 cal yrBP^	CR
Pad1	CR121	*Bos primigenius*	P	Italy	Padova	45.37	12267 - 11844 cal yrBP	CR
Pal1	CR125	*Bos primigenius*	P	Italy	Central Italy	670.03	16287 - 15849 cal yrBP	CR
Rhi3	CR182	*Bos primigenius*	P	Germany	Upper Rhine	1075.22	Pleistocene*	CR
ROS002	ROS002	*Bos primigenius*	P	Kazakhstan	Roschinskoe	7.34	5116 - 4926 cal yrBP	KD/CR
Siv1	CR117	*Bos primigenius*	P	Bulgaria	Vratsa	1036.55	6258 - 6006 cal yrBP	CR
Ska1	CR167	*Bos primigenius*	P	Sweden	Skåne	588.17	9610 - 9207 cal yrBP	MS/CR
Ska2	CR168	*Bos primigenius*	P	Sweden	Skåne	4.18	10294 - 9896 cal yrBP	CR
Ska3	CR169	*Bos primigenius*	P	Sweden	Skåne	1105.77	9731 - 9508 cal yrBP	MS/CR
Ton1	CR172	*Bos primigenius*	P	Denmark	Jejsing	1091.87	3279 - 3074 cal yrBP	MS/CR
Tri1		*Bos primigenius*	P	Germany	Trier-Brüderkrankenhaus	907.19	10291 - 9976 cal yrBP	AS
UKR07	CR107	*Bos primigenius*	P	Ukraine	Starobels	39.91	9000 - 8000 yrBP*	CR
Uzz1	CR126	*Bos primigenius*	P	Italy	Sicily	486.49	7593 - 7480 cal yrBP	CR
	CR122	*Bos primigenius*	P	Italy	Padova	0.68	Late Upper Palaeolithic*	CR

	CR133	*Bos primigenius*	P	Sweden	Bjäreshög	0.41	9555 - 9218 cal yrBP	CR
	CR134	*Bos primigenius*	P	Sweden	Rönneholm	0.49	8837 - 8504 cal yrBP	CR
	CR147	*Bos primigenius*	P	Sweden	Frörum	0.36	9497 - 9107 cal yrBP	CR
	CR148	*Bos primigenius*	P	Sweden	Lund	0.23	10193 - 9611 cal yrBP	CR
	CR149	*Bos primigenius*	P	Sweden	Lund	0.22	10193 - 9507 cal yrBP	CR
	ER003	*Bos primigenius*	P	Sweden	Haväng	14.27	9044 - 8613 cal yrBP	ER/CR
	ER004	*Bos primigenius*	P	Sweden	Haväng	0.90	10779 - 10368 cal yrBP	ER/CR
	ER005	*Bos primigenius*	P	Sweden	Stehag	0.78	9065 - 8719 cal yrBP	ER/CR
	ER023	*Bos primigenius*	P	Sweden	Jussberg	3.25	8392 - 8009 cal yrBP	ER/CR
	ER027	*Bos primigenius*	P	Sweden	Västra Tollstrad	1.49	9073 - 8719 cal yrBP	ER/CR
	ER028	*Bos primigenius*	P	Sweden	Västra Tollstrad	0.37	9067 - 8719 cal yrBP	ER/CR
Bes2	MV048	*Bos taurus*	Q	Iraqi Kurdistan	Bestansur	141.84	624 - 731 cal yrBP	(Verdugo et al. 2019)
Kalisz3	MV196	*Bos primigenius*	Q	Poland	Kalisz	632.67	4477 - 4168 cal yrBP	MV/CR
Men2	MV013	*Bos taurus*	Q	Turkey	Menteşe	131.33	7998 - 7843 cal yrBP	(Verdugo et al. 2019)
Plo1	MV136	*Bos taurus*	Q	Serbia	Pločnik	11.55	7275/7125 - 7100/7000 cal yrBP^	(Verdugo et al. 2019)
Plo2	MV137	*Bos taurus*	Q	Serbia	Pločnik	40.73	7025/6950 - 6850 - 6650 cal yrBP^	(Verdugo et al. 2019)
Gyumri2	MV200	*Bos primigenius*	Q - basal	Armenia	Gyumri district	480.64	7394 - 7239 cal yrBP	(Verdugo et al. 2019)
Var1	CR110	*Bos primigenius*	Q - basal	Kazakhstan	Varfolomeevka	367.71	7641 - 7500 cal yrBP	CR
Th7		*Bos primigenius*	R	Morocco	Taghit Haddouch	21.26	9073 - 8730 cal yrBP	(Verdugo et al. 2019)
Yer1	MV240	*Bos taurus x Bos indicus*	T3	Uzbekistan	Yerqorqan	181.98	2100 yrBP*	MV/CR

Shau1	CR103	*Bos taurus*	T1	Kazakhstan	Shauke	76.14	Early Bronze Age*	CR
Ace1	MV135	*Bos taurus*	T2	Turkey	Acemhöyük	22.46	4923 - 4323 cal yrBP^	(Verdugo et al. 2019)
Has3	MV169	*Bos taurus*	T2	Iran	Hasanlu	231.78	4470 cal BP^	(Verdugo et al. 2019)
Plo4	MV139	*Bos taurus*	T2	Serbia	Pločnik	183.64	7100/7000 - 7025/6950 cal yrBP^	(Verdugo et al. 2019)
Plo8	MV162	*Bos taurus*	T2	Serbia	Pločnik	26.49	7100/7000 - 7025/6950 cal yrBP^	(Verdugo et al. 2019)
Qaz1	MV172	*Bos taurus*	T2	Iran	Tepe Shizar, Qazvin	207.53	4905 - 4659 cal yrBP	(Verdugo et al. 2019)
Bes1	MV047	*Bos taurus x Bos indicus*	T3	Iraqi Kurdistan	Bestansur	144.17	2934 - 2632 yrBP*	(Verdugo et al. 2019)
Ch22		*Bos primigenius*	T3	Turkey	Çatalhöyük	52.72	7750 - 7600 cal yrBP^	(Verdugo et al. 2019)
Horn_11570		*Bos taurus*	T3	Poland	Hunting Horn of last aurochs bull	545.05	700 yrBP*	MS
Horn_382		*Bos taurus*	T3	Denmark	Lykkesholm, Fyn	812.98	600 yrBP*	MS
Horn_617		*Bos taurus*	T3	Norway	Bjørnsund, Norway	747.5	600 yrBP*	MS
Plo3	MV138	*Bos taurus*	T3	Serbia	Pločnik	164.19	7100/7000 - 7025/6950 cal yrBP^	(Verdugo et al. 2019)
Plo7	MV160	*Bos taurus*	T3	Serbia	Pločnik	21.01	7100/7000 - 7025/6950 cal yrBP^	(Verdugo et al. 2019)
Shi1	CR104	*Bos taurus*	T3	Kazakhstan	Shiderty River	100.23	3983 - 3772 cal yrBP	CR
Tango1	MV198	*Bos primigenius*	T3	Former Soviet Union	Unknown	1406.86	4142 - 3914 cal yrBP	MV
Tango2	MV199	*Bos primigenius*	T3	Former Soviet Union	Unknown	308.28	4599 - 4379 cal yrBP	MV
YoA	YoA	*Bos primigenius*	T3	England	Yorkshire	463.56	4909 - 4658 cal yrBP	AH/LF/CR
	CR129	*Bos taurus*	T3	France	Normandy	27.52	Roman Era*	CR
	ER007	*Bison bonasus*	Bb1 (Bison)	Sweden	Skåne	0.82	10232 - 9563 cal yrBP	ER/CR

	ER008	*Bison bonasus*	Bb1 (Bison)	Sweden	Lommaryd	3.09	11254 - 10456 cal yrBP	ER/CR

7.4. Results

7.4.1. Mitochondrial capture

In total, 40 samples were selected for mitochondrial sequence enrichment with varying degrees of success (Appendix Table 7.1). Enrichment was not calculable for 4 samples due to a lack of shotgun sequencing aligned reads. 21 samples had an enrichment of at least 10X and 14 samples had an enrichment of at least 100X. Meanwhile, 6 samples returned no aligned reads after enrichment. The consistently poor preservation in a batch of samples from Sweden discussed in Chapter 3 were reflected in the lower success rate of mitochondrial enrichment. The average enrichment of these samples was ~47X compared to the enrichment of ~408X in other samples. Additionally, 10 of the 29 captured Swedish samples had no mitochondrial enrichment compared to shotgun sequencing which greatly impeded genotyping.

7.4.2. Haplotype-specific alignment

Alignment of sequencing data to haplotype-specific reference genomes resulted in a higher mean genome coverage compared to alignment against the *Bos taurus* reference genome (V00543) for samples that were not haplotype T. Using haplotype specific reference genomes also resulted in less uncalled bases (Ns) in consensus fasta generation at a minimum 3X coverage. This was especially evident in samples that were more divergent from the reference genome like haplotypes R, C and K. A selection of comparative samples is displayed in Table 7.2.

Table 7.2 - Enrichment of coverage when aligning to a haplotype-specific reference genome.

Sample	Haplotype	Mean coverage (haplotype specific)	Number of Ns called (haplotype specific)	Mean Coverage (V00543)	Number of Ns called (V00543)	Enrichment of coverage (X)
Kalisz3	Q	480.64	N/A	451.26	40	1.07
Blec1	P	87	N/A	75.15	184	1.16
Rhi6	G	49.86	N/A	40.1	406	1.24
Rhi5	N	37.19	58	29.15	830	1.28
Th7	R	21.26	126	10	2364	2.13
Bor1	K	221.1	N/A	89.24	71	2.48
Bor2	C	80.36	N/A	32.06	83	2.51

Alignment of VEM212 sequencing data against the N reference genome yielded the highest average coverage and lowest number of Ns called at minimum 3X coverage (Table 7.3). In all cases, the consensus formed a clade with the other N sequences except for the consensus called when aligned to a Bison reference (*Bb1*). In this case, VEM212 outgrouped all *Bos* clades but ingrouped from *Bison*.

Table 7.3 - Coverage and alignment statistics for VEM212 when aligned to a haplotype-specific reference genome.

	T	Q	P	R	C	K	E	N	I	Bb1	G
Coverage	2.91	2.91	2.92	2.93	2.95	2.93	2.9	3.05	2.88	1.99	3.02
Number of Ns @ 3X	9892	9908	9813	9816	9737	9864	9905	9419	9931	12230	9636
Reads aligned	1098	1089	1100	1107	1107	1099	1116	1152	1103	758	1144
ML	N	N	N	N	N	N	N	N	N	OUT	N

7.4.3. *Bovini* phylogeny

A Bayesian phylogeny was constructed using modern and ancient *Bovini* sequences (Figure 7.1). Extant domestic taurine and indicine cattle and extinct aurochs haplotype populate the *Bos primigenius* clade. The mitochondrial sequences from pre-LGM European Pleistocene aurochs (the British VEM212 and Rhine samples) outgroup all extant taurine and indicine animals but remain within *Bos primigenius* variation. These sequences are denoted by a star.

The discordant mitochondrial placement of *Bison* emerges in this phylogenetic tree. European wisent (*Bison bonasus*) are placed closer to the *Bos primigenius* clade than to the extinct Steppe bison (*Bison priscus*) or to the extant American bison (*Bison bison)*. The wisent clade features a deep split between the two major haplogroups, *Bb1* and *Bb2.* All American bison samples form a sub-clade within the larger clade of Caucasian/American bison. Of the three Scandinavia *Bison* mitochondria sequenced in this project, two were *Bb1* while one was *Bb2*. The other *Bos* species (i.e. Gaur, Gayal and Banteng) displayed poor species sorting, as has been previously reported (Sinding et al. 2021). Some Gaur samples form clades with both Banteng and Gayal despite deep divergences for other haplotypes (detailed in Appendix Table 7.3).

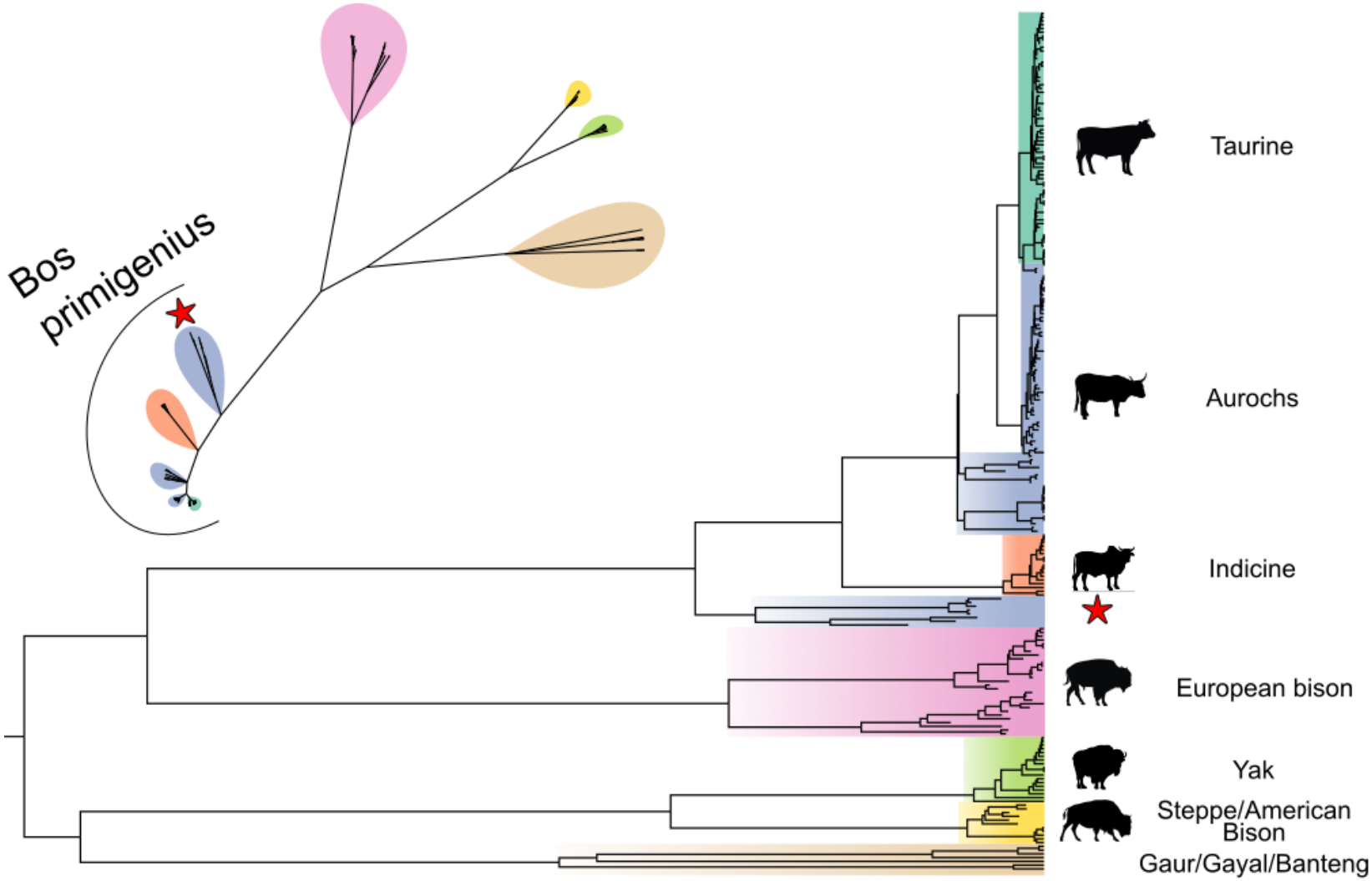


Figure 7.1 - A maximum clade credibility Bayesian phylogeny constructed from mitochondrial sequences of ancient and modern Bovini samples (dataset01).

7.4.4. Bayesian analysis of *Bos primigenius*

The Bayesian *Bos primigenius* phylogenetic tree expands data within the established mitochondrial haplotypes while also uncovering multiple previously undescribed mitochondrial lineages. Here, the newly described, divergent G and N haplotypes (notation from this work) outgroup all other lineages.

The partitioned and unpartitioned BEAST runs for dataset02 (all *Bos primigenius* samples) did not converge after 250 million chains due to poor mixing in some parameters. In particular, the height parameter did not converge for VEM212 or any Rhine samples in the partitioned BEAST runs (ESS <100). Mixing was better in the unpartitioned runs for these parameters but still unconverged for some samples (ESS <100 for height parameter of Rhi4, Rhi6 and Rhi7). The unpartitioned runs were merged with 10% burnin, and a tree was generated. Separately, the partitioned trees were merged with 10% burnin, and a tree was generated. The overall tree topologies were identical except for the divergence of the K and C haplotypes. In the partitioned tree, KC diverge together from REPQT haplotypes (KC|REPQT) while in the unpartitioned tree, the CKRE haplotypes diverge from the PQT haplotypes together (KCRE|PQT). Also, in the partitioned tree, VEM212 outgroups all sequences while it appears to be a very basal N haplotype in the unpartitioned tree. Age estimates of VEM212 and the Rhine samples from the unpartitioned BEAST run are displayed in Table 7.4.

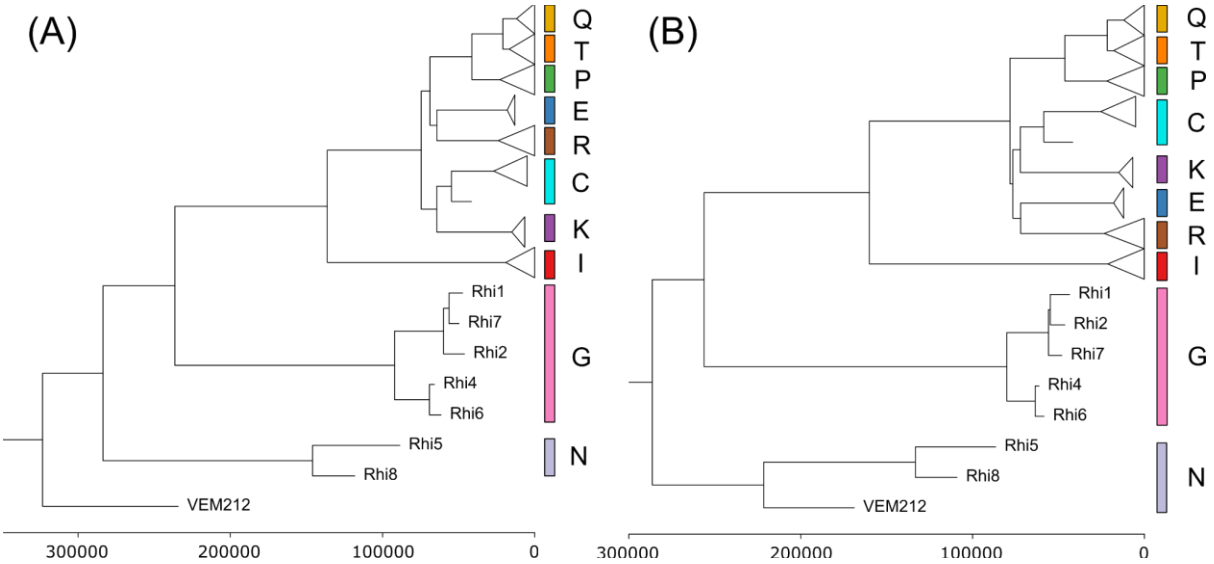


Figure 7.2 - A maximum clade credibility Bayesian phylogeny constructed from mitochondrial sequences of ancient and modern Bos primigenius samples (dataset02) using apartitioned (A) and unpartitioned (B) scheme. The overall tree topology and divergence timings remains similar except for divergence of K and C haplotypes and the placement of VEM212.

Table 7.4 - The age estimates of undated Pleistocene aurochs based on the unpartitioned BEAST run.

Sample	Haplogroup	Unpartitioned 95% HPD interval	Unpartitioned Median value	Unpartitioned ESS
VEM212	Basal N	133,160-349,820	233,670	180
Rhi1	G	25,575-70,131	46,201	146
Rhi4	G	25,176-107,180	64,282	182
Rhi5	N	59,134-145,040	90,023	20
Rhi6	G	21,833-102,470	59,904	190
Rhi7	G	29,556-75,806	48,498	240
Rhi8	N	74,692-150,000	119,410	142

All partitioned and unpartitioned BEAST runs for dataset03 (All *Bos primigenius* excluding the undated Pleistocene aurochs) converged after 250 million chains, with most parameters reaching an ESS of >300. The tree topology for the partitioned and unpartitioned tree differed on the divergence of the KC haplotypes (KC|REPQT in the unpartitioned tree, KCRE|PQT in the partitioned tree), once again reflecting the same pattern observed in Figure 7.X. The maximum clade credibility tree of the partitioned run is displayed in Figure 7.X. The deepest split is G|IKCREPQT, which had a median age of 267,633 years ago, corresponding to MIS 8. After the taurine/indicine split, a rapid emergence of five haplogroups results in what may be described as a pentafurcation. K, C, R, E, and PQT all diverged between 89,840 and 59,952 HPD 95% years ago. All divergences are well-supported for these with a prior <0.9, except for RE|PQT (prior=0.33).

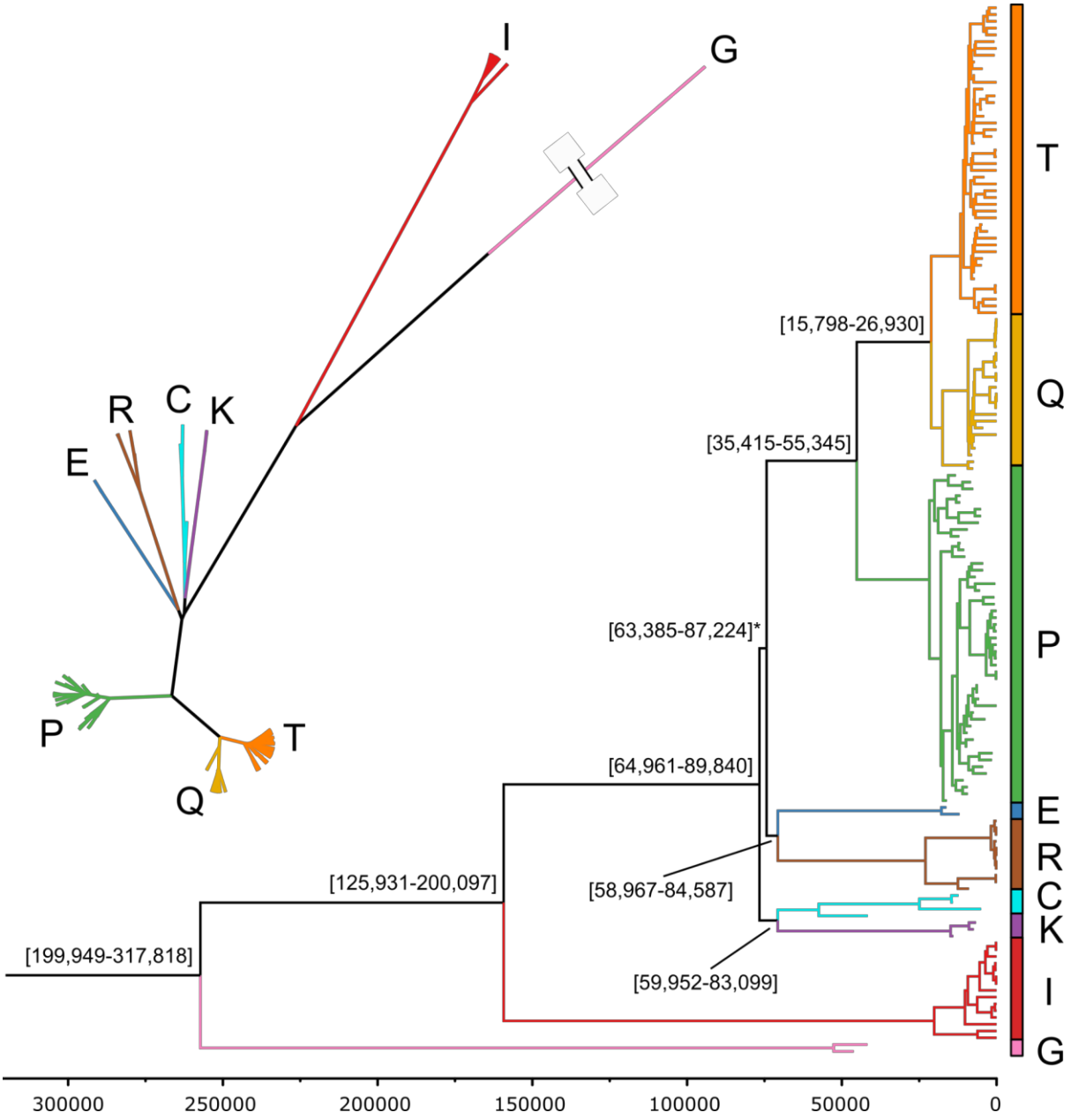


Figure 7.3 - A maximum clade credibility Bayesian phylogeny constructed from mitochondrial sequences of ancient and modern Bos primigenius samples (dataset03) using a partitioned scheme. The newly described G haplotype outgroups all previously known *Bos primigenius* variation. Further along the tree, the well-characterised taurine versus indicine split is observed. A pentafurcation of five *B. p. primigenius* lineages is then observed for the K, C, R, E and PQT lineages. A star-like phylogeny is similarly observed in the P haplotype. Several tooth-comb structures are observed in domestic lineages of T, Q, P, R and I at varying depths in time.

Table 7.5 - Node ages of haplotype divergences based on the partitioned BEAST run of dataset03.

Tree Node	Median node age (yrBP)	Node age 95% HPD (yrBP)	Node Prior
G|IKCREPQT	257280	199,949-317,818	1
I|KCREPQT	159281	125,931-200,097	0.9998
KC|REPQT	76607	64,961-89,840	1
RE|PQT	74322	63,385-87,224	0.3151
K|C	70708	59,952-83,099	0.9441
R|E	70675	58,967-84,587	0.9662
P|QT	45183	35,415-55,345	1
Q|T	21129	15,798-26,930	1

Substitution rates for each partition were calculated for the partitioned run of dataset03 (Table 7.6). The mean substitution rate was highest in the “rest of molecule” partition followed by the D-Loop. The lowest mean mutation rate was observed in the rRNA partition.

Table 7.6 - Substitution rates for each mitochondrial partition used in dataset03.

Partitions	Mean	Median	SD	95% HPD	ESS
D Loop	3.61E-07	3.57E-07	5.21E-08	[2.62E-07,4.63E-07]	1357
C1+C2	2.44E-08	2.43E-08	2.59E-09	[1.94E-08, 2.95E-08]	1023
C3	9.38E-08	9.33E-08	9.71E-09	[7.50E-08,1.12E-07]	863
tRNA	3.38E-08	3.33E-08	6.18E-09	[2.27E-08,4.64E-08]	2350
rRNA	2.22E-08	2.20E-08	3.06E-09	[1.63E-08,2.82E-08]	1170
rest of molecule	2.63E-07	2.54E-07	6.44E-08	[1.5E-07,3.92E-07]	2941

7.4.5. Haplotype survey through time

The European aurochs samples Rhi1 and Rhi2 are radiocarbon dated to the Pre-LGM and belong to the now-extinct haplotype G. Six additional aurochs (VEM212, Rhi4, Rhi5, Rhi6, Rhi7 and Rhi8) are likely also Pre-LGM based on contextual dating. These samples also belong to the extinct haplotypes of G and N, while VEM212 may represent a haplotype which may outgroup all known *Bos* variation. A ~41,000 year old sample (Tula1) is a basal C haplotype. In the Last Glacial, several new haplotypes are first seen in the data across Eurasia. In Europe, four occurrences of P appear in Italian and German samples. The E haplotype is also tentatively molecularly dated in two samples from eastern Ukraine (UKR08 and UKR09). The newly described K haplotype is observed in the Caucasus (Gyu1) while C is observed in two Siberian samples (Baikal1 and Yen1).

The Holocene is the most sampled time series in this dataset. P predominates in Europe with 32 occurrences ranging in age from the beginning of the Holocene (~11,700 yrBP) to the extinction of the aurochs (~600 yrBP). Additionally, a Chalcolithic Polish aurochs (Kalisz3) and a Neolithic British aurochs (YoA) harbour the domestic Q and T3 haplotypes, respectively. In southwest Asia, a basal Q haplotype is observed in a ~7,000 Caucasian aurochs (Gyu2). In Anatolian, a Neolithic animal designated as wild (Ch22) harbours a domestic T3 haplotype. The C haplotype persists in two published Far Eastern sites (Cai et al. 2018; H. Zhang et al. 2013). In Central Asia, four maternal lineages are identified in 10 animals. Additionally, two Holocene aurochs of unknown geographic origin (Tango1 and Tango2) also have the T3 haplotype.

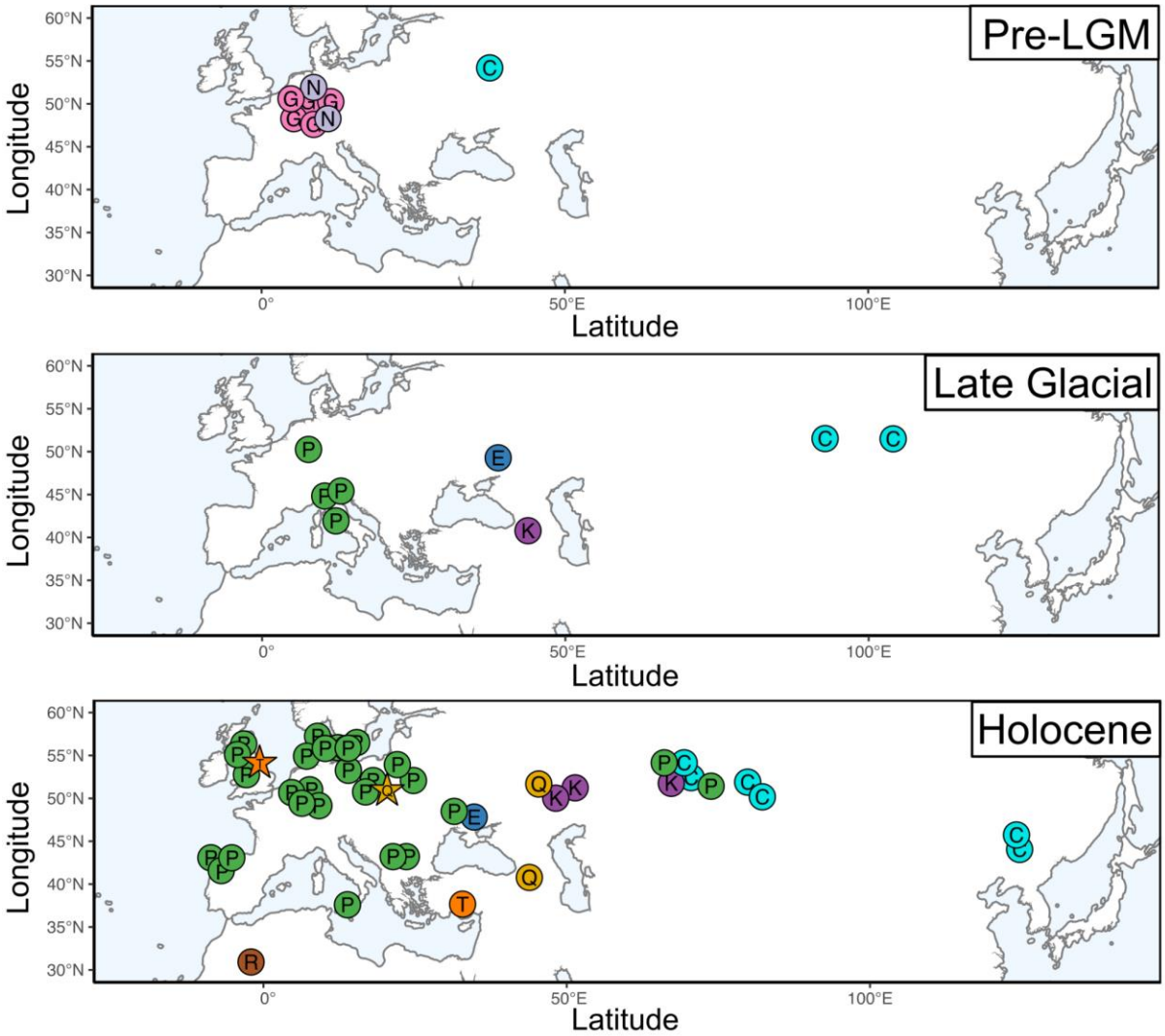


Figure 7.4 - Postglacial haplotypic turnover of mitochondrial lineages. Timing of eras are based on (Leonardi et al. 2022): Pre-LGM occurs prior to 37,000 yrBP, LGM occurs between 27,000 and

18,000 yrBP, Late Glacial occurs between 18,000 and 11,7000 yrBP, Holocene occurs after 11,700 yrBP. LGM is omitted as there is no sampling of this time period.

7.4.6. Tracking European P haplotypes

The six deepest lineages of the P haplotype were identified to have diverged towards the end of the LGM based on the 95% HPD interval (P1-P6). The divergences had high posterior probabilities (>0.99) except for P2 vs [P3-P4-P5] and P5 vs [P3-P4] which were 0.34 and 0.7, respectively. Within these newly defined sub-haplotypes, P1 predominates. Holocene recolonisation of Britain and Scandinavia comes exclusively from this, while much of Central and Eastern Europe also consists of P1. Modern East Asian breeds display at low frequency a subset of this haplotype, labelled here as P1a. Italy displays the highest amount of diversity with three haplotypes detected (P2, P5, P6). In central Asia two ancient maternal lineages of P are found (P1 and P4). An Iron Age Ubezekistani hybrid domesticate is a lineage of P1 separate to P1a. This haplotype (Kok1) forms a clade with P mitochondria found in modern Austrian cattle.

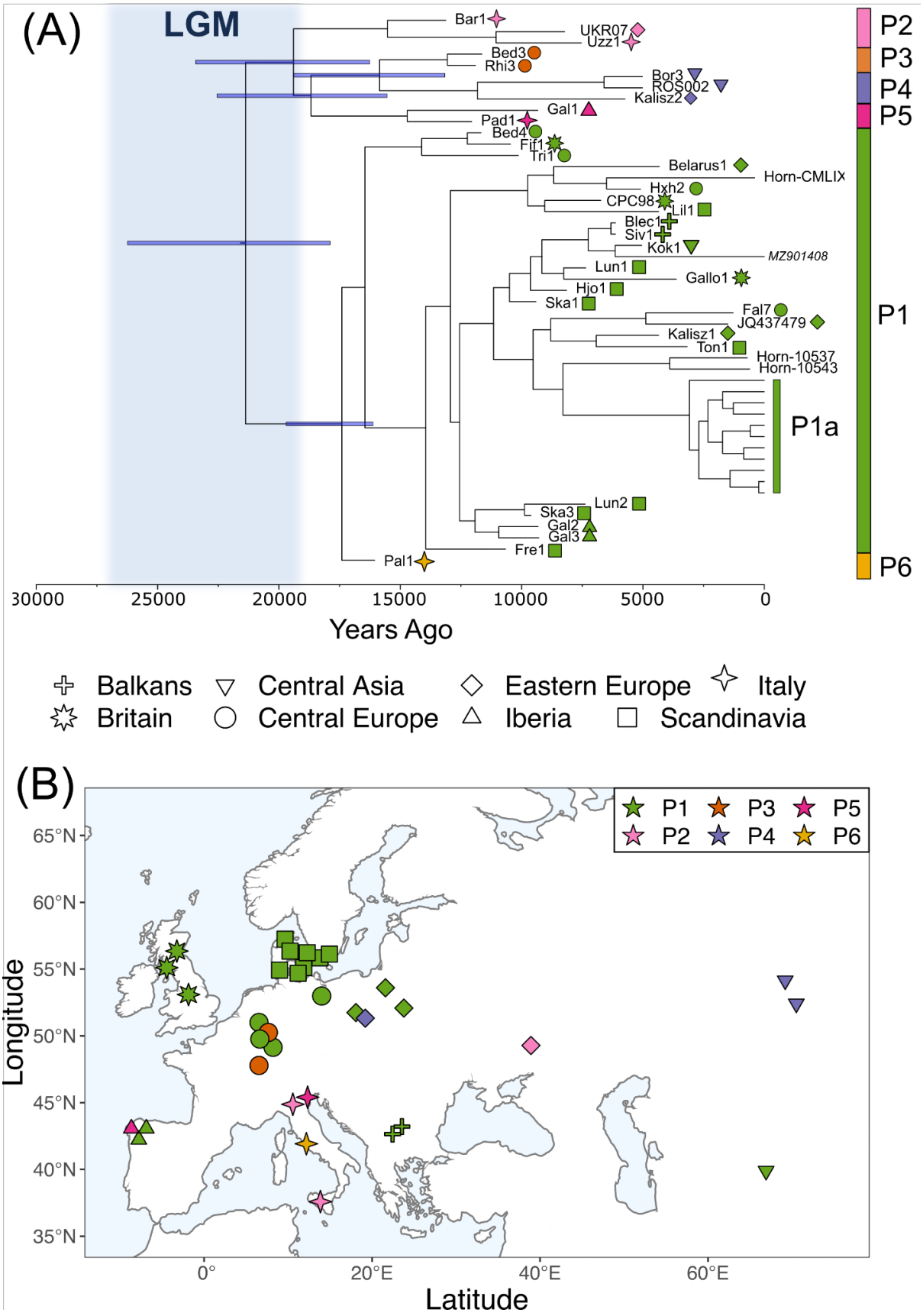


Figure 7.5 - P sub-haplotypes. (A) Defining P sub-haplotypes based on lineages that may have arisen during the LGM. (B) Survey of P sub-haplotypes across Eurasia.

The BEAST runs of dataset04 converged after 250 million chains and a Bayesian skyline plot was constructed. The plot depicts a median effective maternal population size of 10,444 around 18,000 years ago. It slowly recovered, reaching a median peak of ~48,000 around 5,000 years ago.

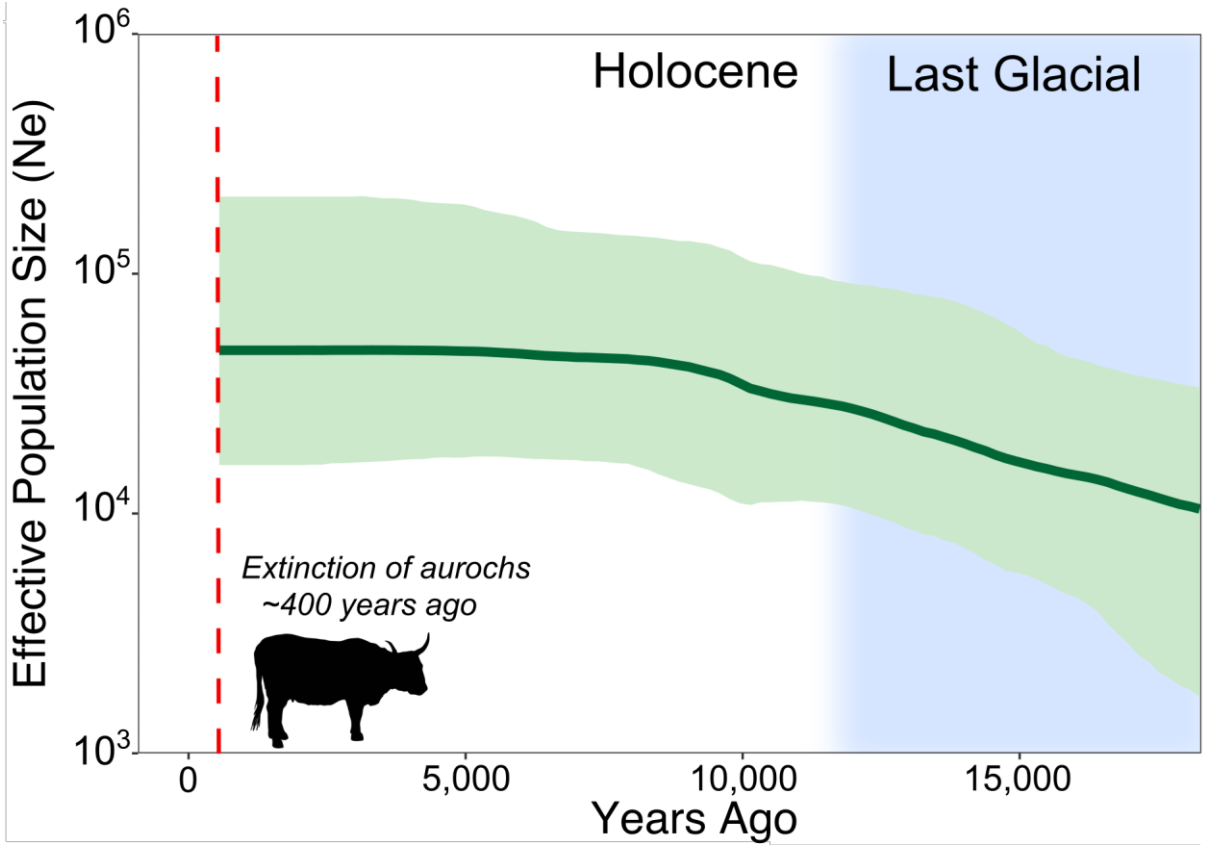


Figure 7.6 - Bayesian Skyline Plot of European P haplotypes.

7.4.7. Domestic expansions

Five star-like phylogenies are observed during the Holocene which may be associated with the domestication process. All sub-haplogroups have a posterior probability of 1 except for T6/T7, T3 and T1 (0.1211, 0.7043 and 0.3733, respectively). These are depicted in Figure 7.7, with the relevant time of possibly related processes highlighted for each.

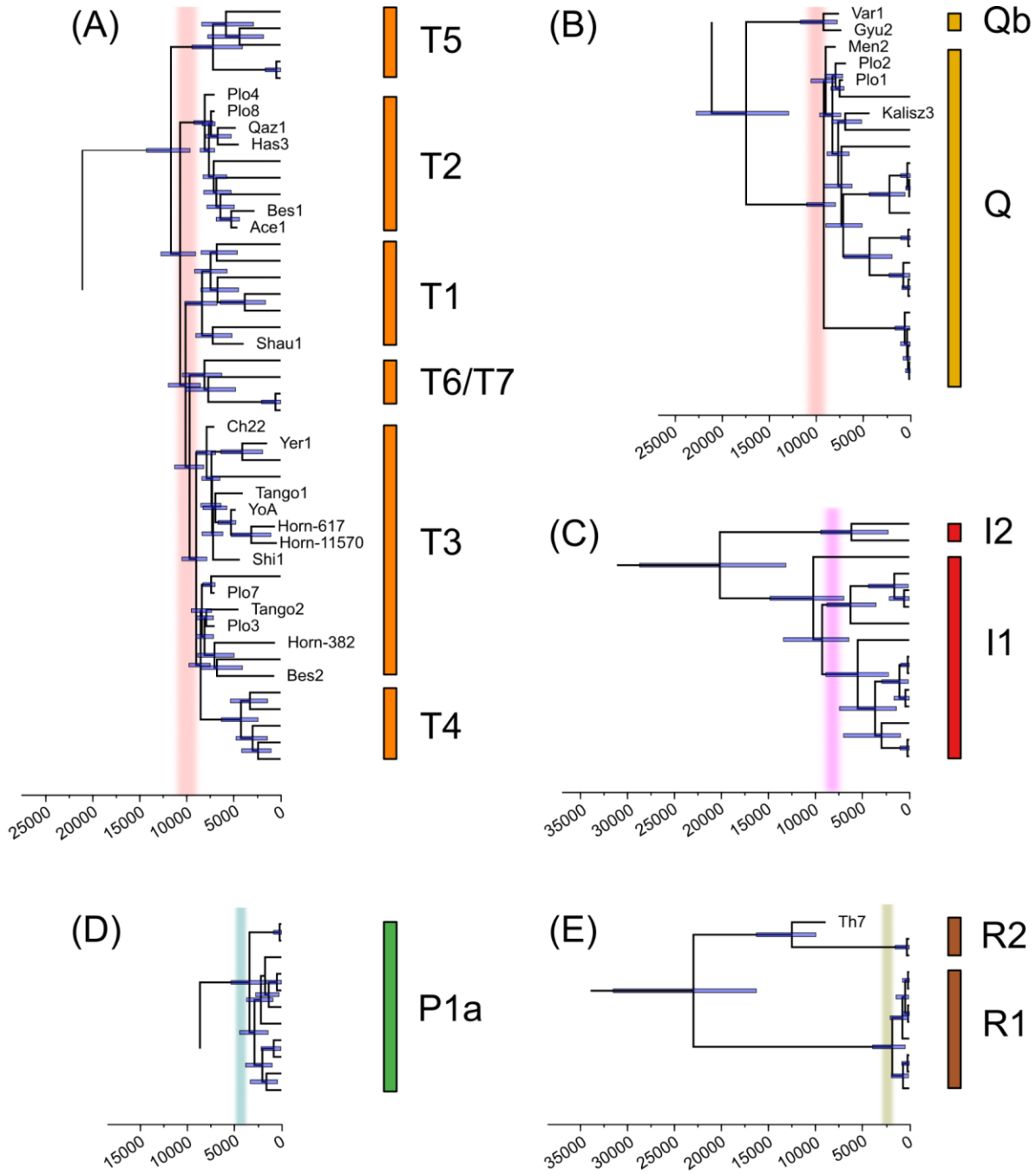


Figure 7.7 - Star-like phylogenies associated with domestic expansions. (A) The T mega-haplotype. The timing of cattle domestication in the Upper and Middle Euphrates Valley approximately 10,500 years ago is highlighted. **(B)** The Q haplotype (wild and domestic shown). **(C)** The I haplotype. Timing of cattle domestication in the Indus Valley is indicated, 8,500 years ago. **(D)** The P1a haplotype. Here the dispersal of domestic cattle from Central Asia into east Asia from 4,000 to 5,000 years ago is highlighted. **(E)** R haplotypes (wild and domestic both shown) with the age of Ancient Roman influence approximately 2,500 to 1,500 years ago indicated.

Table 7.7 - TMRCA of each domestic haplotype and sub-haplotype.

Haplotype	Median node age (yrBP)	HPD 95% node age (yrBP)
Q	9195	7,944-11,004
T	11682	9,649-14,297
T1	8370	6,773-10,199
T2	8089	7,252-9,234
T3	9692	8,193-11,285
T4	4224	2,409-6,307
T5	7204	4,069-9,443
I	20163	13,141-28,717
I1	10222	6,963-14,833
I2	6162	2,259-9,435
R	22976	16,284-31,513
R1	1818	443-3,902
R2	255	0-1,498
P	22008	18,523-25,164
P1a	3389	1,766-5,348

7.5. Discussion

7.5.1. Haplotypes G and N: relics of ancient variation

Seven Upper Rhine Valley aurochs had previously undescribed divergent haplotypes denoted here as G and N. These haplotypes are observed in the oldest radiocarbon dated animals of this database, including Rhi2, which had a calibrated date of 46,160 years old, and Rhi1, which was beyond the upper limit of radiocarbon dating (47,000 yrBP+). As these haplotypes outgroup all known taurine and indicine variation, this may suggest that these Rhine samples are an outgroup sub-species of *Bos primigenius*. This is supported by divergence timing of the G haplotype (median node age = 257,280 yrBP, HPD 95% = 199,949-317,818 yrBP) which precedes TMRCA of *B. p. primigenius* and *B. p. namadicus* based on mitochondrial and nuclear sequence divergence dating (Cubric-Curik et al. 2021; Chen et al. 2018). However, Y chromosome phylogenies and autosomal variation of the Rhine samples firmly place them within the *B. p. primigenius* clade and not as an outgroup species (discussed further in Chapter 8 and 9). This non-monophyly emphasises the importance of not extrapolating mitochondrial trees to be true phylogenies of species (Pamilo and Nei 1988). Two scenarios are explored to reconcile this observation; incomplete lineage sorting or introgression from a ghost outgroup population.

Under an ILS framework, mitochondrial variation within ancient *Bos primigenius primigenius* retained an ancestral mitochondrial lineage that preceded the phylogenetic split of *B. p. primigenius* and *B. p. namadicus* (Figure 7.8A). The retention of ancestral mitochondrial diversity (*i.e.* haplotypes G and N) persisted in pre-LGM taurine animals of Rhine Valley. This ancient variation may have also been retained in ancient herds of *B. p. namadicus* but was also lost during or prior to the establishment of modern zebu cattle maternal lineages. This would explain why the Rhine aurochs have a taurine Y and nuclear genome but an outgroup mitochondrial haplotype. While stochastic lineage sorting of mitochondria is thought to occur more rapidly due to the lower *Ne* and maternal inheritance of the allele, instances of mitochondrial ILS are observed in some species (Funk and Omland 2003; McKay and Zink 2010). For example, ILS is the favoured explanation for the mitochondrial non-monophy observed in the sister Genus *Bison* (Wang et al. 2018).

Alternatively, the presence of this divergent mitochondrial haplotype may be the result of a prior introgressive event from an outgroup ghost population. A suitable candidate for such an outgroup ghost population is found in the ~200 thousand year old British aurochs, VEM212. VEM212 also has a divergent mitochondrial haplotype, which is basal to the N haplotype in ML and unpartitioned Bayesian phylogenetic trees. Although based on very low coverage, nuclear analysis of VEM212 similarly places this animal as an outgroup to both taurine and indicine autosomal variation (Chapter 9). It is plausible that a lineage related to VEM212 may have persisted in Europe and introgressed N and G haplotypes into an ancestor lineage of the Rhine samples. Introgressive hybridisation is a relatively common phenomenon during species range expansions and often proceeds asymmetrically from the local population into the colonising population (Currat et al. 2008). During warmer periods of the Pleistocene the ranges of some warm-adapted species expanded into hybrid zones with cold-adapted sister species, facilitating introgressive hybridisation. For example, extant brown bears harbour polar bear mitochondria from a hybridisation event when brown bears expanded into polar bear ranges during warmer periods of the Pleistocene (Cahill et al. 2015). Similarly, colonising brown hares have been the recipient of the indigenous mountain hare mitochondria in Iberia and Scandinavia when climate amelioration facilitated their expansion (Fredsted, Wincentz, and Villesen 2006; Alves et al. 2008). In this scenario, the ancestor lineage of the Rhine samples expanded into Europe and hybridised with the local VEM212-related population.

While an introgressive hybridisation scenario is favoured by the current evidence, it does not resolve why the introgression was seemingly female-biassed. Only the female uniparental

markers of the Rhine samples are introgressed, while male uniparental markers are taurine (*i.e.* clustering to the exclusion of *Bos indicus)*. Hybridisation could have occurred between a population of *Bos primigenius* and a population related to VEM212. Subsequent population dynamics, *e.g.* back-crossing of the hybrids with migrating male *Bos primigenius* may have ultimately resulted in a hybrid Rhine population that derives most of its ancestry from *Bos primigenius* but retains the outgroup mitochondrial haplotype. A similar framework of hybridisation is proposed in brown bears and polar bears during the Pleistocene (Cahill et al. 2015). The brown bear populations encroaching on polar bear ranges were mostly or exclusively male. This sex-biassed gene flow resulted in an extant population that derived most of its ancestry from brown bears but retained polar bear mitochondria and elevated X chromosome input from polar bears. Like bears, it is likely that aurochs cows were more philopatric than their male counterparts and so were less likely to colonise new niches (Greenwood 1980). Female philopatry is also reflected in the social structures of other *Bovini* species, whereby solitary males tend to leave the herd to seek mates (van Vuure 2005).

A dramatic turnover of mitochondrial haplotypes is observed in post-LGM Europe, as the formerly predominant N and G haplotypes are replaced with P haplotypes. The P haplotype diverged from QT prior to the LGM (median node age = 45,183 yrBP, HPD 95% = 35,415 - 55,345 yrBP) and possibly coexisted with N and G in Europe. A basal P lineage may have been a haplotype that was dispersed more southerly than the G and N haplotypes. The earliest appearance of the P haplotype in this dataset is a Late Glacial Italian aurochs, Pal1, that existed ~16,000 years ago - European aurochs likely survived the LGM in southern refugia (van Vuure 2005). Conversely, the G and N haplotypes may have had a more northern distribution and succumbed to the ice and cold of the LGM.

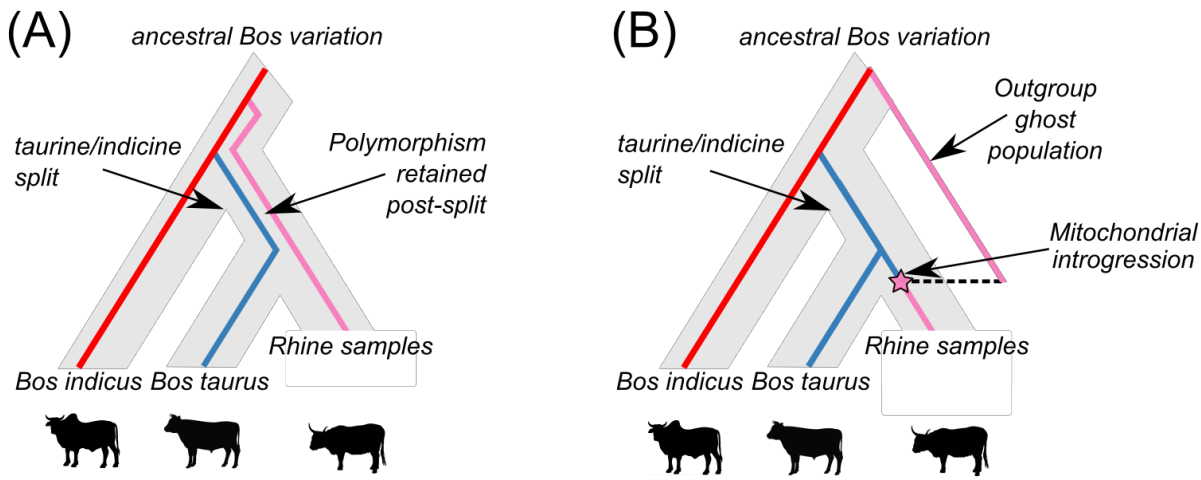


Figure 7.8 - The proposed origin frameworks of the G and N haplotypes. (A) An incomplete lineage sorting framework whereby ancestral mitochondrial Bos variation is retained in the taurine lineage

resulting in non-monophyly. **(B)** An outgroup introgression framework whereby G and N haplotypes were introgressed into *Bos primigenius* during a hybrid event with an outgroup ghost lineage.

7.5.2. *Refugia* of *Bos primigenius* mitochondria

Along the taurine branch, five haplotypic divergences occur in rapid succession resulting in a star-like phylogeny. Such star-like phylogenies can occur due to rapid demographic expansion (Cox et al. 2019; Dai et al. 2011). According to median node ages, these five lineages arose within as few as ~6,000 years, 76,607 to 70,675 years ago (HPD 95% = 89,840 - 58,967 yrBP). These successive divergences culminated in five haplogroups that persisted into the Pleistocene/Holocene transition (E, C, R, K, and PQT). Interestingly the estimated start of this radiation corresponds to marine isotopic stage MIS5a (~82-71 Ka). While this is a period that marks the onset of the Last Glacial Period, it is also characterised by large oscillations in global temperature, sea levels and ice sheet volume (Lambeck, Esat, and Potter 2002; Batchelor et al. 2019). The periodic climate amelioration of MIS5 has been suggested as the cause of range expansion of other warm-adapted mammals in the Northern hemisphere, such as brown bears (Wang et al. 2022). MIS5a, the beginning of this radiation, was a relatively warm period when large ice sheets receded across Northern Eurasia, resulting in increased sea levels. This brief interstadial may have provided an opportunity for an ancestral *Bos primigenius* population to rapidly expand into new niches that were previously inaccessible due to the ice and cold of the prior isotopic stage MIS5b. This time period may also correspond to the expansion of *Bos primigenius* males into Europe, as discussed in the previous section. Such a model of rapid postglacial expansion is observed in Holocene aurochs expanding into postglacial northern Europe and Central Asia (Edwards et al. 2007; Hideyuki Mannen et al. 2020).

The relatively warm period of MIS5a was short-lived and the succeeding cold period MIS4 (~71-57 Ka) was accompanied by a large ice-sheet expansion across northern Eurasia (Lambeck, Esat, and Potter 2002; Batchelor et al. 2019). Through this stadial, much of Northern Eurasia was inaccessible to the warm-adapted aurochs. The MIS4 cold period may have therefore fractured the *B. primigenius* population distribution that had expanded during the warmer MIS5a period. The fragmented aurochs populations retreated to refugia, resulting in genetic isolation between these populations. Such habitat contraction and fragmentation has similarly been modelled in European Neanderthals due to the climatic decline associated with the MIS5a/MIS4 transition (Banks et al. 2021). The putative location of fragmented refugia may be inferred by the persistence of haplotypes into the Late Pleistocene/early Holocene. An eastern refugium of the C haplotype is inferred by its Siberian and Chinese dispersal in the Late Pleistocene/early Holocene (Cai et al. 2018; Zhang et al. 2013). The newly described K

haplotype may be a representative of a Caucasian refugium based on the recovery of a Late Glacial K aurochs in Armenia (Gyu1). The existence of a ~9,000 year old R aurochs in Morocco suggests that this lineage existed in North Africa during the Pleistocene (Verdugo et al. 2019). The refugium of E is less secure due to the lack of exact dates on the two samples but it may have been a lineage that survived in a refugium in eastern Europe (Edwards et al. 2007). The PQT lineage was widely dispersed across Europe and southwest Asia by the Late Pleistocene/early Holocene. During the LGM the P haplotype likely retreated to southern peninsulas of Europe while T/Q were dispersed across southwest Asia. Such genetic isolation of refugia is supported by autosomal and Y chromosomal variation of Upper Pleistocene aurochs, which displays strong genetic structure along an east-west axis (Chapter 8 & 9).

7.5.3. postglacial expansion

The topology of the P mitochondrial tree also has some suggestion of a star-like phylogeny. TMRCA of the sampled P lineages arose during the LGM (median node age = 21,699 yrBP, HPD 95% = 18,253 - 25,164 yrBP) which later diverged into up to 6 haplotypes (P1-P6) during the LGM. The Late Glacial proliferation of these lineages is consistent with a previously suggested postglacial population expansion of P aurochs in Europe (Ceiridwen J. Edwards et al. 2007). The expansion of P aurochs lineages is also reflected in the Bayesian Skyline Plot, which depicts a steady increase in effective maternal population size from ~18,000 yrBP. During the LGM, populations of European aurochs retreated to three putative refugia in France-Iberia, Italy and the Balkans (van Vuure 2005). These southern peninsulas have been well-characterised as refugia for a number of warm-adapted species during the Last Glacial Period, before expanding into postglacial northern territories as the climate warmed (Hewitt 1999). Recolonisation routes for multiple species have been reconstructed based on archaeological remains and extant genetic variation. Broadly, Iberia and the Balkans are the likely refugia from most terrestrial mammals recolonising Europe, while the Alps served as an effective barrier to gene flow leaving Italy (Taberlet et al. 1998). The Alps have previously been suggested as an effective barrier to gene flow for aurochs (Edwards et al. 2007). However, aurochs maternal lineage distributions do not resolve the recolonisation route from southern refugia as there is no evidence in the current dataset of which P haplotypes, if any, Iberia or the Balkans harboured during the Late Glacial.

While refugial recolonisation routes were not discernible with the current P sub-haplotypes for European aurochs, it is clear that there was refugial connectivity since the LGM. P1 existed in both Holocene Iberia and the Balkans, while P2 existed in Pleistocene Italy and Holocene Iberia. TMRCA of the P2 haplotypes of Iberian Gal1 and Italian Pal1 existed during the Late

Glacial (HPD 95% = 12,759-17,174 yrBP) which provides evidence of gene flow between these refugia post-LGM. Explicitly, this implies that the Alps were not an impassable barrier to gene flow for aurochs after the LGM, contrary to previous suggestions (Mona et al. 2010). Such refugial connectivity is congruent with recent paleoecological modelling of suitable habitats of aurochs over the last 47,000 years (Leonardi et al. 2022). Notably, Italy displays the highest amount of haplotypic diversity, with three haplotypes detected (P2, P5, P6). In addition, one of these haplotypes (P6) has thus far only been detected in the peninsula. As postglacial expansion into new niches often results in the loss of genetic diversity (Hewitt 2000), it is possible that Italy was the centre of diversity of P through the LGM. While it remains unclear which refugia the sub-haplotype arose in, Holocene recolonisation of Britain and Scandinavia is exclusively P1 and much of Central and Eastern European sampling also consists of P1. P1 was not detected in four Italian animals but may present in further sampling.

While the P haplotype represents the major extinct European aurochs haplotype, it is also unexpectedly harboured in east Asian cattle breeds (Mannen et al. 2020). This has been interpreted to be due to an eastward expansion of European aurochs across Central Asia in the post-LGM. During the Last Glacial Period, Central Asia was likely too cold to support long term aurochs habitation (van Vuure 2005). However, as the climate warmed post-LGM, Central Asia opened up as a new ecological niche for expanding populations of aurochs to claim. This is also observed in Moose populations that migrated to Central Asia westward in the post-LGM (Meiri et al. 2020). Sampling of Holocene Central Asian aurochs revealed a melting pot of maternal ancestry from European, southwest Asian and east Asian refugia. In addition to two P aurochs sampled here (Bor3, ROS002), three K (Bor1, Ura1, Var2), four C (Bor2, Ust1, ROS001) and one basal Q haplotype (Var1) were encountered. It is suggested that these haplotypes migrated to Central Asia from Europe (P), east Asia (C), and southwest Asia (Qb, K). Y chromosome phylogenies and autosomal variation of Central Asian aurochs also reflect this diversity of maternal ancestries and support hybridization among migrating populations (Chapter 8 and 9).

7.5.4. Origins of aurochs domestication

Star-like topologies have been inferred as the signatures of rapid population expansions in a wide array of contexts including pathogen epidemic spread (Li et al. 2020), human migration out of Africa (Forster 2004), domestication events (Troy et al. 2001) and expansion of wild populations into newly opened ecological niches (Cox et al. 2019; Dai et al. 2011). Several star-like topologies were observed in the mitochondrial phylogeny of aurochs which may indicate proliferation of animals after a domestication event of a small number of breeding

females. Indeed, the starting group of aurochs that were domesticated may have been quite small, with estimates as low as around 80 breeding females for taurine cattle (Bollongino et al. 2012; Scheu et al. 2015). These timings and topologies of these domestication events will now be discussed.

While the T mega-haplogroup dominates taurine cattle, the exact process of domestication of the T sub-haplotypes is equivocal. It has also been suggested that domestication of a basal-T was the original maternal stock for domestication around 10,500 years ago, which diversified into sub-haplotypes thereafter (Achilli et al. 2008, 2009; Bonfiglio et al. 2012). An opposing view is that secondary taurine domestication events may have occurred in North Africa and east Asia based mainly on the geographic structuring of T1 and T4, respectively (Bradley et al. 1996; Bradley et al. 1998; Mannen et al. 2004). A framework of an original southwest Asian domestication event of basal-T cows is supported by the node age of TMRCA of the T mega-haplogroup (median node age = 11,682 yrBP, HPD 95% = 9,649-14,297 yrBP), which overlaps with *Bos primigenius* domestication in southwest Asia. The basal-T haplotype diversified into sub-haplotypes, the timing of which may also be ascribed to bottleneck events through migrations. For example, the expansion time of T1 corresponds to the dispersal of cattle into Africa from the Near East, which occurred around 8,000 years ago (Hanotte et al. 2000), (median node age = 8,370 yrBP, HPD 95% = 6,773-10,199 yrBP). The preponderance of T1 in ancient Levantine samples further favours a framework that T1 arose during the dispersal of the Neolithic package into North Africa (Verdugo et al. 2019). Similarly, T4 expands when cattle were introduced to east Asia, roughly 4,000-5,000 years ago (median node age = 4,224 yrBP, HPD 95% = 2,409-6,307 yrBP). This is further evidenced by T4 appearing in ~4,000 year old cattle from China (Xia et al. 2021).

Overall, divergence dating of the T mega-haplogroup supports a framework of a basal-T domestication in southwest Asia and subsequent dispersal of these sub-haplotypes into Europe, Africa and Asia by Neolithic communities. However, this conclusion directly contradicts a study from 2006 that detected a T3 allele in a ~17,000 year old southern Italian aurochs (Beja-Pereira et al. 2006) (a D loop sequence identical to the most common modern haplotype). Further studies from the same research group found that T was harboured in multiple other Pleistocene Italian aurochs (Lari et al. 2011; Mona et al. 2010). In all, of 15 ancient bones sampled in these studies, twelve aurochs were designated T3, one was designated T2 while two were designated P. However, the existence of these T sub-haplotypes predates TMRCA of the T mega-haplotype by several thousand years and contravenes previous evidence of T origins in southwest Asia (Troy et al. 2001). Significantly, these sample

haplotypes were determined using amplification of sequence-specific PCR primers which are very vulnerable to modern contamination in aDNA studies (Rizzi et al. 2012). T3 is the dominant allele and T2 is also sporadically observed in modern day Italian cattle (Cubric-Curik et al. 2021), and so modern contamination may have been the source of spurious T haplotyping. This also occurred in a separate study where a supposed T3 Pleistocene aurochs from Czechia was detected using PCR amplification which the authors concede is likely modern contamination (Kyselý and Hájek 2012). Since the transition to NGS technologies, aDNA is more easily authenticated by short DNA fragment lengths and deamination patterns (Jónsson et al. 2013) and T has not been detected in Pleistocene Europe since. For these reasons, there is currently no convincing evidence of T aurochs in Pleistocene Europe.

The domestic Q haplotypes are found sporadically in modern breeds of taurine cattle (Achilli et al. 2008; Olivieri et al. 2015). First discovered in modern beef breeds of southern Italy, it was suggested that this haplotype may have originated from a local European aurochs introgression. However, the domestic Q is also found in Men2, an Anatolian domestic animal that precedes the Neolithisation of Europe. Similarly, early Neolithic southeast European animals (including Plo3, Plo4 in this dataset) also harbour this haplotype (Scheu et al. 2015). The presence of a basal Q haplotypes in southwest Asian aurochs (Gyu2) further supports that Q was not recruited from the wild in Italy but instead was domesticated in southwest Asia. The divergence time of T/Q was prior to the domestication of taurine cattle, perhaps during the LGM (median node age = 21,129 yrBP, HPD 95% = 15,798 - 26,930 yrBP) confirming that they were not domesticated from a common maternal ancestor. The timing of TMRCA of Q supports a recruitment timing very early in cattle domestication (median node age = 9,195 yrBP, HPD 95% = 7,944 - 11,004 yrBP), perhaps contemporaneously to the T mega-haplotype domestication. It is striking that there is no firm evidence that more than two maternal lineages were captured and bred in the origins of *Bos taurus,* perhaps speaking to the difficulty of domesticating this large, dangerous mammal.

The separate domestication of indicine cattle occurred later than taurine cattle in the Indus Valley, around 8,000 years ago (Chen et al. 2010). In a study of 844 zebu mitochondrial sequences two haplogroups emerged in extant breeds (I1 and I2). These haplotypes diverged prior to the domestication of indicine cattle (median node age = 20,163 yrBP, HPD 95% = 13,141=28,717 yrBP). Genetic diversity of I1 peaks within the Indus Valley, which is proposed for the origins of domesticated indicine cattle. This is supported by the dating of TMRCA of I1 which overlaps with this time (median node age = 10,222 yrBP, HPD 95% = 6,963-14,833 yrBP). The geographic origin of I2 was more ambiguous, although its diversity also peaks in

the northern part of the Indian subcontinent (Chen et al. 2010). Based on TMRCA of I2, a framework of contemporaneous recruitment during the domestication or of secondary recruitment is supported (median node age = 6,162 yrBP, HPD 95% = 2,259-9,435 yrBP).

7.5.5. Maternal gene flow between wild and domesticated cattle

Star-like topologies are observed in aurochs-like haplotypes, but these expansions are likely the result of incorporation of wild maternal lineages into domestic stock. As discussed, the European haplotype P is found in several east Asian cattle breeds (Hideyuki Mannen et al. 2020). In ancient Central Asia maternal lineages of P are found in two ~5,000 year old wild Central Asian samples - Roschinskoe and Borly, Kazakhstan (ROS002 and Bor3) - providing explicit evidence of postglacial migration of European aurochs across Eurasia. However, based on the defined P sub-haplotypes, these ancient samples are not maternal ancestors to the modern domestic P haplotype (P1a). TMRCA of ROS002 and Bor3 (P4) and modern domestic P haplotype (P1a) was during the LGM (median node age = 22,008 yrBP, 95% HPD = 18,523-25,164 yrBP). Therefore, the direct ancestor of the P1a lineage has not yet been sampled in Asia. Interestingly, the expansion of the P1a haplotype occurred approximately 3,500 yrBP (median node age = 3,389 yrBP, HPD 95% = 1,766-5,348 yrBP) correspondings to a time after taurine cattle are believed to have spread from Central Asia to East Asia between 5,000 and 4,000 years ago (Flad et al. 2007). In addition, P was introgressed into an Iron Age Ubezekistani hybrid domesticate (Kok1) and a modern Austrian animal. These two samples form a clade together, with TMRCA existing post-domestication (median node age = 6,903 yrBP, 4,519-8,719 yrBP). It is therefore possible that they are ancestors from the same maternal introgression event.

The domestic R haplotypes are found exclusively (but rarely) in modern southern Italian breeds. This has led to suggestions that the R haplotype is representative of an independent Italian domestication event or wild Italian introgression (Bonfiglio et al. 2010). The only ancient R sample is Th7, an ~9,000 year old aurochs from Morocco. TMRCA between the modern R2 and Th7 existed during the Last Glacial or early Holocene (median node age = 12,500 yrBP, HPD 95% = 9,952-16,228 yrBP). Such a fragmented distribution of R2 aurochs is difficult to reconcile considering the absence of a land bridge between the Mediterranean and North Africa in the Late Pleistocene. A more parsimonious explanation is that the R haplotype was imported to Italy from North Africa by humans and introgressed into the local domesticated herds there. Notably, the expansion of R1 overlaps with the age of the Roman empire (median node age = 1,818 yrBP, HPD 95% = 443-3,902 yrBP). Domestic cattle were an important commodity during Roman civilisation, with cattle dealers having a defined occupation in Roman

society (*pecuarii*). Strontium isotope analysis of Roman cattle teeth demonstrated that animals were transported through Roman road and maritime networks, which extended across much of Europe, North Africa and the Near East (Minniti et al. 2014). Furthermore, exotic animals such as lions, tigers, and camels are known to have been imported from North Africa to the Mediterranean basin during the Roman ages (Azaza and Colominas 2020). Therefore, it is possible that R cows were transported into Roman agricultural communities as an exotic improvement to local landraces. Sampling of ancient Roman cattle may provide insight into the origin of the R haplotype.

In addition to recruitment of wild haplotypes into domestic stock, there are several instances of clear domestic introgression into the wild. The earliest instance is found in a ~4,300 year old Polish Eneolithic aurochs harbouring a domestic Q haplotype (Kalisz3). Similarly, a ~4,200 year old Neolithic aurochs from England was genotyped to a T3 allele (YoA). Notably, interbreeding of ancient domesticates using local aurochs bulls is inferred in Neolithic Poland and England based on Y chromosomal phylogenies (Chapter 8). This may have been accidental or perhaps due to an intentional strategy of early farmers (Park et al. 2015; Orlando 2015). By allowing domestic and wild animals to hybridise, some female hybrids may have escaped into the wild and back-crossed with aurochs, resulting in the retention of female domestic uniparental markers. Strikingly, in England, all genotyped Neolithic domestic cattle harbour a Y chromosome that is characteristic of European aurochs. The existence of a domestic female haplotype in the wild further emphasises the extent of hybridisation between wild and domestic cattle in ancient farming communities.

8. Y chromosomal analysis of *Bos primigenius*

8.1. Introduction

Arising approximately 180 million years ago from a proto-sex chromosome, the Y chromosome underlies sex determination in all placental and marsupial lineages (Cortez et al. 2014). Millions of years of sequence erosion and stepwise recombination arrest have left only a small segment of the Y chromosome that maintains recombination with the X chromosome. These homologous sequences are collectively known as the pseudoautosomal region (PAR) and are generally located in the terminal portions of the Y chromosome. The majority of Y chromosomal sequences are located in the male-specific region (MSR) which does not recombine with the X chromosome during meiosis. As the MSR is paternally inherited, it makes the Y chromosome an attractive genetic marker for reconstructing male migrationary patterns and sex-specific gene flow (Jobling and Tyler-Smith 2003). However, technical limitations have dictated that the Y chromosome is much less studied compared to its maternally-inherited counterpart, the mitochondrial genome. The small, circularised mitochondrial genome has a high copy number and rapid mutation rate, making it a candidate for phylogenetic studies in recently diverged populations (Ballard and Whitlock 2004). In contrast, the Y chromosome is a large, highly repetitive chromosome present in only half the population with relatively low genetic variability (Greminger et al. 2010). These technical aspects of Y chromosomal studies have limited its use to select few model organisms.

Domestication of taurine cattle occurred around 10,500 years ago from what was likely a small group of cows in Upper Euphrates Valley (Bollongino et al. 2012). The maternal lineage of the first domesticated cows in Anatolia were likely T and Q based on modern dispersal of mitochondrial haplotypes and molecular dating of the expansion of the macro-T haplogroup (Chapter 7). In contrast, very little is known about the origin of the bovine Y chromosome. Genetic surveys of modern cattle revealed a deep divergence within taurine breeds (between haplotypes Y1 and Y2) and indicine breeds (Y3) (Götherström et al. 2005; Ginja, Telo da Gama, and Penedo 2009; Edwards et al. 2011). In addition, geographic structure was found within modern taurine populations with the Y1 haplogroup found mostly in northwestern European breeds while the Y2 haplotype was found predominantly in southern European and Anatolian breeds. An initial hypothesis to explain the geographic structure posited that Y2 was the original Anatolian haplotype domesticated in the Fertile Crescent while Y1 was a wild European haplotype that was introgressed into domestic herds of northern European cattle ((Götherström et al. 2005). Later studies using aDNA refuted a wild European origin for Y1, as

the Y2 haplotype was genotyped in several Neolithic European aurochs carrying P mitochondria (Bollongino et al. 2008). This suggested either widespread male-domesticate introgression occurred into Neolithic aurochs, or more parsimoniously, that Y1 does not define the European aurochs Y chromosome. More recently, (Chen et al. 2018) described Y chromosomal diversity in a worldwide survey of modern cattle breeds using NGS data. Increased resolution in the Y2 clade revealed two geographic sub-haplogroups: a Y2b clade manifested in east Asian taurines, while a Y2a clade was found mostly in southern European taurines. Y3 was also partitioned into Y3a found in Chinese indicine breeds and Y3b in South Asian indicine breeds.

While dating the expansion of these modern haplotypes using Bayesian analysis may help elucidate domestication routes and possibly reveal wild aurochs introgression, strong genetic drift in Y chromosomes and sex-specific behaviours can confound genetic inferences when using only modern variation (Jobling and Tyler-Smith 2017). Human control of cattle breeding patterns has intensified since domestication through widely-used modern artificial insemination practices (Foote 2002), so incorporation of ancient variation is crucial to understand the complex dispersal of bovine Y chromosomes. Phylogenetic tree construction and molecular divergence dating using ancient whole mitochondrial genomes is widely applied and has led to well-established conclusions on population history of wild and domestic animals and hominids (Fu et al. 2013; van der Valk et al. 2021; Daly et al. 2018). Ancient Y chromosomal phylogenetic trees have also been generated, although these have been produced using only SNP sites, likely due to the larger chromosomal size (Daly et al. 2018; Librado et al. 2021; Smeds, Kojola, and Ellegren 2019; Karmin et al. 2015). Combining ancient and modern high quality Y chromosomes may help to better understand the complex dispersal of bulls through the last glacial period, domestication and beyond.

8.2. Aims

1. To characterise the relationship between different populations of *Bos primigenius* using Y chromosomal variation.
2. To construct Bayesian Y chromosomal phylogenies of *Bos primigenius* lineages to detect phylogeographic structure.
3. To time the expansions of several Y chromosomal lineages that are associated with *Bos* domestication events.

8.3. Materials and methods

8.3.1. Materials

Due to the sparsity of high-quality ancient genomes, several unpublished domesticate cattle genomes processed by Dr Victoria Mullin and Deborah Diquelou are included in this study. In total, the ancient dataset comprises 73 ancient wild and domestic males of at least 0.3X average genome coverage (Table 8.1). These samples ranged in age from the Middle Ages (<1,000 yrBP) to pre-LGM (>47,000 yrBP).

149 modern bulls are included in this study (Appendix Table 8.1). These BAMs were provided by Dr Victoria Mullin and Dr Ningbo Chen. 105 of which are taurine animals, 19 are indicine while the remainder are Asian hybrids. The majority (75) of modern animals are European breeds.

8.3.2. Region for Y chromosome SNP calling

The bovine MSY consists of three major regions: the Y-transitional, the Y-ampliconic and the X-degenerate (Chang et al. 2013). The X-degenerate region spans from 2.5-3.9 Mb and 42.2-43.3 Mb along the bovine Y assembly, GCF_000003205.7. Variants were called only in the two sub-regions of the X-degenerate region for a total callable length of 2.5 Mb.

Additionally, a mappability filter based on Heng Li's SNPable filter (http://lh3lh3.users.sourceforge.net/snpable.shtml) was applied to the X-degenerate region. This filter masks sites of the X-degenerate regions that are susceptible to alignment of short sequences found elsewhere in the genome. While this reduces the length of the sequence available for SNP-calling, it is necessary to avoid distorting results caused by incorporating false variation caused by technical artefacts (Karmin et al. 2015). This reduced the usable region to 1,443,572 total bases.

8.3.3. SNP calling pipelines

Two SNP calling pipelines were used to produce SNP sets for different phylogenetic analyses. The first SNP calling pipeline was devised to produce high quality SNP calls for molecular divergence dating based on modified methods in (Karmin et al. 2015). Three datasets were curated with ancient and modern bulls above a mean genome coverage of 8X (Appendix Table 8.2):

- All ancient and modern *Bos* genomes 8X and above (reference01)
- All ancient and modern *Bos* genomes 8X and above, excluding Y3 animals (reference02)
- All ancient and modern *Bos* genomes 8X and above, excluding Y3 and Y4 animals (reference03)

Samtools mpileup was used to call variants in bam files (-C 50 -q 30 -Q 20 -s) and these variant calls were converted to a merged VCF file using bcftools call (-mv -f GQ,GP). Indels were removed using vcftools --remove-variants to leave only SNPs in the VCF file. Only bialleic SNPs were retained using vcftools (--max-alleles 2 --min-alleles 2) and heterozygous calls were removed in samples using bcftools filter (-S . -e 'GT=="het"').

The total number of SNPs called for each dataset are as follows:

-6,479 SNPs for reference01
-5,647 SNPs for reference02
-3,096 SNPs for reference03

The filtered VCF file was converted to a tabbed VCF file using vcftools vcf-to-tab. This tabbed VCF file was then converted into a multifasta file using an inhouse Python script.

The second SNP calling pipeline was intended to produce a SNP set to investigate tree topology and haplotypic diversity of modern and ancient samples of mixed coverages based on modified methods detailed in Daly et al. 2018 and Chen et al. 2018. Variants were called on a reduced dataset of ancient and modern bulls to attempt to balance the sampling bias of mainly European taurine breeds and to avoid retention of breed-specific variation. To this end, two modern bulls of at least 8X coverage from each of the haplotypes identified in (Chen et al. 2018) were randomly chosen. Additionally, two African Y2b bulls were included. All available ancient genomes above 8X (11) were also added for a total of 28 samples (Appendix Table 8.3).

Variants were called on this reduced dataset and filtered as above. The filtered VCF file was converted to plink files using vcftools. A number of filtering steps were then performed on the plink files to remove breed specific variation. These steps were applying a genotyping filter of 0.1 (-geno 0.1), removing SNPs that were present in less than 90% of the samples (-mind 0.9), and applying a MAF filter of 5% (-MAF 0.05). This resulted in a total of 2,943 SNPs to be called

in all low coverage samples. The plink file was converted to a SNP bed file and used to call variants in all ancient and modern samples, as above. The resulting VCF file was filtered to remove indels and triallelic sites, and set all heterozygous sites to missing, as above. Transition SNPs were removed from the final filtered VCF file to produce a separate dataset that excluded the possibility of ancient damage causing false variation.

The two filtered VCF files (all SNPs and transversions only) were converted to tabbed VCF files using vcftools vcf-to-tab. These tabbed VCF files were converted into separate multifasta files (all SNPs and transversions only) using an inhouse Python script.

8.3.4. Bayesian analysis

Each high quality SNP call multifasta was converted to a Nexus file using SeaView Version 5.0.1 and loaded into BEAUTi. For ancient samples, tip dates were set to the midpoint age of the radiocarbon 95% confidence interval (Appendix Table 8.2). The site model was chosen by bModelTest while the clock model was set to Relaxed Clock Log Normal. A Coalescent Bayesian Skyline model was chosen as the tree model and a BEAST xml file was generated.

Caution is necessary in phylogenetic tree construction with only SNP sites, as this can result in overestimated branch lengths and biassed phylogenetic inferences (Lewis 2001; Leaché et al. 2015). As the loaded file only contained variant sites (*i.e.* SNPs), the number of invariant/constant sites was specified in the XML file (https://github.com/andersgs/beast2_constsites). To compare the effect of this setting on the Bayesian analysis, reference02 was also run without invariant/constant sites specified.

BEAST analyses of the three high coverage sample databases were performed using BEAST v2.6.6. For each dataset, two independent BEAST runs of 350 million chains were performed. The MCMC chain was sampled every 35,000 iterations, with 30% burn-in. Log files were assessed using Tracer to inspect parameter convergence. Successful runs were combined using LogCombiner and a maximum clade credibility tree was constructed with median node heights using TreeAnnotator. The resulting tree was visualised with FigTree.

8.3.5. Topology SNP processing

The two reduced dataset multifastas (all SNPs and transversions only) were loaded separately into SeaView Version 5.0.1 and a Maximum Likelihood (ML) tree was computed for each using PhyML. The model chosen was GTR with 100 bootstraps. The resulting trees were visualised with FigTree.

Table 8.1 - Summary of the ancient *Bos* Y chromosomes generated in this study. AH = Andrew Hare, AS = Amelie Scheu, BL = Bastien Llamas, CR = Conor Rossi, DD=Deborah Diquelou, KM = Kieren Mitchell, MV = Marta Verdugo, RM = Rita Monteiro, VEM = Victoria E. Mullin.

Sample ID	Species	Given Context/Age	Country	Region/Site	Haplotype	Coverage	Laboratory Work
Dzh1	*Bos taurus*	Neolithic	Bulgaria	Dzulhunitsa	Y2b-earlydom	9.68	AS/VEM
Che1	*Bos taurus*	Neolithic	Bulgaria	Chernomorets	Y2 (unplaced)	0.41	CR
Den2	*Bos taurus*	Neolithic	Denmark	Syltholm	Y2b-earlydom	3.93	VEM/CR
Bis1	*Bos taurus*	Neolithic	England	Bishops Cannings	Y2c	3.02	VEM
Bis2	*Bos taurus*	Neolithic	England	Bishops Cannings	Y2c	2.74	VEM
Shu1	*Bos taurus*	Neolithic	Georgia	Shulaveris Gora	Y2 (unplaced)	1.53	KD/VEM
New1	*Bos taurus*	Neolithic	Ireland	Newgrange	Y2b-earlydom	16.38	VEM
New3	*Bos taurus*	Neolithic	Ireland	Newgrange	Y2b-earlydom	5.20	VEM
New5	*Bos taurus*	Neolithic	Ireland	Newgrange	Y2b-earlydom	4.70	VEM
Par2	*Bos taurus*	Neolithic	Ireland	Parknabinia	Y2b-earlydom	2.02	VEM
Lud3	*Bos taurus*	Neolithic	Poland	Ludwinowo	Y2b-earlydom	1.80	MV/VEM
Kie2	*Bos taurus*	Neolithic	Poland	Kierzkowo	Y2a-basal	0.97	VEM
Ness5	*Bos taurus*	Neolithic	Scotland	Ness of Brodgar	Y2c	2.73	VEM
Plo4	*Bos taurus*	Neolithic	Serbia	Pločnik	Y2b-earlydom	2.86	(Verdugo et al. 2019)
Plo3	*Bos taurus*	Neolithic	Serbia	Pločnik	Y2b-earlydom	2.63	(Verdugo et al. 2019)
Sub1	*Bos taurus*	Neolithic	Turkey	Suberde	Y2b-earlydom	13.85	(Verdugo et al. 2019)
Men1	*Bos taurus*	Neolithic	Turkey	Menteşe	Y2b-earlydom	1.85	(Verdugo et al. 2019)
Ulu3	*Bos taurus*	Neolithic	Turkey	Ulucak	Y2a-basal	0.55	DD
Lec4	*Bos taurus*	Bronze Age	England	Lechlade Skatepark Gloucester	Y2a-basal	2.46	VEM/CR

Irt6	*Bos taurus*	Bronze Age	England	Irthlingborough	Y2a-basal	1.29	RM/VEM
Irt3	*Bos taurus*	Bronze Age	England	Irthlingborough	Y2a-basal	1.02	RM/VEM
Has3	*Bos taurus*	Bronze Age	Iran	Hasanlu Tepe	Y2 (unplaced)	1.50	(Verdugo et al. 2019)
Enk4	*Bos taurus*	Bronze Age	Netherlands	Enkhuizen-Kadijken	Y1-basal	2.14	VEM/CR
Enk1	*Bos taurus*	Bronze Age	Netherlands	Enkhuizen-Kadijken	Y1-basal	1.91	VEM/CR
Cla2	*Bos taurus*	Bronze Age	Scotland	Cladh Hallan	Y2b-earlydom	1.86	VEM
Stra1	*Bos taurus*	Bronze Age	Scotland	Strathglebe, Skye	Y2b-earlydom	1.25	RM/VEM
New2	*Bos taurus*	Iron Age	Ireland	Newgrange	Y2b-earlydom	3.93	VEM
Bun2	*Bos taurus*	Iron Age	Netherlands	Bunnik	Y1-basal	1.76	MV/VEM
Bisk1	*Bos taurus*	Iron Age	Poland	Biskupin	Y2b-east	2.43	VEM
Bor1-VEM	*Bos taurus*	Iron Age	Scotland	Bornais	Y2b-earlydom	2.68	VEM
Kok1	*Bos taurus x Bos indicus*	Iron Age	Uzbekistan	Kok Tepe	Y3	0.48	MV/CR
Bri1	*Bos taurus*	Roman Age	Netherlands	Britsum	Y2a-Eu	7.38	MV/VEM
Viz2	*Bos taurus*	Roman Age	Netherlands	Vianen-Zijdervel	Y1	5.75	MV/VEM
Otb1	*Bos taurus*	Roman Age	Netherlands	Oosterbeintum	Y1-basal	2.58	MV/VEM
Hou1	*Bos taurus*	Mediaeval	Netherlands	Houten-Loerik	Y2a-Eu	2.56	MV/VEM
Els1	*Bos taurus*	Roman Age	Netherlands	Elst	Y1-basal	2.52	MV/VEM
Yor7	*Bos taurus*	Mediaeval	England	Coppergate	Y2a-Eu	6.63	VEM
Yor12	*Bos taurus*	Mediaeval	England	Hungate	Y2a-Eu	4.48	VEM
Yor2	*Bos taurus*	Mediaeval	England	Hungate	Y2a-Eu	3.91	VEM
Yor5	*Bos taurus*	Mediaeval	England	Hungate	Y2a-Eu	2.96	VEM

Yor3	*Bos taurus*	Mediaeval	England	Coppergate	Y2a-Eu	2.50	VEM
Bes2	*Bos taurus x Bos indicus*	Mediaeval	Iraqi Kurdistan	Bestansur	Y3	13.80	(Verdugo et al. 2019)
Dro1	*Bos taurus*	Roman Age	Netherlands	Dronrijp-zuid	Y2b-earlydom	2.84	MV/VEM
Rhi1	*Bos primigenius*	47,000+ cal yrBP	Germany	Upper Rhine Valley	Y1-Y2-out	9.54	CR
Rhi2	*Bos primigenius*	47009 - 45312 cal yrBP	Germany	Upper Rhine Valley	Y1-Y2-out	2.12	CR
Tula1	*Bos primigenius*	40746 - 42460 cal yrBP	Russia	Tula	Y4a	3.51	MV/CR
Gyu1	*Bos primigenius*	14097 - 13863 cal yrBP	Armenia	Gyumri	Y4a	18.18	MV/CR
Yen1	*Bos primigenius*	14193 - 14019 cal yrBP	Russia	Yenisei River	Y4b	14.90	
Baikal1	*Bos primigenius*	12015 - 12529 cal yrBP	Baikal	Russia	Y4b	13.58	MV
Rhi3	*Bos primigenius*	Pleistocene	Germany	Upper Rhine Valley	Y2c	3.14	CR
Tri1	*Bos primigenius*	10291 - 9976 cal yrBP	Germany	Trier-Brüderkrankenhaus	Y2a-basal	2.24	AS
Ska3	*Bos primigenius*	9731 - 9508 cal yrBP	Sweden	Skåne	Y2c	2.95	CR
Gal1	*Bos primigenius*	9538 - 9109 cal yrBP	Spain	Galicia	Y2c	17.08	CR
Gal3	*Bos primigenius*	9481 - 9100 cal yrBP	Spain	Galicia	Y2c	2.93	CR
Gal2	*Bos primigenius*	9481 - 9100 cal yrBP	Spain	Galicia	Y2a-basal	0.85	CR
Ch22	*Bos primigenius*	7750 - 7600 cal yrBP^	Turkey	Çatalhöyük	Y2a-basal	0.31	(Verdugo et al. 2019)
Hjo1	*Bos primigenius*	8245 - 8053 cal yrBP	Denmark	Hjørring, Tofte Bæk	Y2c	1.14	CR
Var1	*Bos primigenius*	7641 - 7500 cal yrBP	Kazakhstan	Varfolomeevka	Y2a-basal	4.55	CR
Uzz1	*Bos primigenius*	7593 - 7480 cal yrBP	Italy	Sicily	Y2c	2.00	CR
Var2	*Bos primigenius*	7486 - 7238 cal yrBP	Kazakhstan	Varfolomeevka	Y4a	9.00	CR
Gyu2	*Bos primigenius*	7394 - 7239 cal yrBP	Armenia	Gyumri	Y2a-basal	1.70	(Verdugo et al. 2019)
Cpc	*Bos primigenius*	6884 - 6741 cal yrBP	England	Carsington Pasture Cave	Y2c	6.00	(Park et al. 2015)

Bor1	*Bos primigenius*	6740 - 6521 cal yrBP	Kazakhstan	Borly	Y4a	0.30	CR
Siv1	*Bos primigenius*	6258 - 6006 cal yrBP	Bulgaria	Vratsa	Y1-basal	9.03	CR
Blec1	*Bos primigenius*	Neolithic	Bulgaria	Vratsa	Y2a-basal	0.99	CR
Kalisz2	*Bos primigenius*	5813 - 5668 cal yrBP	Poland	Kalisz	Y2a-basal	1.79	(Verdugo et al. 2019)
Tango2	*Bos primigenius*	4599 - 4379 cal yrBP	Former USSR	Unknown	Y2a-basal	2.09	MV
Kalisz1	*Bos primigenius*	4487 - 4229 cal yrBP	Poland	Kalisz	Y2a-basal	0.65	CR
Kalisz3	*Bos primigenius*	4477 - 4168 cal yrBP	Poland	Kalisz	Y2a-basal	2.60	MV/CR
Belarus1	*Bos primigenius*	4476 - 4173 cal yrBP	Belarus	Brest district	Y2a-basal	1.78	MV
Tango1	*Bos primigenius*	4142 - 3914 cal yrBP	Former USSR	Unknown	Y2a-basal	1.78	MV
Gallo1	*Bos primigenius*	3764 - 3479 cal yrBP	Scotland	Galloway	Y2c		CR
Fal7	*Bos primigenius*	Mediaeval	Germany	Falkenwalde	Y2a-basal	1.53	AS

8.4. Results

8.4.1. Bayesian phylogenetic analysis

Neither reference01 BEAST run converged after 350 million chains. The ESS of several parameters were below 100 with 30% burn in, indicating poor mixing. In particular, the rate.mean and ucldMean did not converge in either run. For this reason, this Bayesian tree was not included in further analysis.

In contrast, all independent runs of reference02 and reference03 datasets converged after 350 million chains. All parameters of reference03 converged with an ESS of >300, except for height(Rhi1) which converged with an ESS of ~160 and TreeHeight with >200 in each run. In reference02, all parameters converged with an ESS of >1000. The tree topologies of reference02 and reference03 agree, although there are some differences in divergence times. Discussion of divergence times will be based on reference02 due to the better mixing. The substitution rates of the reference02 with and without invariant sites specified are listed in Table 8.2.

Table 8.2 - Substitution rates of reference02 with and without invariant sites, as per Tracer.

	Invariant sites specified	Invariant sites not specified
rate.mean	3.66E-09	1.27E-06
stderr of mean	3.35E-11	1.76E-07

The generated taurine tree depicts lineages described in Chen et al. 2018 with several undescribed Y chromosomal lineages also emerging.

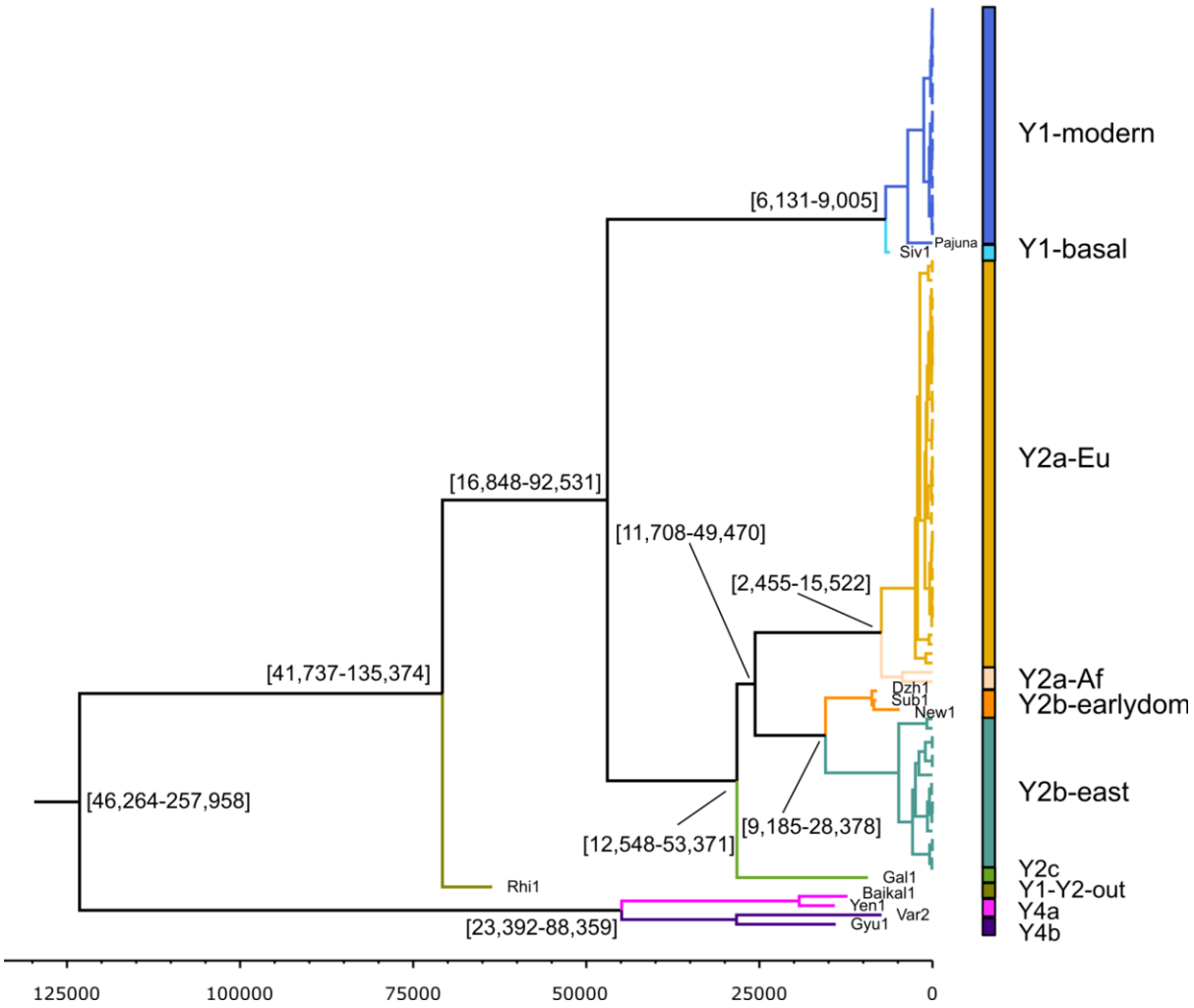


Figure 8.1 - highest likelihood Bayesian phylogeny generated using data from reference02 with HPD 95% of internal node splits indicated.

The deepest divergence splits east Eurasian aurochs lineages (Y4) from European aurochs and all extant and extinct taurine domestic animals (Y1-Y2). This divergence occurred pre-LGM but has a wide margin of divergence (median node age = 123,183 yrBP, HPD 95% = 46,264 - 257,958 yrBP). Within the Y4 branch, a divergence in clades (Y4a versus Y4b), occurring during or before the LGM is observed (median node age = 44,894 yrBP, HPD 95% = 23,392 - 88,359 yrBP). Geographic structuring within the clade is observed with Y4a comprising Siberian Pleistocene aurochs and Y4b comprising aurochs from Pleistocene Caucasus (Gyu1) and Holocene Central Asia (Var2).

A Pre-LGM aurochs (Rhi1) outgroups all lineages along the Y1-Y2 taurine branch. This aurochs exceeded radiocarbon dating limits and so populated central Europe prior to the LGM, over 47,000 years ago. Further along the branch, Y1 diverged from Y2 in the pre-Holocene (median node age = 46,962 yrBP, HPD 95% = 16,848 - 92,531 yrBP). The Y1 branch is

punctuated by Siv1 (a Bulgarian Mesolithic aurochs) which sits along the base of the branch (Y1-basal). Within the Y1-modern clade, a single Iberian modern breed (Pajuna) outgroups all Y1 breeds of Britain, Ireland or northwestern Europe (median node age = 3,579 yrBP, HPD 95% = 1,247 - 6,223 yrBP).

A newly described Y2c lineage also emerges along the Y2 branch, outgrouping both Y2a and Y2b clades. This lineage is populated by a single Iberian Holocene aurochs (Gal1). Y2c diverged from the common ancestor of Y2a and Y2b during the Last Glacial Period (LGP) (median node age = 28,247 yrBP, HPD 95% = 12,548 - 53,371 yrBP). The modern haplotypes of Y2a and Y2b diverged slightly later, although still during the LGP (median node age = 28,247 yrBP, HPD 95% = 11,708 - 49,470 yrBP).

Modern geographic structure is observed in the Y2a clade. A Y2a clade consisting of mostly southern European taurines (Y2a-Eu) and is outgrouped by a smaller clade of West African taurine cattle (Y2a-Af). This divergence predates modern breed formation (median node age = 7,390 yrBP, HPD 95% = 2,455 - 15,552 yrBP). The Y2b clade similarly displays previously undescribed structure. A clade of ancient domesticates (Y2b-earlydom) comprises three Neolithic bulls from Anatolia (Sub1) and Europe (Dzh1; Bulgaria and New1; Ireland). These lineages form a distinct clade that diverged from the modern Y2b haplotype (Y2b-east) prior to domestication or very early in the domestication process (median node age = 15,450 yrBP, HPD 95% = 9,185 - 28,378 yrBP).

Several star-like expansions are observed in the tree topology, best observed in Figure 8.2. Node dating of these putative haplotype expansions are listed in Table 8.3.

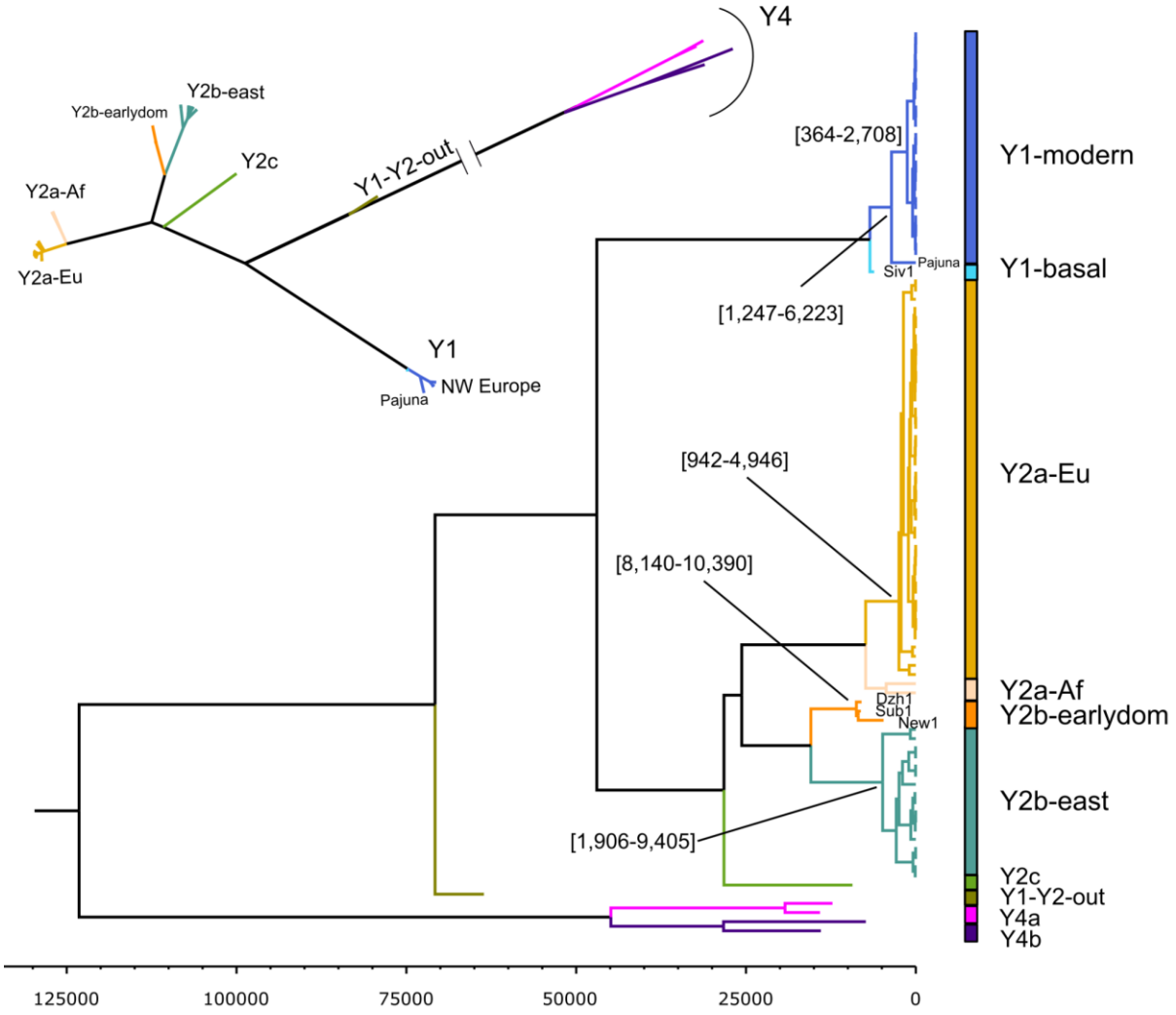


Figure 8.2 - Most likely Bayesian phylogeny generated using data from reference02 with several putative domestic tooth-comb phylogenies depicted.

Table 8.3 - TMRCA of putative domestic tooth-comb phylogenies.

Haplotype	Median node age of TMRCA (yrBP)	HPD 95% (yrBP)
Y2b-earlydom	8754	8,140-10,390
Y2b-east	4883	1,906-9,405
Y2a-Eu	2487	924-4,946
Y1-modern	3579	1,247-6,223
Y1-modern (northwestern European breeds only)	1264	364-2,708

8.4.2. Low coverage phylogenies

The general tree topology of the low coverage ML phylogenetic trees reflects that of the Bayesian tree. This ML tree also expands to include Y3, which was excluded from the Bayesian analysis due to poor MCMC convergence. Geographic structure emerges within Y3. The Y3a sub-haplotype consists of Chinese indicine and hybrid breeds while the Y3b clade consists of south Asian indicine breeds, as described in (N. Chen et al. 2018). A Mediaeval Iraqi hybrid (Bes2) falls into the Y3b clade. In the transversions only ML tree, an Iron Age Uzbekistan sample (Kok1) has an insecure placement as basal to Y3 in this phylogeny. However, in the tree including transitions, Kok1 falls within the Y2c clade.

Y4 now expands to include Tula1, a ~41,000 year old aurochs recovered in Tula, Russia. Tula1 appears to sit on the base of the branch leading to the Y4a versus Y4b divergence, suggesting these lineages diverged after Tula1. Similarly, Rhi2 sits along the branch leading to Y1 and Y2 lineages, suggesting that all Y1 and Y2 samples are ancestral to these Rhine samples. Rhi1 now also sits basal along the Y1/Y2 branch while it outgrouped these lineages in the Bayesian phylogeny.

The Y1-basal branch is now populated by ancient Dutch domesticates from the Bronze Age (Enk1 and Enk4), Iron Age (Bun2) and Roman era (Els1). The single modern *Pajuna* from southern Iberia again outgroups the modern Y1 clade. Roman and Mediaeval domesticates (Otb1 and Viz2) fall into the Y1-modern clade. In addition, a ~6,000 year old aurochs from central Asia (Bor1-CR) is basal to all samples along the Y1 branch in the full ML tree. However, this placement is not well-supported as Bor1-CR switches to the Y2c clade in a transversions-only ML tree.

The Y2c group now expands to include thirteen ancient samples, including seven pre-domestic contact European aurochs. These Ys are widely dispersed across Europe, in Iberia, Italy, Britain, Central Europe and Scandinavia. One Scandinavian sample (Ska3) changes positions to a basal Y2a in the transitions-only ML tree. In post-Neolithic Europe, only one wild Y2c was detected in a Bronze Age Scottish sample (Gallo1). Y2c also appears exclusively in domestic animals of Neolithic Britain (Bis1, Bis2, Ness5). Y2c appears geographically confined to Europe except for two domesticates: the Anatolian Neolithic Ulu3 and the insecure transitions-included placement of the Central Asian Iron Age domesticate, Kok1.

The Y2a branch now includes numerous animals that are basal to the modern Y2a-Eu and Y2a-Af clades. These haplotypes are tentatively referred to as Y2a-basal. Both domestic and

wild animals are observed here ranging in age from ~10,000 years old (Tri1) to ~1,000 years old (Fal7). A Bronze Age Iranian domesticate (Has3) outgroups the clade but moves to Y2c in the transversions-only ML tree. Further along the Y2a branch is the aforementioned divergence of Y2a-Eu and Y2a-Af. Along this branch now sits a modern Anatolian breed (South Anatolian Red). A number of ancient animals now appear within the Y2a-Eu clade. These animals are from northwestern Europe and postdate the Iron Age. The earliest animal is a British domesticate from the Roman era (Bri1), while the remainder are from Medievael Britain (Yor2, Yor3, Yor5, Yor12) and Medievael Netherlands (Hou1).

The Y2b group expands to include 20 ancient domestic animals. In the ML tree 5 of these animals (Lud3, Pot3, Cla2, Plo3, Che1) are basal to Y2b-east/Y2b-earlydom split. However, 4 of these samples (Lud3, Pot3, Cla2, Plo3) fall into the Y2b-earlydom haplogroup in the transitions-only ML tree. Neolithic/Chalcolithic animals from the Anatolia (Sub1, Men1), Ireland (Par2, New1, New5), Scandinavia (Den2), Poland (Lud2) and the Balkans (Dzh1, Plo3, Plo4, Che1) are also placed here. Later two Scottish Bronze Age domesticates (Stra1, Cla2) appear here. While two British (Pot3, Bor1) and one Irish (New2) domesticates from the Iron Age also appear in the Y2b-earlydom clade. The last occurrence of Y2b-earlydom is the Roman era Dutch sample, Dro1.

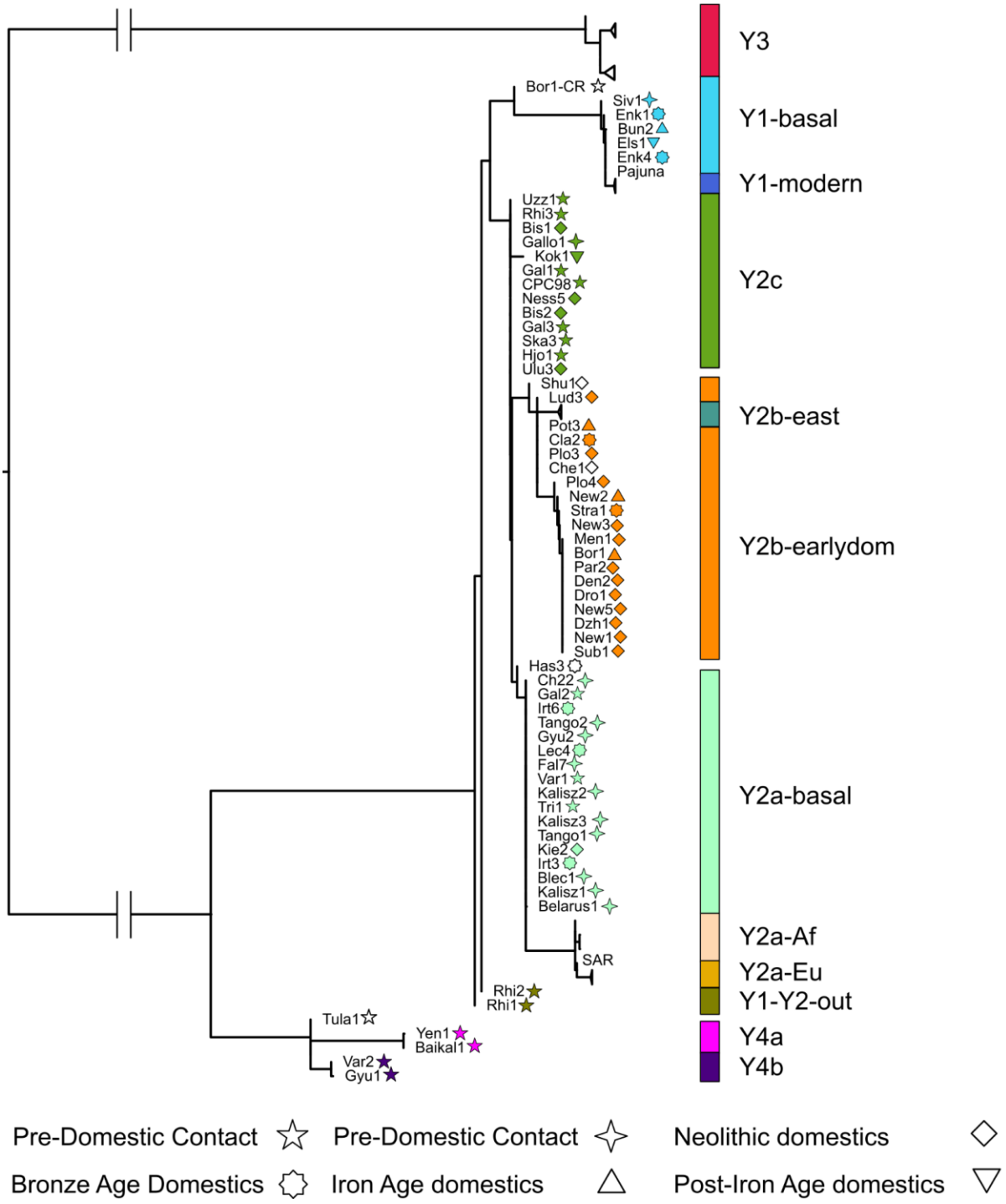


Figure 8.3 - A Maximum Likelihood phylogeny of ancient and modern bovine Y chromosomes.

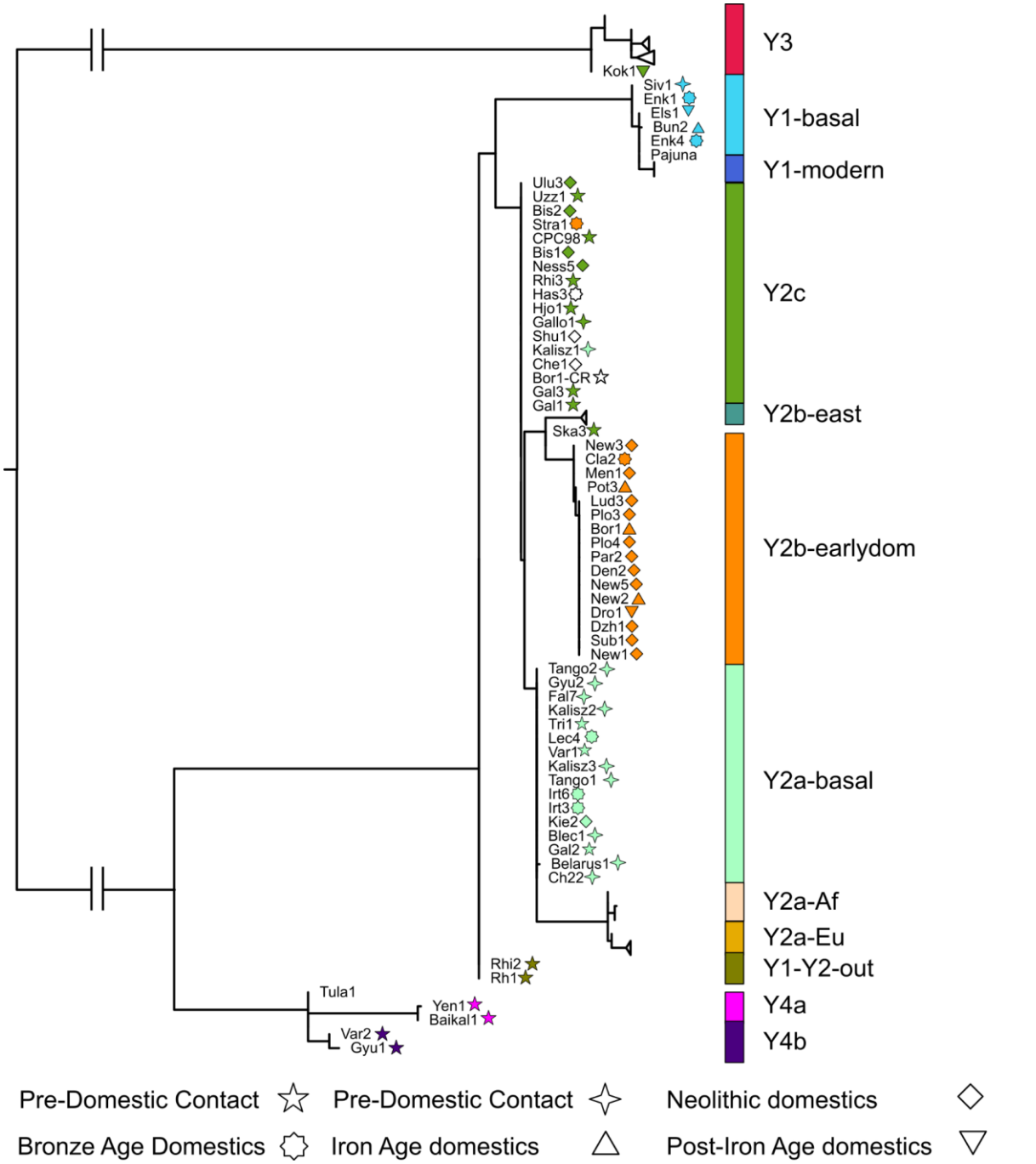


Figure 8.4 - A Maximum Likelihood phylogeny of ancient and modern bovine Y chromosomes, excluding transitions.

Table 8.4 - Earliest and last appearances of Y chromosomal haplotypes in this dataset.

Haplotype	Earliest appearance	Last appearance
Y1-modern	Netherlands Mediaeval (1028 - 861 cal yrBP)	Modern
Y1-basal	Bulgarian Neolithic wild (6258 - 6006 cal yrBP)	Dutch Roman Era domesticate (1900 - 1700 yrBP)
Y2c	German Mesolithic wild (~12000 yrBP)	British Bronze Age wild (3764 - 3479 cal yrBP)
Y2b-earlydom	Anatolian domesticate (8244 - 8046 cal yrBP)	Dutch Roman Era domesticate (1900 - 1700 yrBP)
Y2b-east	Polish Iron Age	Modern
Y2a-basal	German Mesolithic wild (~10,000 yrBP)	German Mediaeval wild (~1,000 yrBP)
Y2a-Eu	Dutch Roman domesticate (1766 - 1601 cal yrBP)	Modern
Y2a-Af	Modern	Modern

8.5. Discussion

8.5.1. Technical considerations

When constructing the Bayesian phylogenetic tree, several considerations were necessary to obtain accurate results. Firstly, better MCMC convergence was observed when invariant sites were accounted for without much extra computational burden. The substitution rate when accounting for invariant sites was also much closer to the substitution rates of other mammalian Y chromosomes, suggesting a more accurate tree was obtained this way (*e.g.* a substitution rate of 1E-09 has been reported for human Y chromosomes (Xue et al. 2009)). Notably, Y chromosomal divergence dating using BEAST has been done in humans using only SNP calls without acknowledgement of invariant sites (Batini et al. 2015; Karmin et al. 2015). Disregarding the invariant sites may have implications for topological accuracy (Leaché et al. 2015).

Poor MCMC convergence was observed when constructing Bayesian phylogenetic trees using *B. indicus* sequences. Since their divergence approximately 200 thousand years ago, the size and morphology of *B. taurus* and *B. indicus* Y chromosomes have become distinct, with differing locations of the PAR (Meo et al. 2005). A lack of convergence may have been due to poor homology between *B. indicus* Y chromosomes and the reference Y chromosome (GCF_000003205.7), which is derived from a Y1 Hereford bull. Alignment of haploid chromosomes to a single reference genome has been shown to introduce systematic biases

and error in phylogeny reconstruction (Bertels et al. 2014). Therefore, a more accurate bovine Y chromosomal phylogeny may require a specific *B. t. indicus* reference Y chromosome. Indeed, mitochondrial genome consensus calling was greatly improved when alignment was against a haplotype specific reference (Chapter 7). However, the larger size and repeat-rich composition of Y chromosomes make haplotype-specific Y chromosome reference construction a more difficult task (Cortez et al. 2014).

In addition to high-quality Bayesian phylogenies, simplified ML phylogenies of low and high depth genomes were constructed using a filtered SNP set. Several potential biases were identified using this approach. Firstly, genetic inferences based on pre-ascertained SNPs can introduce bias, distorting genetic inferences (Lachance and Tishkoff 2013). This may explain some inconsistencies that arose when comparing the Bayesian trees to the ML phylogenies. For example, it was observed that Rhi1 outgroups Y1-Y2 clades in the Bayesian phylogenetic tree while it sits along the branch leading to Y1-Y2 clades in the reduced-SNP ML tree. In the reduced-SNP ML tree, the lineage-specific variation of Rhi1 was removed using a MAF filter which may have caused this branch shortening. Additionally, while SNPs were called on a curated database of animals in an attempt to balance lineage specific variation that may have arisen during breed formation, this database was still biassed toward modern samples. A SNP ascertainment bias may have also had an even larger effect on lineages that had no representatives in the curated SNP call database, such as Y2a-basal. Y2a-basal may therefore be a poorly defined group as a “basal” haplotype to Y2a-Eu/Y2a-Af. While some of the Y2a-basal samples may truly be the paternal ancestor to the modern Y2a haplotypes, it is also noted that several of the Y2a-basal animals postdate TMRCA of Y2a-Eu/Y2a-Af. For example, the Y2a-basal Fal7 is an approximately 1,000 year old Mediaeval German aurochs yet it is basal to the divergence of Y2a-Eu/Y2a-Af which occurred at least 1,500 years previous (median node age = 7,390 yrBP, HPD 95% = 2,455 - 15,552 yrBP). For this reason, these lineages should not be thought as the progenitor of the modern Y2a haplotypes but rather derived representatives of a sister clade.

Two reduced-SNP ML trees were constructed - a tree with all sites and a tree with transversions sites only. This was to account for the potential of false variation caused by transitions due to aDNA damage. In total 7 lower coverage samples had inconsistent placements (Bor1-CR, Kok1, Shu1, Stra1, Ska3, Has3, Che1). This may be due to incorporation of aDNA damage as false variation, as these low coverage samples would be more vulnerable to this effect (Parks and Lambert 2015). Equally, the low coverage aspect of these samples may have meant that there was insufficient information to place the samples

with the lower number of transversion sites. In either case, caution is necessary when interpreting the placement of these samples.

8.5.2. Deep origins of the aurochs Y

The deepest node in the ML Y trees describes the divergence of taurine Y chromosomal variation versus indicine Y chromosomal variation. Unlike the mitochondrial sequences, this deep divergence of indicine versus taurine animals could not be dated by BEAST due to the poor convergence of the reference01 dataset. A mediaeval Iraqi hybrid (Bes2) is placed in Y3b clade, which is populated by south Asian indicine. Bes2 is a hybrid animal deriving approximately equal parts ancestry from taurine and indicine lineages (Verdugo et al. 2019). Bes2 has a taurine mitochondrial haplotype, reflecting the current maternal and paternal dichotomy of many Near Eastern hybrids (Kantanen et al. 2009). As previously suggested, this may indicate that drought resistance was incorporated into Near Eastern herds primarily via male-mediated indicine introgression (Verdugo et al. 2019).

The deepest node within *B. primigenius primigenius* variation describes a divergence of east Eurasian aurochs (Y4) versus European and southwest Asian (Y1 and Y2) aurochs. This divergence occurred before the onset of the LGM (median node age = 123,183 yrBP, HPD 95% = 46,264-257,958 yrBP). This timing may be contemporaneous with the rapid radiation of maternal lineages pre-LGM between the years of 92,767 and 55,930 BP (Chapter 7). However, due to the large confidence intervals, more recent divergences may also be contemporaneous with the maternal radiation - Y1 vs Y2 (median node age = 46,962 yrBP, HPD 95% = 16,848-92,531 yrBP) and Y4a vs Y4b (median node age = 44,894 yrBP, HPD 95% = 23.302-88,359 yrBP). The larger confidence intervals in deeper time underpin a weakness of Y chromosomal divergence dating. In the reduced-SNP ML tree, three putative Y-MRCA or 'Y chromosomal Adams' are uncovered for eastern and western lineages in pre-LGM animals. A ~41,500 year old east Eurasian aurochs (Tula1) sits along the branch leading to Y4a and Y4b. Similarly, Rhi1 and Rhi2 sit along the branch leading to Y2 and Y1. Their position along the branch of this clade suggests that Y1 diverged from Y2 after Rhi1/Rhi2 and Y4a diverged from Y4b after Tula1. However, an important consideration is that Rhi1 actually outgroups the Y1-Y2 lineages in the Bayesian trees. As discussed above, the method of SNP ascertainment in the reduced-SNP ML removes SNPs that arise on a single lineage. Therefore, considering Rhi1, Rhi2 and Tula1 as the Y-MRCA of their respective lineages may be unjustified.

The pre-LGM aurochs Rhi1 and Rhi2 have a clearly taurine Y chromosome that ingroups from indicine variation, unlike their mitochondrial lineage. In hybrid zones, introgressive

hybridisation often proceeds asymmetrically from the local population into the colonising population (Currat et al. 2008). For example, in the case of ancient bears on the ABC Islands, the local population of polars bears hybridised with a migrating, mostly male brown bear population (Cahill et al. 2015). This resulted in a hybrid population that derived most of their ancestry from the immigrant brown bears, including Y chromosomes, but retained the indigenous polar bear mitochondrial haplotype. A similar non-monophyly of Y chromosomal and mitochondrial lineages is observed in Rhi1 and Rhi2. This may have also resulted from a population of mostly male immigrants expanding their range into a hybrid zone of pre-LGM Europe. Such sex-biassed gene flow reflects in the social structures of other *Bovini* species, whereby solitary males tend to leave the female herd, seeking mates (van Vuure 2005). Similarly, as in most mammals, it is likely that aurochs cows were more philopatric than their male counterparts (Greenwood 1980). In this scenario, the immigrant males encountered an undescribed, or ghost, population of *Bos primigenius* in Europe and subsequently hybridised. Based on the divergent mitochondrial haplotype of the Rhine aurochs (G and N), it is possible that the local European population was an outgroup to taurine and indicine animals.

8.5.3. Holocene wild structure

The newly described lineage Y2c is the most common aurochs Y haplogroup in pre-domestic Europe and was widely dispersed across the continent in the early Holocene. This haplotype diverged from the Y2a-Y2b lineages during the Last Glacial Period (median node age = 28,247, HPD 95% = 12,548 - 53,371) - possibly during the LGM when much of northern Europe was covered in ice. Y2c may therefore be analogous to the mitochondrial haplotype P, which also arose during the LGM and was restricted to European aurochs. However, unlike the P haplotype, the natural range of Y2c appears to extend into southwestern Asia. Y2c was also genotyped in a single domestic animal (Ulu3) from a western Neolithic Anatolian site, Ulucak. As Ulucak was occupied prior to the Neolithisation of Europe, the Y2c haplotype must have been recruited from a wild population that existed in southwest Asia at least 8,000 years ago. In post-domestic contact Europe, the observations of Y2c diminishes to one single aurochs, Gallo1. This ~3,600 year old Scottish aurochs is among the final aurochs on this island, which are thought to have been completely extirpated from Britain only a few hundred years later (van Vuure 2005).

Y2a-basal is a group of Y haplotypes that place along a branch leading to modern Y2a haplotypes. Y2a-basal is observed in early Holocene aurochs from central Europe (Tri1), Iberia (Gal2) and Anatolia (Ch22). The contextual dating of Ch22 also precedes the Neolithisation of Europe which implies that Y2a-basal was present in wild populations of southwest Asian

aurochs. Most post-domestic contact aurochs, including the most recent aurochs in our dataset, are Y2a-basal. As Y2a-basal is found in both pre-domestic contact Europe (Tri1, Gal2) and European domesticates (Kie2, Lec4, Itr3, Itr6), it is unclear if this preponderance of Y2a-basal in post-domestic contact European aurochs is the result of domestic bulls introgressing into the wild, a wild expansion across Europe, or both. As discussed, SNP ascertainment bias may have led to poor resolution of lineages along the Y2a-basal branch. Higher depth sequencing of Y2a-basal bulls and further sampling may help to resolve ambiguity associated with the origin and spread of Y2a-basal.

Taken together, the dispersal of Y2c and Y2a-basal bulls indicates that wild paternal ranges overlapped in European and southwest Asian populations post-LGM. This contrasts with the observed maternal geographic structure of post-LGM aurochs in European and southwest Asia. To date, the European mitochondrial P haplotype has never been observed in southwest Asia while similarly, robust evidence of pre-Neolithic T/Q haplotypes in European aurochs is not known. Limited gene flow between the Anatolian Peninsula and southeast Europe has been observed in a number of terrestrial mammal species (Demirbaş et al. 2019).

Migration of disparate wild male populations to Central Asia is also observed by the presence of Y haplotypes. Aurochs did not inhabit Central Asia during much of the LGP, likely due to an unfavourable habitat caused by the cold climate (van Vuure 2005). In the post-LGM, much of the northern ice-sheet receded which opened up Central Asia as a suitable habitat for expanding populations of aurochs. Central Asia was revealed as a Holocene aurochs hybrid zone of European, southwest Asian and east Asian refugia by tracking maternal lineages (Chapter 7). While only three aurochs bulls were genotyped from Holocene Central Asia, each had a distinct genotype (Y4b, a tentative Y1-basal, Y2a-basal), further suggesting male migration from multiple eastern and western refugia occurred in the Holocene.

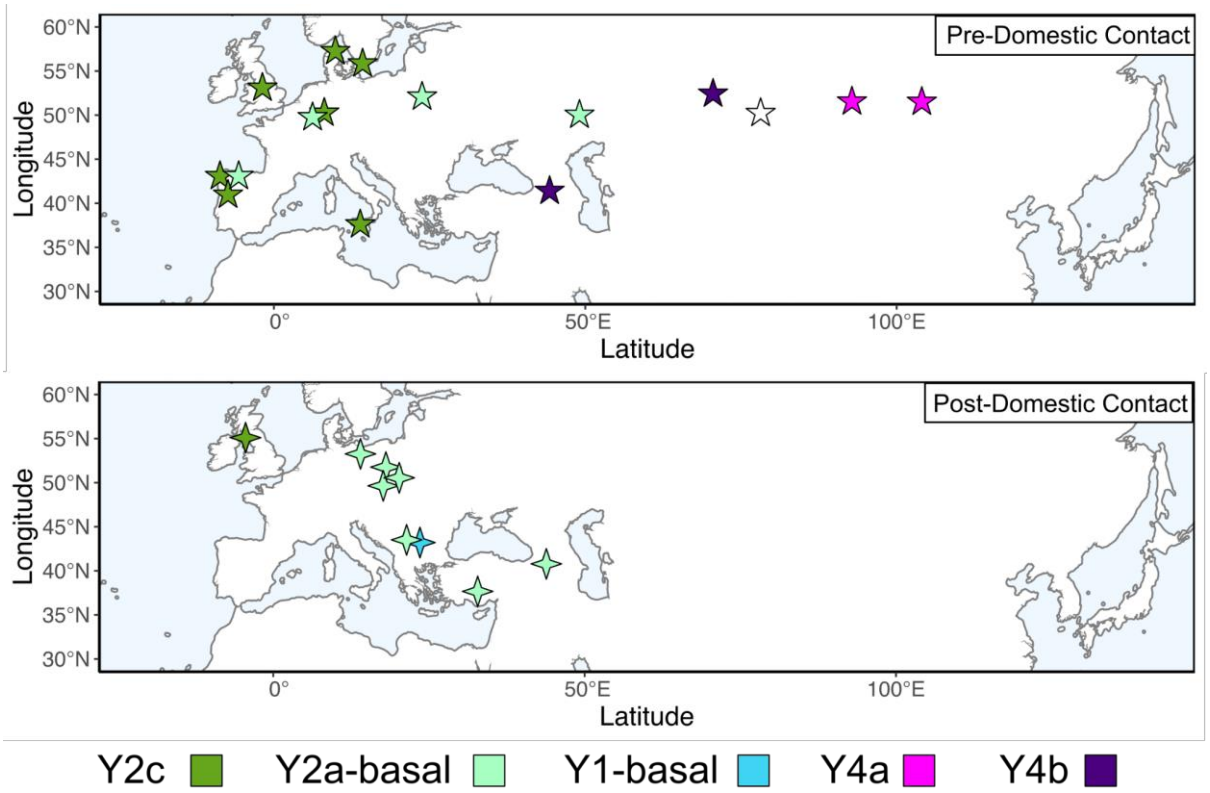


Figure 8.5 - Aurochs Y haplotypes across Eurasia from the LGM to their extinction.

8.5.4. Domestication of a Neolithic Y chromosome

After domestication in the Upper Euphrates Valley, taurine cattle moved with early farmers across southern, southwestern and northwestern Anatolia as part of the Neolithic package (Arbuckle et al. 2014). This westward dispersal was complex and the four core animals of the Neolithic package did not always move together. The domestic bulls of central and northwestern Anatolian sites (Suberde and Menteşe) are in the Y2b-earlydom clade. The node age of Y2b-earlydom expansion (median node age = 8,956 yrBP, HPD 95% = 8,140-10,390 yrBP) accords with the time of domestication, supporting a scenario that TMRCA of Y2b-earlydom could have been within the bulls used in the domestication process. In addition, a western coastal Anatolian domestic bull from Ulucak (Ulu3) has a Y2c haplotype, previously only seen in European aurochs. As the TMRCA of Y2b and Y2c pre-dates domestication by at least two thousand years (median node age = 28,247 yrBP, HPD 95% = 12,548 - 53,371 yrBP), the paternal ancestors of these two clades of Neolithic bulls may not have been domesticated from the same population of bulls.

While Y2b-earlydom is a very early paternal lineage of Y chromosomes that was dispersed westward through the Neolithisation of Europe, modern east Asian taurine breeds belong almost exclusively to the Y2b-east haplotype. Therefore, two scenarios of domestication of

Y2b are possible. The first scenario is that a common ancestor of Y2b-earlydom and Y2b-east was domesticated in the Near East and these paternal lineages diverged as the animals were dispersed eastward and westward. The second scenario is that Y2b-earlydom and Y2b-east were already diverged prior to their recruitment to the domestic stock. Both scenarios are supported based on divergence dating of TMRCA of Y2b (median node age = 15,450 yrBP, HPD 95% 9,185 - 28,378 yrBP). In the expanded ML tree, several ancient domesticates (Lud3, Pot3, Cla2, Plo3, Che1) appear to be basal to Y2b-east and Y2b-earlydom which would support a Y2b-east/Y2b-earlydom common ancestor domestication origin of Y2b. However, this basal group may be a technical artefact due to aDNA damage as it disappears in the transversions-only ML tree. Very little ancient data is available east of the Upper Euphrates Valley which greatly reduces the power to trace lineages. A single Neolithic Georgian bull (Shu1) from Shulaveris Gora also reveals a poorly placed Y2 haplotype. The sample switches from a placement as a Y2b outgroup with transitions to Y2c without transitions. It is possible that Shu1 reflects the basal Y2b diversity that existed shortly after domestication. Overall, deeper sequencing and more diverse sampling is required to better understand the dispersal of domestic bulls from the Upper Euphrates Valley.

Whether cattle management in Çatalhöyük occurred is an open question. Cattle remains in this site are considered hunted wild aurochs due to their large size and horn shape remaining stable over a millennium of site occupation. However, some signs of animal management are also noted such as early kill-off patterns and overrepresentation of adult females in assemblages. For this reason, it has been suggested that animal management consisted of backcrossing domestic females with wild males (Peters, Arbuckle, and Pöllath 2014). The mitochondrial haplotype of Ch22 is T3, a firmly domestic haplotype but its Y chromosome is Y2a-basal and may be wild. Therefore, the genotyping of this animal supports such an animal management pattern.

From Anatolia, early farming communities spread westward into the Aegean Basin and the Danubian corridor approximately 8,000 years ago, likely through demic diffusion (Marchi et al. 2022). The exact route of migration into Europe is equivocal and may have taken a seafaring route from southwestern Anatolia into the Aegean or across the Bosphorus into the Balkans (Arbuckle et al. 2014). Although the sample size is small, it is notable that the well-supported domesticates from Neolithic Balkans are exclusively Y2b-earlydom, like the Neolithic domestics of central and northwestern Anatolian sites. From the precipice of southeast Europe, Neolithisation spread across Europe along two main dispersal routes: the Mediterranean route and the Danubian route (Marchi et al. 2022). The Mediterranean route brought Neolithic

technologies along the southerly coasts of Europe, eventually reaching Iberia while the Danubian route expanded along the Danube river to central Europe and Scandinavia. Tracing the Danubian route to Neolithic Poland (Ludwinowo) and Denmark (Syltholm) reveals a consistent Y2b-earlydom signal. The exception to this is one Y2q-basal bull (Kie2) which has significant wild input (Victoria Mullin, *per. comms.*). Overall, this suggests that Y2b-earlydom represents the paternal signal of the Neolithic package that spread across Europe via the Danubian route.

At the western fringes of Europe, a contrast in Neolithic bovine paternity is observed on the islands of Britain and Ireland. Bulls from Neolithic and Chalcolithic Ireland (Newgrange and Parknabinnia) are exclusively Y2b-earlydom while British Neolithic domesticates are exclusively Y2c. Significantly in this survey, local British aurochs were also exclusively Y2c. Neolithic farmers of Britain may have taken the opportunity to breed their imported Anatolian domestic cows with the local British Y2c aurochs. Archaeological and genomic evidence indicate that the first farmers of Britain and Ireland were derived from the same population of agriculturalists who sailed from Normandy and the western coast of France (Rivollat et al. 2020). While no Neolithic boats from these regions have been recovered, examples from other regions and periods suggest that only a few fully grown cattle would have been possible to transport across the English Channel at a time (Cummings and Morris 2022). For this reason, early farmers may have prioritised the movement of breeding females from France and opted to use the local wild population to restock the herds. This is reflected in Early Neolithic British site assemblages, where there is a clear sex bias toward cows (Serjeantson 2011). In addition, the more aggressive nature of males may have been a liability on fragile boats crossing the English channel. Overall, this provides evidence of wild restocking of domestic cattle in Neolithic Britain.

The preponderance of Y2b-earlydom bulls in Neolithic Ireland is intriguing when considering that all Neolithic British bulls sampled thus far have been Y2c. The proposed British local aurochs bull-imported domestic cow breeding strategy would not be available to Neolithic farmers in Ireland as aurochs never colonised Ireland post-LGM (van Vuure 2005). Multiple strands of Neolithisation in Ireland are proposed from both Britain and France, although this is currently debated (Sheridan 2010). The earliest known presence of cattle in Britain or Ireland was in the third quarter of the fifth millennium BC in Ireland based on seven cattle bones recovered from a Mesolithic site at Ferriter's Cove on the Dingle peninsula. This was interpreted as a possible false start strand of Neolithisation in Ireland, arriving from western France rather than Britain. Evidence of domestic cattle management would not appear again

in Ireland until the fourth millennium BC. These later Neolithic sites were part of the Carinated Bowl Neolithic culture, which had already spread across much of Britain. This would suggest either that Y2b-earlydom bulls were also present in Neolithic Britain or that Y2b-earlydom bulls were transported to Ireland from continental Europe. The existence of Y2b-earlydom in Ireland also hints at domestic dispersal along the Mediterranean route. British and Irish Neolithic human genomes have affinity to Iberian early farmers, indicating that their ancestry is most likely derived from Neolithic people who migrated from Anatolia via the Mediterranean route (Rivollat et al. 2020; Cassidy et al. 2020; Brace et al. 2019). While there is no genomic data from bulls of Neolithic Italy or Iberia, the existence of Irish Neolithic Y2b-earlydom bulls indicates that this haplotype was likely transported along this southern route as well as the Danubian route.

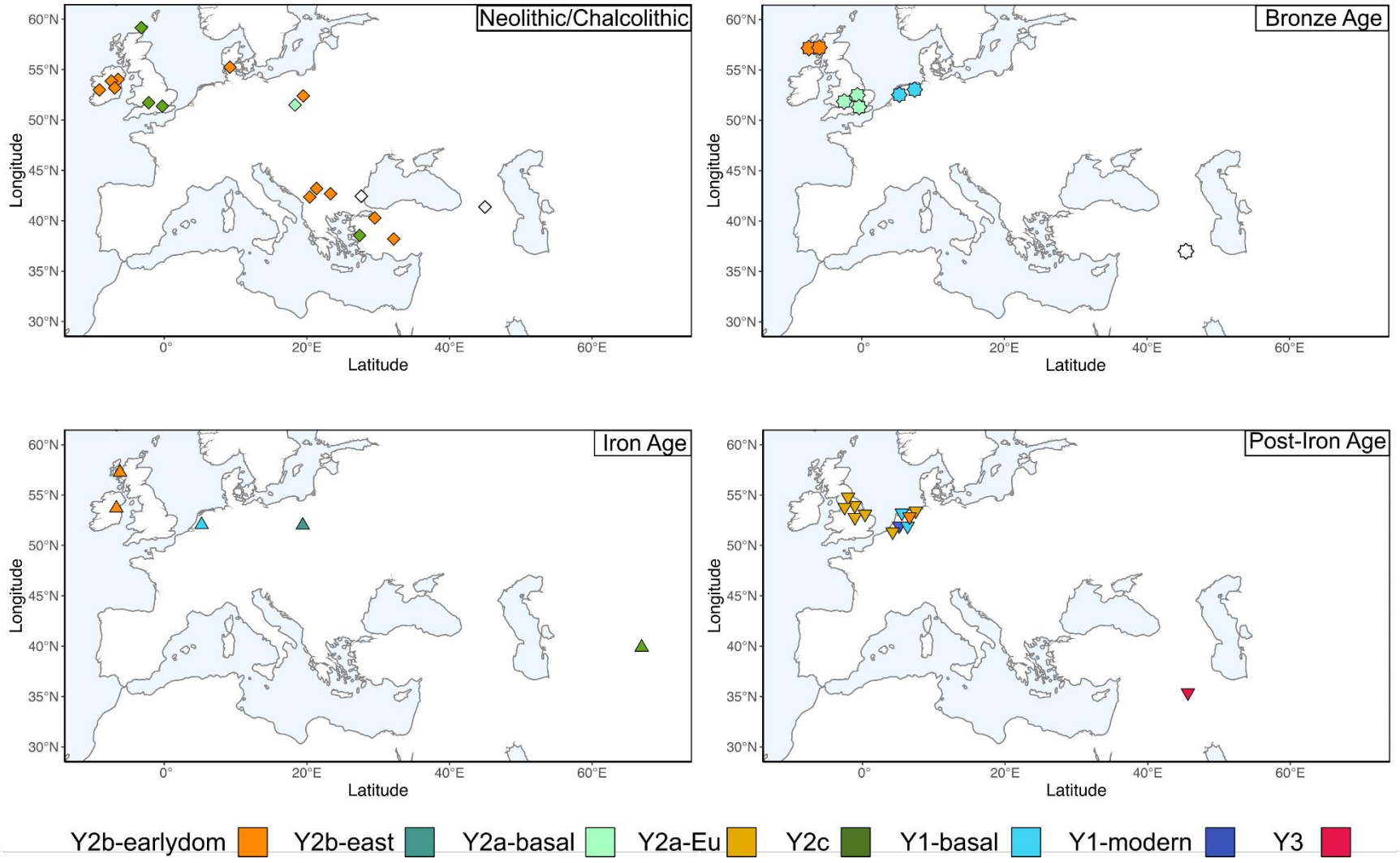


Figure 8.6 - Turnover of domesticate haplotypes through time based on the haplotypes assigned in the ML tree.

8.5.5. Wild recruitment and modern cattle haplotype formation

From the limited data in this study, it appears that Y2b-earlydom is a suitable candidate for the original Neolithic bull haplotype that spread across Europe with other agricultural technologies and traditions. Despite this, Y2b-earlydom is apparently extinct in modern cattle, indicating that a near-complete turnover of male ancestry has occurred since the Neolithic in Europe. This contrasts with the maternal lineages where most extant European cattle retain the T mitochondrial haplotype, the likely original domestic maternal lineage. As the ancestral haplotypes of resulting European variants are found in this analysis among European aurochs, this provides explicit evidence of wild recruitment into domestic European herds. In well-sampled regions like Britain, Ireland and the Netherlands multiple turnovers of Y chromosomal haplotypes are observed through time. It has previously been suggested that local wild restocking of cattle may have been a purposeful strategy of early farmers, possibly to acquire new advantageous traits that would adapt the imported Anatolian cattle to their new surroundings (Park et al. 2015; Orlando 2015). Although equally, restocking from wild surrounding bulls may have also been due to accidental encounters. In any case, wild restocking was strongly male-mediated as domestic cattle retain originally domesticated maternal lineages. The origin and spread of modern haplotypes will now be discussed.

Y1 recruitment and spread

The earliest occurrence of the Y1 haplogroup is found in a Y1-basal ~6,000 year old Neolithic Bulgarian aurochs (Siv1). With only one wild occurrence, it is not possible to ascertain if its presence in Siv1 is because this is the natural range of the haplogroup or if it is due to introgression into the wild population by domesticates. Nevertheless, the timing of divergence between Y1-basal and Y1-modern is concordant with Siv1 being a paternal ancestor of all extant Y1 domestic bulls.

The first domestic detection of a Y1-basal haplotype in domestic animals is in Bronze Age Dutch cattle (Enk1 and Enk4). This Y1-basal signal persists in the Netherlands through the Iron Age (Bun2) and Roman era (Els1 and Otb1). The Y1-modern haplotype also first appears here in the Dutch domestic sample, Viz2. This bull has been dated to 1028 - 861 cal yrBP, a period of the Middle Ages that directly succeeds the Viking Age. The Viking Age was characterised by Scandinavian seafaring people undertaking large-scale trading, raiding and colonisation throughout coastal Europe and parts of North America. Previously, a connection between Viking conquest with the spread of Y1 across northwestern Europe and Scandinavia has been suggested due to the preponderance of Y1 in Icelandic cattle (Edwards et al. 2011).

Cattle were first imported to Iceland by Vikings approximately 1,000 years ago and modern breeds show greatest affinity to Nordic breeds (Gautason et al. 2020). The modern dispersal and low genetic diversity of Y1 could be consistent with a Viking dispersal.

However, several lines of evidence contradict a widespread dispersal of Y1-modern by Vikings. In this dataset, ancient domestic samples from British Mediaeval sites (Hungate and Coppergate) have not been genotyped to Y1-modern despite these sites postdating Viking occupation. Similarly, Mediaeval bulls from Sweden were not Y1 (Svensson and Götherström 2008), while domesticates from Viking settlements in Dublin had a larger affinity to modern cattle of Britain and Ireland than to Scandinavia (Edwards et al. 2003). Overall, this suggests that while Vikings may have managed Y1 bulls, they were not the agents of their current dispersal. Notably, the single modern Y1 sample not derived from a breed from northwestern Europe (an Iberian Pajuna animal) is the only animal not falling within the more recent northwest European expansion. TMRCA of this Pajuna animal and the northwest European breeds likely lived thousands of years (median node age = 3,579 yrBP, HPD 95% = 1,247 - 6,223 yrBP). Comparatively, the expansion of the northwest European Y1 occurred much later (median node age = 1,263 yrBP, HPD 95% = 345-2,708 yrBP). With the knowledge that a domestic basal Y1 existed since at least Bronze Age Netherlands, the most parsimonious explanation is that Y1 has been through a very recent bottleneck in northwest European breeds. This explains both why Y1 was not widespread in Middle Ages Britain or Scandinavia and why basal diversity persists outside of northwest Europe in Pajuna breeds. It is therefore likely that the low diversity and very recent common ancestor of Y1-modern in northwestern breeds is due to modern breeding practices propagating a narrow pool of male lineages.

Y2a recruitment and spread

The wide and sporadic occurrence of Y2a-basal in wild and domesticate bulls makes the establishment of Y2a difficult to discern. As discussed, Y2a-basal is a poorly defined group of haplotypes and may not be informative for understanding modern Y2a recruitment and dispersal until higher quality sequences are generated. However, in modern Y2a, geographic structure emerges with a clear Y2a subclade of African taurines (Y2a-Af). Such a sub-clade of Y2 had previously been reported in Sub-saharan taurine cattle (Pérez-Pardal et al. 2010). TMRCA of Y2a-Af and Y2a-Eu predates modern breed formation (median node age = 7,390 yrBP, HPD 95% = 2,455 - 15,522 yrBP). This divergence may be attributed to the dispersal of taurine cattle into Africa from the Near East, which occurred around 8,000 years ago (Hanotte et al. 2000). A single modern Near Eastern breed (South Anatolian Red) also sits along the branch of Y2a which suggests that the ancestors of modern Y2a may have a Near Eastern or

North African origin. More extensive sampling of North African and Near Eastern modern breeds may reveal more diversity within the modern Y2a haplogroup.

Y2a-Eu has a modern distribution across southern European breeds, which has led to a suggestion that it may be linked to the southern Mediterranean route of Neolithisation (Edwards et al. 2011). However, the node age of TMRCA of the Y2a-Eu clade indicates a relatively recent expansion (median node age = 2,686 yrBP, HPD 95% = 924 - 4,946 yrBP). This time period overlaps with the emergence of ancient Roman civilisation. At the height of its influence, the Roman state extended its control across much of Europe, North Africa and the Near East. Notably, the first appearance of the Y2a-Eu haplotype is also the Roman era Dutch sample, Bri1. The transition to Roman control was associated with a local increase of size in domesticated animals, in particular cattle (Colominas, Schlumbaum, and Saña 2014). Indeed, two Roman domestic samples in this dataset (Bri1 and CR129) were both incorrectly identified as aurochs due to their large stature. The increased size of Roman cattle has been postulated to be due to intensified animal husbandry practices or the importing of exotic animals, although these reasons are not mutually exclusive (Colominas, Schlumbaum, and Saña 2014). It is possible that Y2a-Eu bulls were imported from the Near East or North Africa and brought across the Roman Empire to improve local herds. As discussed in Chapter 7, the R haplotype in modern Italian cattle likely corresponds to Roman innovation of importing North African diversity into local domestic stocks. Similarly, exotic male diversity may have been imported from the fringes of the Roman state. While the R haplotype remains confined to sporadic occurrences in Italian breeds, Y2a-Eu has very effectively swept across most southern Europe breeds, reflecting the strength of founder effects in Y chromosomes compared to mitochondria in breed formation (Carolino and Gama 2008). While the modern distribution of Y2a-Eu and timing of the expansion of Y2a haplotype support an ancient Roman origin, sampling of southern Europe domesticate breeds from this era will be necessary to support this hypothesis.

Y2b recruitment and spread

As discussed, Y2b-east may represent the eastern ancestors of the original recruitment of aurochs in the Middle and Upper Euphrates Valley. Taurine cattle are believed to have spread from Central Asia to East Asia between 5,000 and 4,000 years ago (Flad et al. 2007). Indeed, TMRCA of Y2b-east has an age that corresponds to the dispersal of cattle into Central and Eastern Asia (median node age = 4,883 yrBP, HPD 95% = 1,906-9,405 yrBP). Currently, the only ancient Y2b-east is genotyped in a domesticate (Bisk1) from Biskupin, Poland - an Iron Age site occupied from the 8th Century BCE onward. While this may seem to contradict an eastern dispersal of Y2b-east, the occupation of Biskupin succeeds massive human migrations

from the Western Eurasian steppes into Europe during the third millennium BC (Haak et al. 2015). The expansion of these peoples across Eurasia has been proposed to be enabled by increased diarying of cattle, sheep and goats (Wilkin et al. 2021). It is possible that these migrating people brought their animals westward as they moved through Europe. The existence of Y2b-east in Buskupin may be a remnant of the movement of eastern ancestry.

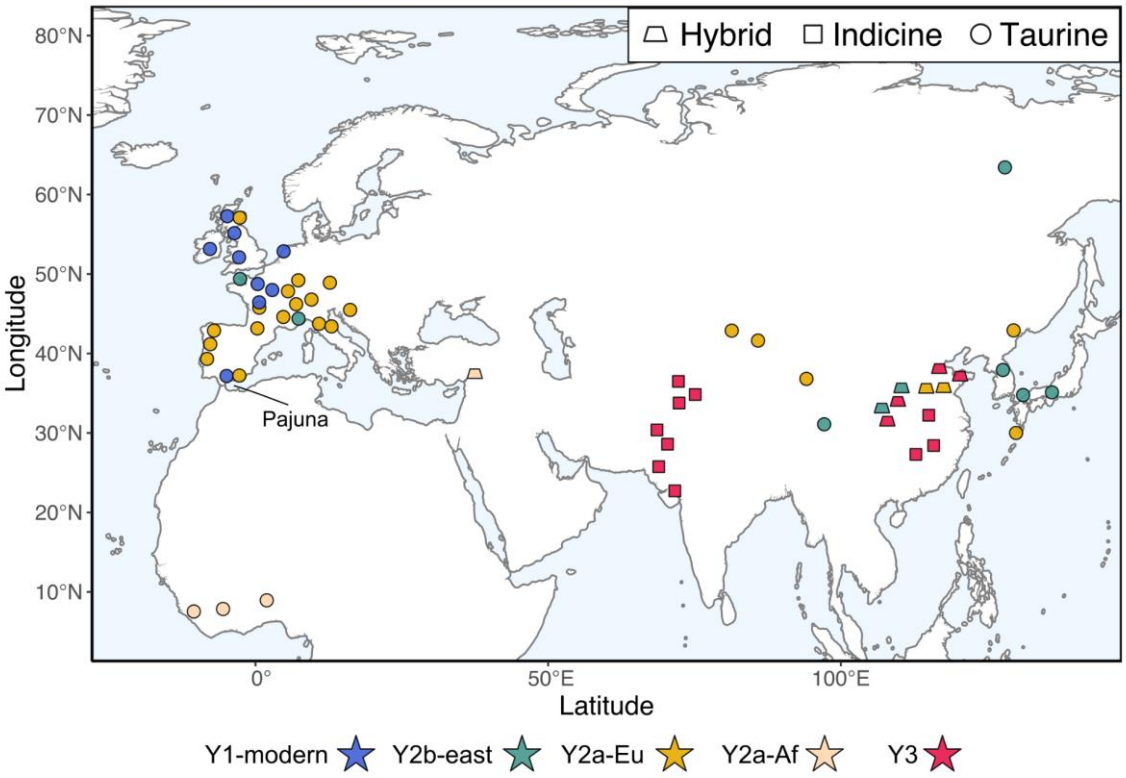


Figure 8.7 - Modern distribution of haplotypes.

9. Autosomal analysis of *Bos primigenius*

9.1. Introduction

The *Bos taurus* nuclear genome comprises nearly 3 billion bases pairs over 29 autosomes, presenting a wealth of information for genetic analysis (Snelling et al. 2007). Phylogenetic inferences made with autosomal DNA are robust due to the availability of millions of independently evolving loci to moderate individual instances of incomplete lineage sorting (Xu and Yang 2016). In contrast, uniparental loci provide a more limited view of ancestry and are vulnerable to stochastic drift that can obscure true species trees (Ballard and Whitlock 2004). Additionally, autosomes are diploid and recombining which preserves multiple ancestries over many generations. Studying patterns of these loci can inform on demographic events. For example, regions of homozygosity along the genome can be the result of inbreeding events whereby both haplotypes were inherited from the same ancestor sequence. Driven by recombination and mutation, the patterns of non-random inheritance of haplotypes caused by linkage disequilibrium (LD) offer extra dimension of information in autosomes. Reconstructing genealogies of all haplotypes from a single diploid genome can estimate a complete demographic history of a species, demonstrating the power of autosomal DNA (Li and Durbin 2011). As the decay of LD is dependent on generation time, periods of admixture can be dated based on the size distribution of preserved haplotypes (Hellenthal et al. 2014).

Autosomal analysis of cattle began with restriction fragment length polymorphism experiments that studied inheritance of traits for cattle breed improvement (Beckmann and Soller 1983). As PCR became ubiquitous in the 1990s, microsatellites studies rose in popularity and were widely used to study the diversity within and between cattle breeds. These studies revealed patterns of ancient admixture and dispersal based on allelic diversity of extant breeds, as well as building upon mitochondrial evidence of a putative domestication in Southwest Asia (Loftus et al. 1999; Cymbron et al. 2005; Loftus et al. 1999). While these studies began to illuminate the complexities of cattle domestication and dispersal, a perceived weakness was the difficulty in combining microsatellite data from different sources (Freeman et al. 2006). The first *Bos taurus* genome was published in 2009 using DNA from a Hereford cow (Zimin et al. 2009) and signalled a new era of cattle population genomics, improving the power to study selection and ancestry (Gautier, Laloë, and Moazami-Goudarzi 2010). Using ~50 thousand SNPs ascertained on 497 modern taurine and indicine breeds (BovineSNP50), a global genetic description of cattle breeds was presented (Bovine HapMap Consortium et al. 2009). Population analyses based on the genomic SNP data reiterated the three major groups of

cattle diversity first described with microsatellites: European taurine, African taurine and indicine cattle (MacHugh et al. 1997). Migratory routes of cattle from their centres of origins across the world were retraced with more detail and localised studies of cattle breeds abound (Gautier, Laloë, and Moazami-Goudarzi 2010; Decker et al. 2014; Karimi et al. 2016).

Reduction of sequencing cost saw later studies opting for whole genome sequencing rather than SNPs ascertained on curated datasets of modern breeds. This uncovered more subtle patterns of ancestry in modern cattle breeds, including delineating that Central and East Asian taurine breeds had a considerable amount of unique variation not present in European taurines (Chen et al. 2018). While these genomic studies greatly improved the understanding of extant cattle variation, a fundamental limitation remained; that of using modern genomes to infer ancient population dynamics. As population bottlenecks associated with breed formation were thought to obscure the dynamics of the domestication of *Bos primigenius*, direct evidence from the past was desirable (Bovine HapMap Consortium et al. 2009). This was addressed with the publication of 67 ancient nuclear genomes that provided proximal evidence of the origins of *Bos taurus* and confirmed that the centre of cattle domestication existed in Southwest Asia (Verdugo et al. 2019). Other aspects of cattle domestication that were observed in modern breeds, such as ancient indicine admixture, were also elucidated with time-stamped precision. While these studies improved the understanding of the history of domestic cattle, a question remained regarding the diversity of wild cattle. As domestic cattle were likely derived from a small initial stock, it's likely that domesticates only capture a fraction of wild diversity that existed in the early Holocene (Bollongino et al. 2012; Scheu et al. 2015). This was confirmed with the analysis of an approximately 6,750-year-old British aurochs bull nuclear genome, which outgrouped all extant European taurine breeds (Park et al. 2015). Furthermore, the authors were also able to show that it shared the most alleles with modern British and Irish breeds, suggesting that recruitment of local aurochs occurred during early cattle management. To date, only four aurochs whole genome sequences have been published from a limited temporal and geographic range (Park et al. 2015; Verdugo et al. 2019). Further sampling across Eurasia is necessary to characterise the genetic variation of *Bos primigenius* and influence on modern breeds.

9.2. Aims

1. To characterise the autosomal genetic structure of aurochs populations through time.
2. To detect introgressive signals of local wild variation in domesticated cattle.
3. To investigate Runs of Homozygosity in higher-coverage samples.

4. To model demography of aurochs population using PSMC.

9.3. Materials and methods

9.3.1. Materials

A total of 108 ancient *Bos* nuclear genomes are included in this study, of which 62 are newly presented here (Table 9.1). The mean genome coverage ranges from 0.01X to 22.47X, with an average coverage of 2.53. 48 genomes are above an average 1X coverage.

These genomes were supplemented by 69 modern *Bos* samples that were downloaded from NCBI, each representing a unique breed (Appendix Table 9.1). Additionally, one Water Buffalo (*Bubalus bubalis*) genome was included as an outgroup. Curation of modern genomes was performed by Dr Victoria Mullin.

9.3.2. Laboratory methods

Sample preparation is detailed in Section 2.1. DNA extraction and sequencing library construction are detailed in Section 2.2 and 2.3, respectively. Screening results for samples are described in Chapter 3.

9.3.3. Read-processing

After taxonomic identification (Chapter 3), trimmed sequencing reads were aligned to the *Bos taurus* reference genome (ARS-UCD1.2) using bwa 0.7.5 with relaxed parameters (Section 2.4.3.). The reference genome was edited to include a bovine Y chromosome (GCF_000003205.7). Post-alignment read processing was performed according to Section 2.4.3.

9.3.4. Variant discovery

Variant discovery using modern and ancient *Bos* genomes was performed by Dr Victoria Mullin. All ancient and modern *Bos* genomes (*Bos taurus, Bos indicus, Bos primigenius*) above an average of 1X coverage were included in the variant calling dataset. With this dataset, haploid SNP calls were called on autosomes with ANGSD (version: 0.936) by sampling a random base called using the SAMtools genotype likelihood algorithm (-doHaploCall 1 -doCounts 1 -GL 1) (Korneliussen, Albrechtsen, and Nielsen 2014). Sites were subsequently passed through several filtering steps for quality control, including removing transitions (-rmTrans 1 -SNP_pval 1e-6 -rmTriallelic 1e-4 -C50 -uniqueOnly 1 -remove_bads 1 -trim 4 -

minMapQ 30 -minQ 25). The minor (ancestral) allele was defined using a consensus fasta of Water Buffalo aligned to the *Bos taurus* reference genome, ARS-UCD1.2 (-doMaf -doMajorMinor 5).

The resulting genotype calls were filtered for coverage (1% and 99% quantiles of the position-wise coverage distribution) using an in-house script and converted to PLINK format using ANGSD haploToPlink. This PLINK file was then filtered for sites that were present in at least 80% of individuals. This SNP set was then used as a site file (-site) to individually call variants in all samples below a mean coverage of 1X and the Water Buffalo outgroup, as above. Each low coverage was then filtered for SNPs that were not defined as major or minor above. All samples were merged using PLINK. A minor allele filter of 2% was applied (-maf 0.2) and linkage disequilibrium pruning was performed (--indep-pairwise 50 5 0.2). This resulted in a total of 3,255,954 filtered SNPs.

9.3.5. Identity-By-State

Identity-By-State (IBS) matrices were computed using ANGSD (version: 0.936) for all samples above 0.01X by sampling a random base (-makeMatrix 1 -doIBS 1) at the filtered SNP sites (-doMajorMinor 3 -sites) (Korneliussen, Albrechtsen, and Nielsen 2014). Quality filters were used, as above (-SNP_pval 1e-6 -rmTriallelic 1e-4 -C50 -uniqueOnly 1 -remove_bads 1 -trim 4 -minMapQ 30 -minQ 25). The resulting IBS matrix was then loaded into R and a Neighbour-Joining tree was computed using the R package ape (Version 5.7) (Emmanuel Paradis, Claude, and Strimmer 2004). The Water Buffalo genome specified as the outgroup using the root() function.

9.3.6. Principal component analysis

PCA was computed using ANGSD (version: 0.936) for all samples above 0.1X using genotype likelihoods using the GATK model (-GL 2) (Korneliussen, Albrechtsen, and Nielsen 2014). Sites were supplied from the SNP calling set, above (-domajorminor 3 -sites). All quality filtering was specified as above (including a MAF of 2% and missingness of 20%) and the file was dumped to a Beagle likelihood file (-doGlf 2). The Beagle file was then used to compute a covariance matrix using PCAngsd v1.0 (Meisner and Albrechtsen 2018).

9.3.7. D Statistics

D statistics were computed on the pseudohaploid dataset using ADMIXTOOLS2 (https://github.com/uqrmaie1/admixtools) using the Water Buffalo as H4/Outgroup. All available SNPs were used in each individual D statistic computed (allsnps = FALSE).

9.3.8. ADMIXTURE

Unsupervised model-based clustering was performed using ADMIXTURE (Version 1.3.0) (Alexander, Novembre, and Lange 2009). Based on methods in (Verdugo et al. 2019), all samples below 1X coverage were first removed from the PLINK dataset. A missingness filter of 5% was then applied (-geno 0.95) to leave 339,482 sites. With this dataset, ADMIXTURE was run with a random seed (-s 12345) across a range of ancestral components (K=1-10). Each run was repeated 10 times and the cross-validation error was inspected (--cv).

Samples between 0.3-1X coverage were merged to the filtered PLINK dataset and ADMIXTURE was run for each sample individually. For each of these samples ADMIXTURE was run with a random seed (-s 12345) with K=4 ancestral components.

9.3.9. Runs of Homozygosity

Runs of homozygosity were calculated for ancient samples above 6X coverage (pre-mapping quality filter) using ROHan (Renaud et al. 2019). All samples were downsampled to a mean coverage of 6X using SAMtools view -s to control for coverage effects. Ancient damage was accounted for using bam2prof. Briefly, damage patterns were calculated for 1,000,000 reads at a minimum fragment length cut-off ranging from 30 to 51 bp (-length) at a minimum quality cut-off ranging from 0 to 30 (-minq) for the 3p and 5p end. Damage rates were inspected and an appropriate filter was selected when substitution rates levelled off. This was repeated for each ancient BAM file and specified in the ROH computation (--deam5p, --deam3p). ROHan was run on a selection of modern samples without specification of ancient damage. All samples were run with --rohmu 2e-5 for consistency.

9.3.10. MSMC

MSMC2 was used to produce PSMC' curves for all ancient samples above an average genome coverage of 12X and a selection of modern genomes (Schiffels and Wang 2020). Genotypes were called as per MSMC2 instructions using samtools and bcftools (version 1.9) to call to single site variants only (Li et al. 2009). A mappability filter based on Heng Li's SNPable filter was also produced for each chromosome, filtering for sites that scored highly for unique

mapping (score of 3) (http://lh3lh3.users.sourceforge.net/snpable.shtml). The genotype calls were converted to VCF format with coverage and mappability filters applied using bamCaller.py from the msmc-tools repository. VCFs were phased individually using Beagle 5.2 (version: 28Jun21.220) with no imputation (impute=false) and effective population size set to 20,000 (ne=20000) (Browning et al. 2021).

In ancient genomes, it was observed that the *Ne* was greatly inflated in the final time segment. As this pattern was not observed in modern samples, it was likely caused by ancient damage creating false patterns of recombination (Figure 9.2). In ancient populations genetic studies, transition variants are often filtered out due to confounding factors of aDNA damage. However, simply setting transition variants as monomorphic would result in large underestimation of divergences of haplotypes, skewing the inferences of *Ne*. To this end, PSMC' curves were computed with and without any transitions for 10 modern genomes (Figure 9.1). The average curve area of the transitions removed was 0.303663313 that of the full PSMC'. This is roughly in line with the rate to transition:transversion rate in humans and what has been previously published for ancient wolf genomes (Bergström et al. 2022). Scaling the transitions removed PSMC' by this transversions-only mutation rate resulted in *Ne* curves that are comparable to the full PSMC' but without the inflation of the final time segment.

Therefore to produce the final PSMC' curves, transitions were removed for all VCFs (moderns and ancients). MSMC2 was run on each individual genome with 20 bootstraps and the time segment pattern: --timeSegmentPattern 25*1+1*2+1*3. Curves were scaled with a mutation rate of 1.26e-8 and generation time of 6, as in (Chen et al. 2018). All curves were then scaled by the transversions-only mutation rate of 0.303663313 and ancients were scaled by age.

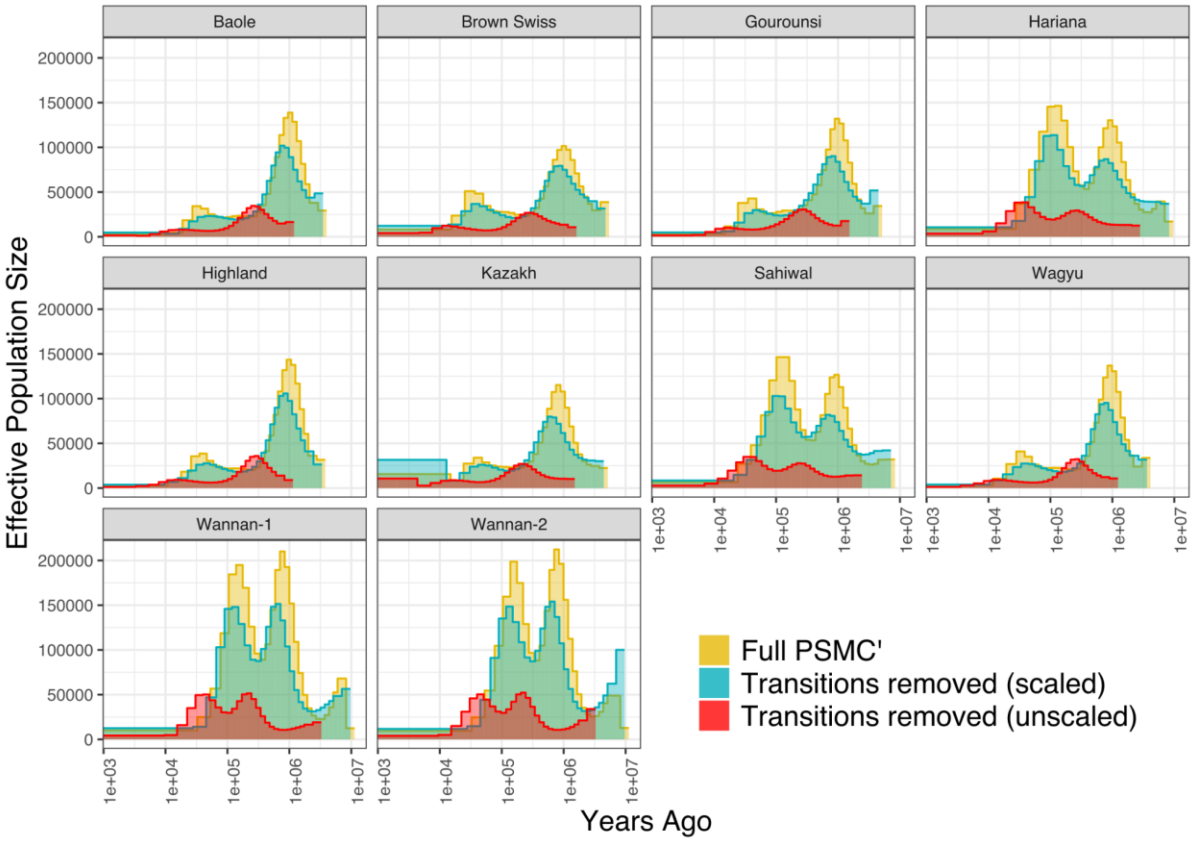


Figure 9.1 - Calculating the transversions-only mutation rate using 10 modern *Bos genomes.*

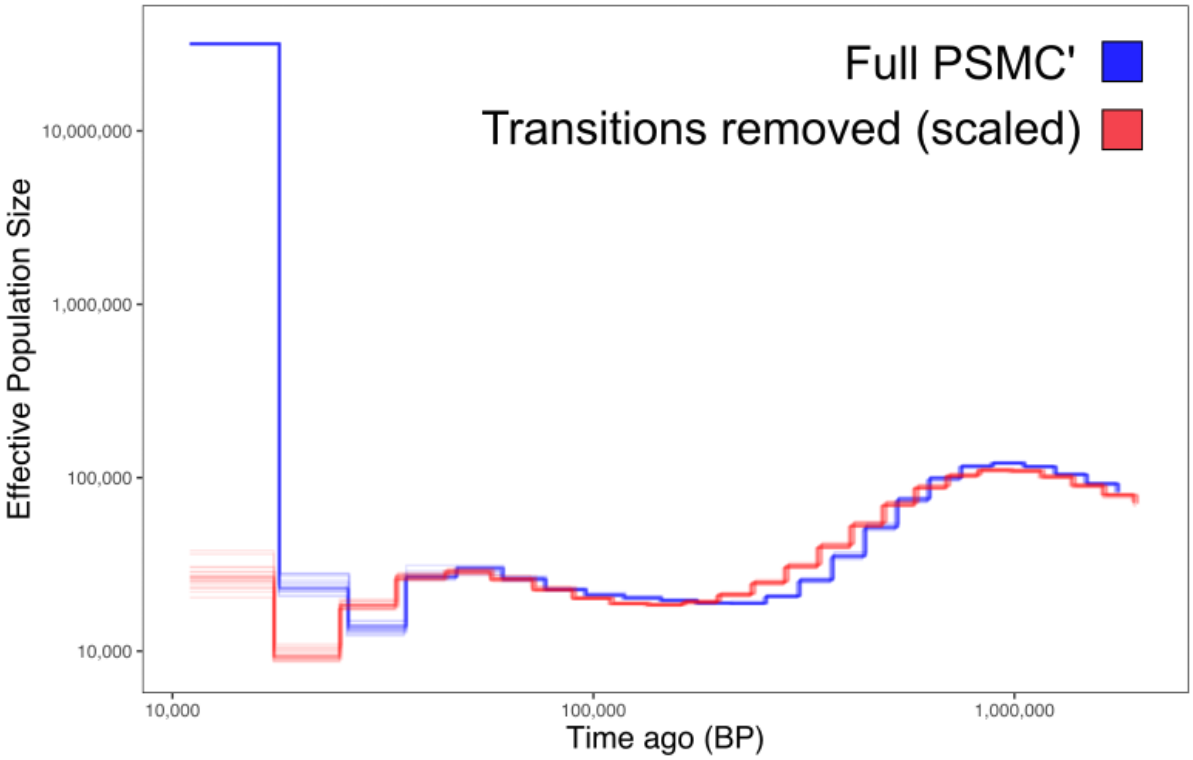


Figure 9.2 - Ancient genome Bed3 displaying transversions-only mutation rate.

Table 9.1 - Summary of the ancient *Bos* genomes used in this study. Context is reproduced as received from the curator unless the sample is radiocarbon dated. * = sample dated by context; ^ sample radiocarbon dated indirectly. AH=Ashleigh Haruda, AS = Amelie Scheu, BL = Bastien Llamas, CR = Conor Rossi, LF = Laurent Frantz, KM = Kieren Mitchell, MS = Mikkel Sinding, MV = Marta Verdugo, VEM = Victoria E. Mullin.

Sample ID	Country	Region/Site	Context	Age	Average coverage	Laboratory Work
Ace1	Turkey	Acemhöyük	Bronze Age	4923 - 4323 cal yrBP^	0.11	(Verdugo et al. 2019)
Baikal1	Russia	Lake Baikal	Pre-domestic contact aurochs	12015 - 12529 cal yrBP	14.00	MV
Bar1	Italy	Mantova	Pre-domestic contact aurochs	13246 - 13005 cal yrBP	0.21	MS/CR
Bed3	Germany	Bedburg-Königshoven	Pre-domestic contact aurochs	11875 - 11399 cal yrBP	18.49	AS
Bed4	Germany	Bedburg-Königshoven	Pre-domestic contact aurochs	11949 - 11476 cal yrBP	2.09	AS
Bel1	Serbia	Belovode-Veliko	Neolithic	7050/6925 - 6900/6800 cal yrBP^	0.09	(Verdugo et al. 2019)
Bel2	Serbia	Belovode-Veliko	Neolithic	7050/6925 - 6900/6800 cal yrBP^	0.13	(Verdugo et al. 2019)
Belarus1	Belarus	Brest district	Post-domestic contact aurochs	4476 - 4173 cal yrBP	1.78	MV
Bes1	Iraqi Kurdistan	Bestansur	Iron Age	2934 - 2632 yrBP*	2.12	(Verdugo et al. 2019)
Bes2	Iraqi Kurdistan	Bestansur	Mediaeval	624 - 731 cal yrBP	14.00	(Verdugo et al. 2019)
Bla1	Serbia	Blagotin	Neolithic	8200 - 8000 yrBP*	0.02	(Verdugo et al. 2019)
Blec1	Bulgaria	Vratsa	Post-domestic contact aurochs	6258 - 6006 cal yrBP^	0.99	CR
Bor1	Kazakhstan	Borly	Pre-domestic contact aurochs	6740 - 6521 cal yrBP	0.30	CR
Bor2	Kazakhstan	Borly	Pre-domestic contact aurochs	4913 - 5099 cal yrBP	0.19	CR
Bor3	Kazakstan	Borly	Pre-domestic contact aurochs	5115 - 4925 cal yrBP	0.71	CR
Bub1	Serbia	Bubanj	Neolithic	8000 - 7600 yrBP*	0.43	(Verdugo et al. 2019)
Bul1	Uzbekistan	Tilla Bulak	Bronze Age	4000 - 3700 cal yrBP^	0.05	(Verdugo et al. 2019)
Cas1	Israel	Horvat Castra	Byzantine	1676 - 1362 yrBP*	0.07	(Verdugo et al. 2019)

Ch22	Turkey	Çatalhöyük	Post-domestic contact aurochs	5750 - 5600 cal yrBP^	0.30	(Verdugo et al. 2019)
Che1	Bulgaria	Chernomorets	Neolithic	7123 - 6818 cal yrBP	0.41	MS/CR
Cpc	England	Carsington Pasture Cave	Pre-domestic contact aurochs	6884 - 6741 cal yrBP	6.00	(Park et al. 2015)
Fal7	Germany	Falkenwalde	Post-domestic contact aurochs	1000 yrBP*	1.53	AS
Far1	Iran	Miyanroud	Neolithic	8000 - 7500 yrBP*	0.03	(Verdugo et al. 2019)
Fif1	Scotland	Newburgh, Fife	Pre-domestic contact aurochs	10825 - 10496 cal yrBP	0.15	MS/CR
Gal1	Spain	Galicia	Pre-domestic contact aurochs	9548 - 9113 cal yrBP	17.08	MS/CR
Gal2	Spain	Galicia	Pre-domestic contact aurochs	9481 - 9100 cal yrBP	0.85	MS/CR
Gal3	Spain	Galicia	Pre-domestic contact aurochs	9538 - 9109 cal yrBP	2.93	MS/CR
Gallo1	Scotland	Galloway	Post-domestic contact aurochs	3764 - 3479 cal yrBP	2.12	MS/CR
Gyu1	Armenia	Hasanlu	Pre-domestic contact aurochs	14097 - 13863 cal yrBP	18.18	MV/CR
Gyu2	Armenia	Hasanlu	Post-domestic contact aurochs	7394 - 7239 cal yrBP	1.70	(Verdugo et al. 2019)
Has1	Iran	Hasanlu	Bronze Age	3900 - 3600 yrBP*	0.03	(Verdugo et al. 2019)
Has3	Iran	Hasanlu	Bronze Age	4470 cal BP^	1.20	(Verdugo et al. 2019)
Has4	Iran	Hasanlu	Iron Age	3050 - 2800 yrBP*	1.40	(Verdugo et al. 2019)
Has5	Iran	Hasanlu	Achaemenid / Seleuco-Parthian	2500 - 2275 BP*	0.03	(Verdugo et al. 2019)
Hjo1	Denmark	Hjørring, Tofte Bæk	Pre-domestic contact aurochs	8245 - 8053 cal yrBP	1.14	MS/CR
Horn-10537	Poland	Hunting Horn of last aurochs bull	Post-domestic contact aurochs	400 yrBP*	1.06	MS
Horn-10543	Denmark	Lykkesholm, Fyn	Post-domestic contact aurochs	600 yrBP*	0.35	MS
Horn-382	Norway	The royal Horn of Norway	Post-domestic contact aurochs	700 yrBP*	0.09	MS

Horn-617	Unknown	Unknown provenance	Post-domestic contact aurochs	600 yrBP*	0.04	MS
Horn-CMLIX	Norway	Bjørnsund, Norway	Post-domestic contact aurochs	600 yrBP*	0.42	MS
Horn11570	Germany	Firemen's Guild, Büsum, Holstein	Post-domestic contact aurochs	400 yrBP*	0.08	MS
Hxh2	Germany	Herxheim	Post-domestic contact aurochs	5000 yrBP*	22.47	AS
Kalisz1	Poland	Kalisz	Post-domestic contact aurochs	4487 - 4229 cal yrBP	1.53	CR
Kalisz2	Poland	Kalisz	Post-domestic contact aurochs	5813 - 5668 cal yrBP	2.60	(Verdugo et al. 2019)
Kalisz3	Poland	Kalisz	Post-domestic contact aurochs	4477 - 4168 cal yrBP	1.79	MV/CR
Kaz1	Georgia	Tamara Fort	Mediaeval	1000 - 0 yrBP*	3.12	(Verdugo et al. 2019)
Kaz2	Georgia	Tamara Fort	Mediaeval	1200 - 900 yrBP*	0.29	(Verdugo et al. 2019)
Kaz3	Georgia	Tamara Fort	Mediaeval	1200 - 1100 yrBP*	0.46	(Verdugo et al. 2019)
Kaz4	Georgia	Tamara Fort	Mediaeval	1200 - 1000 yrBP*	0.47	(Verdugo et al. 2019)
Kaz5	Georgia	Tamara Fort	Mediaeval	1349 - 1234 yrBP*	0.02	(Verdugo et al. 2019)
Kho1	Iran	Nishapur Qohandez	Parthes-Islamic	2300 - 700 yrBP*	0.20	(Verdugo et al. 2019)
Kok1	Uzbekistan	Kok Tepe	Iron Age	5000 - 3500 yrBP*	0.48	MV/CR
Lil1	Denmark	Randers, Bønnerup Strand	Post-domestic contact aurochs	4484 - 4228 cal yrBP	5.59	MS/CR
Lun1	Sweden	Skåne	Pre-domestic contact aurochs	9744-9525 / 9045-8656 cal yrBP^	0.22	CR
Lun2	Sweden	Skåne	Pre-domestic contact aurochs	9744-9525 / 9045-8656 cal yrBP^	0.07	CR
Mar1	Iran	Maral Tappeh	Neolithic	7000 yrBP*	0.04	(Verdugo et al. 2019)
Men1	Turkey	Menteşe	Neolithic	8000 - 7500 yrBP*	2.97	(Verdugo et al. 2019)
Men2	Turkey	Menteşe	Neolithic	7998 - 7843 cal yrBP	1.88	(Verdugo et al. 2019)
NVL1	Russia	Altai Krai	Pre-domestic contact aurochs	5300 - 5000 yrBP*	0.01	KD

NVL3	Russia	Altai Krai	Pre-domestic contact aurochs	5300 - 5000 yrBP*	0.06	KD
Pad1	Italy	Padova	Pre-domestic contact aurochs	12267 - 11844 cal yrBP	0.16	CR
Pal1	Italy	Central Italy	Pre-domestic contact aurochs	16287 - 15849 cal yrBP	2.56	CR
Plo1	Serbia	Pločnik	Neolithic	7275/7125 - 7100/7000 cal yrBP^	0.13	(Verdugo et al. 2019)
Plo2	Serbia	Pločnik	Neolithic	7025/6950 - 6850 - 6650 cal yrBP^	0.35	(Verdugo et al. 2019)
Plo3	Serbia	Pločnik	Neolithic	7100/7000 - 7025/6950 cal yrBP^	2.66	(Verdugo et al. 2019)
Plo4	Serbia	Pločnik	Neolithic	7100/7000 - 7025/6950 cal yrBP^	2.74	(Verdugo et al. 2019)
Plo5	Serbia	Pločnik	Neolithic	6850/6650 - 6600/6500 cal yrBP^	0.08	(Verdugo et al. 2019)
Plo6	Serbia	Pločnik	Neolithic	7275/7125 - 7100/7000 cal yrBP^	0.12	(Verdugo et al. 2019)
Plo7	Serbia	Pločnik	Neolithic	7100/7000 - 7025/6950 cal yrBP^	0.11	(Verdugo et al. 2019)
Plo8	Serbia	Pločnik	Neolithic	7100/7000 - 7025/6950 cal yrBP^	0.20	(Verdugo et al. 2019)
Qaz1	Iran	Tepe Shizar	Bronze Age	4905 - 4659 cal yrBP	1.70	(Verdugo et al. 2019)
Rhi1	Germany	Upper Rhine Valley	Pre-domestic contact aurochs	47,000 cal yrBP+	9.54	CR
Rhi2	Germany	Upper Rhine Valley	Pre-domestic contact aurochs	47009 - 45312 cal yrBP	2.12	CR
Rhi3	Germany	Upper Rhine Valley	Pre-domestic contact aurochs	Pleistocene*	3.14	CR
Rhi4	Germany	Upper Rhine Valley	Pre-domestic contact aurochs	Pleistocene*	0.18	CR
Rhi5	Germany	Upper Rhine Valley	Pre-domestic contact aurochs	Pleistocene*	0.16	CR
Rhi6	Germany	Upper Rhine Valley	Pre-domestic contact aurochs	Pleistocene*	0.12	CR
Rhi7	Germany	Upper Rhine Valley	Pre-domestic contact aurochs	Pleistocene*	0.07	CR
Rhi8	Germany	Upper Rhine Valley	Pre-domestic contact aurochs	Pleistocene*	0.03	CR
ROS002	Kazakhstan	Upper Rhine Valley	Pre-domestic contact aurochs	5116 - 4926 cal yrBP	0.09	KD/CR

Sac3	Iran	Tapeh-Sang-e-Caxmaq	Neolithic	7373 cal yrBP^	0.02	(Verdugo et al. 2019)
Shi1	Kazakhstan	Shiderty River	Bronze Age	3983 - 3772 cal yrBP	0.46	CR
Shimao	China	Shimao	Neolithic	3975 - 3835 cal yrBP	2.73	(Chen et al. 2018)
Siv1	Bulgaria	Vratsa	Post-domestic contact aurochs	6258 - 6006 cal yrBP	9.03	MS/CR
Ska1	Sweden	Skåne	Pre-domestic contact aurochs	9610 - 9207 cal yrBP	1.33	MS/CR
Ska3	Sweden	Skåne	Pre-domestic contact aurochs	9731 - 9508 cal yrBP	2.95	MS/CR
Stu1	Serbia	Stubline	Neolithic	6800 - 6500 cal BP^	0.15	(Verdugo et al. 2019)
Sub1	Turkey	Suberde	Neolithic	8221 - 8024 cal BP	13.50	(Verdugo et al. 2019)
Tango1	Former Soviet Union	Unknown	Post-domestic contact aurochs	4142 - 3914 cal yrBP	1.78	MV
Tango2	Former Soviet Union	Unknown	Post-domestic contact aurochs	4599 - 4379 cal yrBP	2.09	MV
Th7	Morocco	Taghit Haddouch	Pre-domestic contact aurochs	9073 - 8730 cal yrBP	0.07	(Verdugo et al. 2019)
Tmq3	Israel	Tel Miqne-Ekron	Iron Age	3000 - 2925 BP*	0.02	(Verdugo et al. 2019)
Tmq4	Israel	Tel Miqne-Ekron	Bronze Age-Iron Age	3400 - 3200 yrBP*	0.02	(Verdugo et al. 2019)
Ton1	Denmark	Jejsing	Post-domestic contact aurochs	3279 - 3074 cal yrBP	2.20	CR
Tri1	Germany	Trier-Brüderkrankenhaus	Pre-domestic contact aurochs	10291 - 9976 cal yrBP	10.00	AS
Tsa1	Israel	Tel es-Safi	Iron Age	2925 - 2720 BP*	0.97	(Verdugo et al. 2019)
Tsa3	Israel	Tel es-Safi	Iron Age	2925 - 2720 BP*	0.02	(Verdugo et al. 2019)
Tula1	Russia	Tula district	Pre-domestic contact aurochs	40746 - 42460 cal yrBP	3.51	MV/CR
Tyq1	Israel	Tel Yoqneam	Hellenistic	2332 - 2167 yrBP*	0.06	(Verdugo et al. 2019)
Ura1	Kazakhstan	Uralsk	Unknown	Unknown	0.01	CR
Uzz1	Italy	Sicily	Pre-domestic contact aurochs	7593 - 7480 cal yrBP	2.00	CR

Var1	Kazakhstan	Varfolomeevka	Pre-domestic contact aurochs	7641 - 7500 cal yrBP	4.55	CR
Var2	Kazakhstan	Varfolomeevka	Pre-domestic contact aurochs	7486 - 7238 cal yrBP	9.00	CR
VEM212	England	Norton Bottoms	Pre-domestic contact aurochs	MIS 7 (by stratigraphy)	0.02	VEM
Yen1	Russia	Yenisei River	Pre-domestic contact aurochs	14193 - 14019 cal yrBP	14.90	KM/BL
Yer1	Uzbekistan	Yerqorqan	Achaemenid	2100 yrBP*	1.68	MV/CR
YoA	England	Yorkshire	Post-domestic contact aurochs	4909 - 4658 cal yrBP	6.53	AH/LF/CR
Zah1	Israel	Tel Zahara	Islamic Era	2063 - 1670 yrBP*	0.03	(Verdugo et al. 2019)

9.4. Results

9.4.1. *Identity-By-State*

A Neighbour-Joining tree computed from IBS matrices revealed several groupings of *Bos* samples based on geographic and temporal identity (Figure 9.3-5). These serve as useful discussion points to begin to characterise the diversity and genetic flux of *Bos*. Group 1 comprises European aurochs dating from over 47,000 years ago (Rhi1) to among the last confirmed aurochs that died approximately 400 years ago (Horn-CMLIX), providing evidence of a degree of continuity of ancestry in Europe through the Pleistocene/Holocene transition. Within this clade, four groups of animals have been highlighted based on their geographic and temporal identity: **N**orthern, **I**berian and **C**entral **H**olocene **E**urope (hereafter referred to as NICHE) aurochs, Italian aurochs, Balkan aurochs and a group of Pleistocene Rhine Valley aurochs. One aurochs is of unknown geographical origin outgroups the group (Tango1). The oldest genomes within the NICHE aurochs clade are German, Scandinavian and Iberian genomes dating approximately to the Holocene transition and form an inner clade. This clade of older genomes is outgrouped by post-domestic contact animals. A subclade of the youngest samples (Fal7 and the three Horn genomes) outgroup all other NICHE genomes.

Group 2 consists of genomes east of the meridian 38° east. This “East Eurasian” group includes both domestic and wild animals. Within this group, a clade of Pleistocene aurochs emerges despite their sampling from a large geographical range; from Tula in Eastern Europe (Tula1) to Lake Baikal in Siberia (Baikal1), and as far south as Gyumri in the Caucasus (Gyu1). The temporal range is similarly extensive; the oldest sample (Tual1) is approximately 41,500 years old while the youngest (Baikal1) just precedes the Holocene transition. Holocene Central Asian aurochs also populate this Group, outgrouping the core Pleistocene clade in a stepping stone manner. Another aurochs of unknown geographical origin (Tango2) similarly outgroups the core Pleistocene clade. Finally, a separate clade in this Group comprises a domestic Kazakhstan Bronze Age animal (Shi1) and an Kazakhstan aurochs (Ura1).

Group 3 is populated mostly by African animals. All modern West African taurine breeds cluster here. A single Iron Age Levantine domesticate from Tell es-Safi (Tsa3) forms a clade with these modern breeds. Notably, other samples from Tell es-Safi do not fall within this Group, instead grouping with hybrid animals in Group 7. Finally, outgrouping these taurine domesticates is an approximately 9,000 year old aurochs from Morocco (Th7).

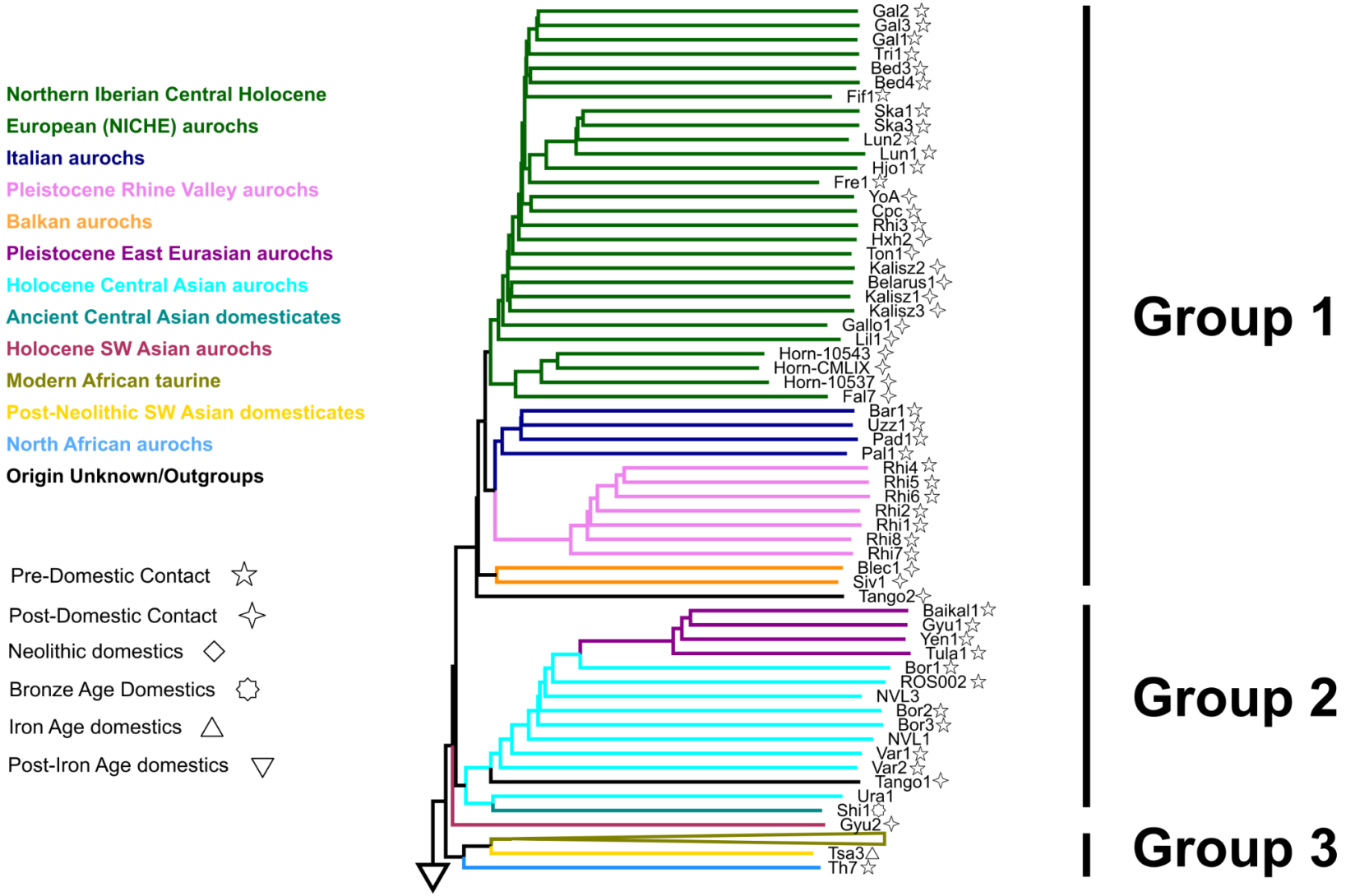


Figure 9.3 - Neighbour-Joining Tree of Identity-By-State Matrix computed with modern and ancient *Bos* genomes displaying clustering of Groups 1-3.

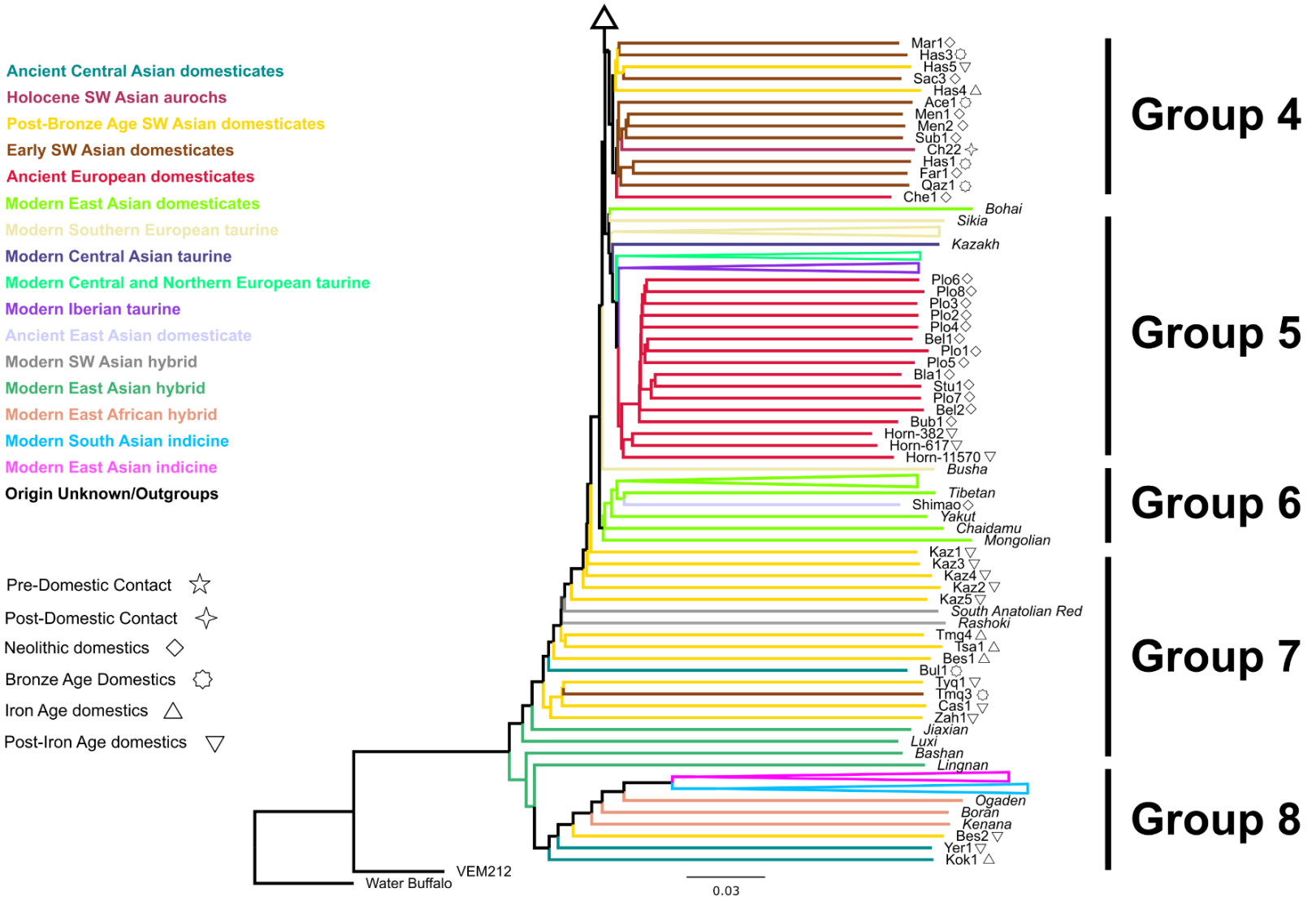
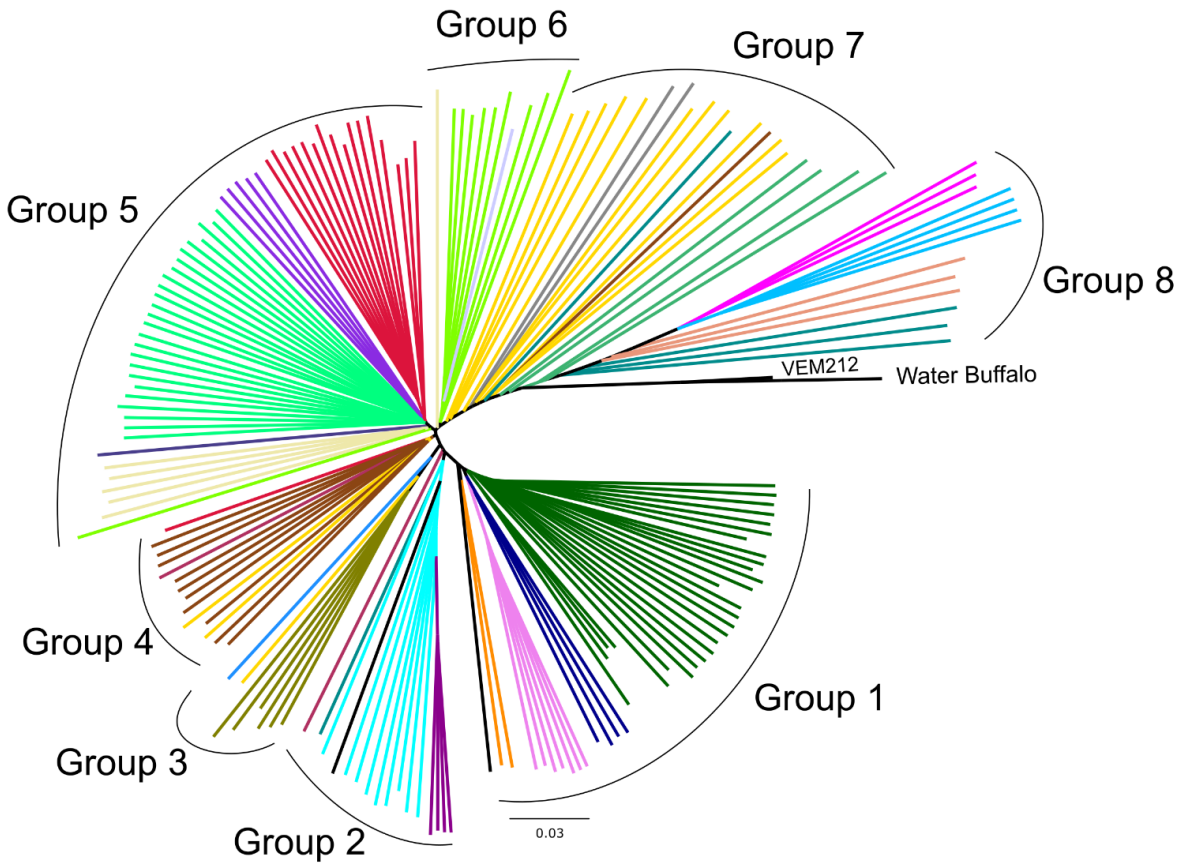


Figure 9.4 - Neighbour-Joining Tree of Identity-By-State Matrix computed with modern and ancient *Bos* genomes displaying clustering of Groups 4-8.

Group 4 comprises the earliest domesticated taurine cattle in this dataset. Here, all Neolithic Anatolian (Sub1, Men1, Men2) and Iranian (SAC3, Far1, Mar1) animals cluster. An animal designated as an aurochs from the Neolithic Anatolian site of Çatalhöyük (Ch22) also clusters here. Bronze Age Iranian (Mar1, Qaz1, Has1, Has3) and Anatolian (Ace1) animals are placed in this Group. Additionally, two post-Bronze Age Iranian animals appear here (Has4 and Has5). The only non-Southwest Asian sample in this group is the Neolithic Bulgarian domesticate, Che1.

Group 5 is a large group consisting almost entirely of ancient and modern European domesticates. Geographic structuring of modern European breeds is evident, with Southern European breeds, Central and Northern European breeds and Iberian breeds forming discrete clades. Two non-European breeds also cluster in this group; the Central Asian Kazakh breed and a Chinese Bohai breed. These modern clusters outgroup a clade of ancient European domesticates. These animals are from Neolithic Southeast Europe except for three historical drinking horn samples that were designated as aurochs. Group 6 comprises modern East Asian taurines and a single ancient Chinese Neolithic domestic (Shimao). Three East Asian taurine breeds (Yakut, Chaidamu, Mongolian) outgroup this clade.

Group 7 is a diverse group of hybrid (taurine x indicine) animals that fall along the branch leading to indicine animals. Ancient domesticates from a range of cultural periods populate this branch including Bronze Age Uzbekistan (Bul1) and Levant (Tmq3), Iron Age Levant (Tmq4, Tsa1) and Iraq (Bes1), and Post-Iron Age Mediaeval Georgia (Kaz1, Kaz2, Kaz3, Kaz4, Kaz5) and Levant (Tyq1, Cas1, Zah1). Two modern Southwest Asian breeds also sit along this branch. Group 8 consists of animals that may be described as indicine or hybrid animals. Ancient Iron Age (Kok1) and Post-Iron Age (Yer1, Bes2) domesticates and East African hybrid breeds outgroup the main indicine clusters. Finally, two indicine clades emerge based on geographic origin (South and East Asian). The tree was rooted by the outgroup Water Buffalo genome. Branching from this outgroup root is the oldest sample in this dataset, a ~200,000 year old British aurochs (VEM212).



NICHE aurochs
Italian aurochs
Pleistocene Rhine Valley aurochs
Balkan aurochs
Pleistocene East Eurasian aurochs
Holocene Central Asian aurochs
Ancient Central Asian domesticates
Holocene SW Asian aurochs
Modern African taurine
Post-Neolithic SW Asian domesticates
North African aurochs
Neolithic SW Asian domesticates
Ancient European domesticates

Modern East Asian domesticates
Modern Southern European taurine
Modern Central Asian taurine
Modern Central and Northern European taurine
Modern Iberian taurine
Ancient East Asian domesticate
Modern SW Asian hybrid
Modern East Asian hybrid
Modern East African hybrid
Modern South Asian indicine
Modern East Asian indicine
Origin Unknown/Outgroups

Figure 9.5 - Neighbour-Joining Tree of Identity-By-State Matrix computed with modern and ancient *Bos* genomes displaying clustering of all Groups.

A second IBS matrix was computed using only genomes of animals that were determined to be best representatives of pre-Holocene populations due to their age and average genome coverage. The NJ tree rooted by the Water Buffalo outgroup is displayed in Figure 9.6 with a heat map displaying the pairwise IBS distance.

The previously described *Bos primigenius primigenius* groups are depicted in geographic clades displaying low pairwise distance between members. Within the European clade, the Pleistocene aurochs (Rhi2, Rhi1) outgroups all other European diversity (Hxh2, Gal1, Gal3, Tri1, Ska3, Bed4, Bed3, Rhi3, Uzz1, Pal1). A group of three Neolithic Anatolian domesticates (Sub1, Men1, Men2) then form a sister clade to all European genomes. Outgrouping all European and Southwest Asian diversity is Th7, a ~9 ,000 year old sole representative of *Bos primigenius mauritanicus*. Finally, a clade of East Eurasian aurochs (Tula1, Gyu1, Baikal1, Yen1) re-emerges within *Bos primigenius* variation. These samples display very low pairwise IBS distance to each other. Excluding other East Eurasian Asian genomes, the oldest sample in this clade, Tula1, displays the lowest pairwise IBS distance with Th7 and the Rhine samples.

Structure is once again observed within the indicine breeds based on providence: South Asian (Gir, Tharparker, Sahiwel, Hariana) and East Asia (Jian, Jiangxi, Wannan). All *Bos primigenius* have a high pairwise IBS distance to the indicine breeds. Low IBS distance is observed within the indicine breeds. Finally, the tree is rooted by a Water Buffalo outgroup. Also outgrouping the *Bos primigenius* is the ~200,000 year old aurochs, VEM212. After the Water Buffalo genome, VEM212 had the lowest pairwise IBS distance with the Rhine samples, followed by Tula1.

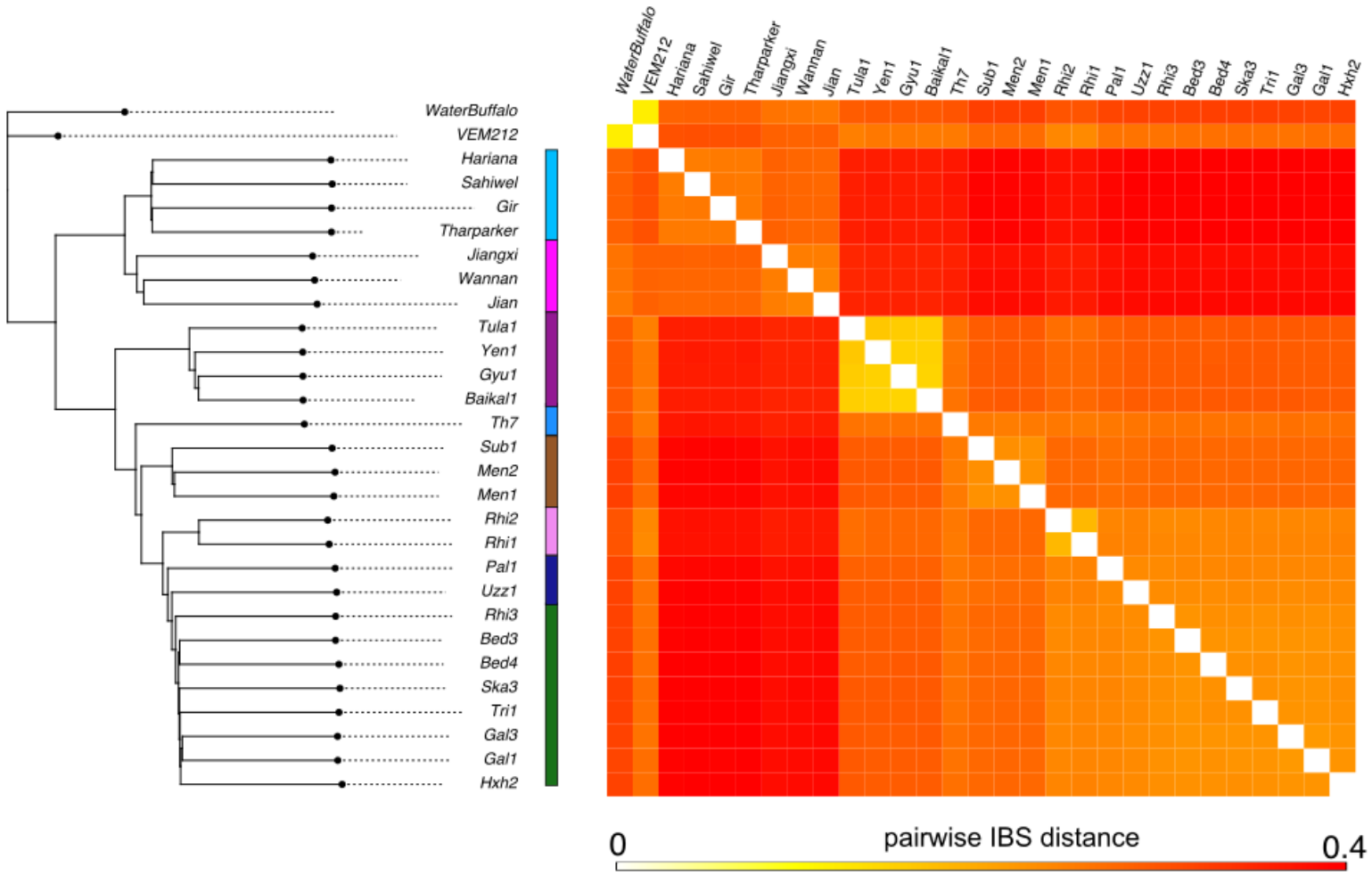


Figure 9.6 - Neighbour-Joining Tree of Identity-By-State Matrix computed with modern and ancient *Bos* genomes of animals that were determined to be best representatives of pre-Holocene populations.

9.4.2. Principal Component Analysis

A PCA comprising 147 *Bos* genomes was computed using genotype likelihoods. This PCA, which included both modern and ancient variation, revealed patterns of structure, hybridisation and genetic isolation.

PC1 differentiates taurine versus indicine ancestry. Modern South Asian and East Asian indicine breeds fall to the extreme right of this PC while modern and ancient Southwest Asian and European domesticated taurine animals fall to the left. Pleistocene and Holocene European aurochs also fall to the extreme left of this PC, with pre-LGM Rhine Valley aurochs displaying some flux, being pulled slightly closer to the indicine end of the PC space. Eastern aurochs also cluster on the left PC1 side, although not quite as extreme as European aurochs. Several taurine breeds drift from the main taurine cluster toward the indicine side of the PC1. These include modern taurine breeds of East Asian, Southern European and West African. The two modern Southwest Asian breeds (South Anatolian Red and Rashouki) also drift from the taurine cluster toward the midpoint of PC1. Ancient Post-Bronze Age Southwest Asian samples similarly drift from this taurine cluster in a stepping stone fashion. In roughly the centre of PC1 are animals that may be described as taurine x indicine hybrids including modern East Asian hybrid animals. Two ancient Central Asian domesticates (Kok1, Yer1) also fall here. Modern East African breeds are placed closer toward the indicine cluster along with a Mediaeval Iraqi domesticate (Bes2).

PC2 separates European wild and Southwest Asian wild and domestic taurine ancestry. Pleistocene and Holocene European aurochs cluster at the extreme top of this PC. Some geographic structure within this cluster is discernible and discussed further below. A group of ancient and modern taurine cattle fall to the extreme bottom of this PC. These samples are exclusively domestic except for Ch22, which is designated as a wild hunted bull from Neolithic Anatolia. Modern domestic breeds from Europe, West Africa, and Asia cluster here. Several individuals pull slightly away from this domestic cluster, including Ch22, and ancient and modern East Asian and European breeds. Ancient Eastern aurochs and modern indicine samples fall roughly at the midpoint of this PC. Modern and ancient taurine x indicine hybrid domesticates are pulled closer to the taurine end of PC2.

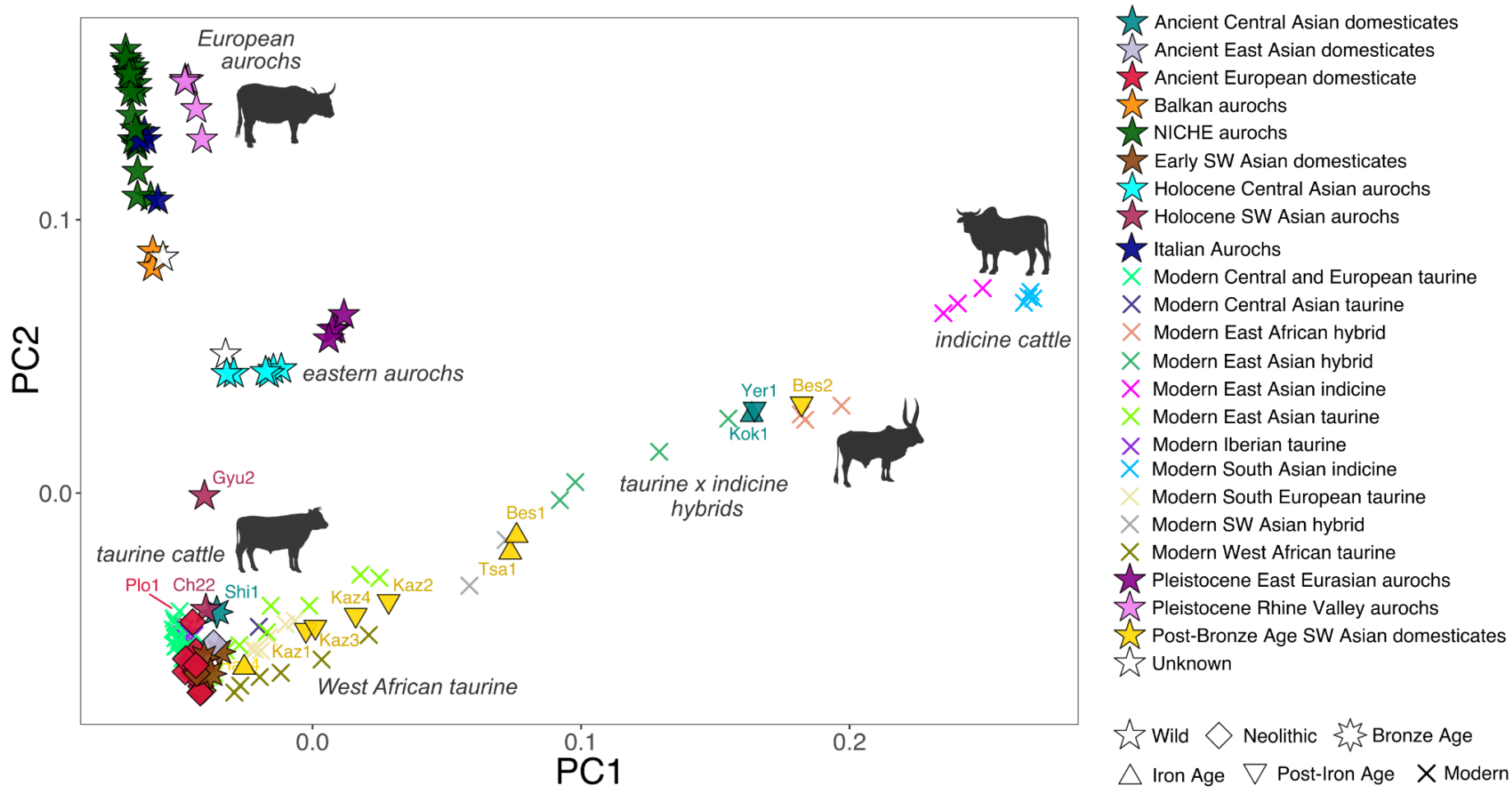


Figure 9.7 - A PCA comprising 147 *Bos* genomes computed using genotype likelihoods with PC1 and PC2 depicted. PC1 is driven by taurine versus indicine variation while PC2 depicts a cline of European wild versus Southwest Asian wild and domestic taurine breeds.

PC3 is driven by East Eurasian Pleistocene variation creating a West-East axis. East Eurasian Pleistocene aurochs from Pre-LGM (Tula1) and Late Glacial (Yen1, Gyu1, Baikal1) cluster to the extreme right while Holocene Europe aurochs, modern indicine breeds, European and Southwest Asian taurine breeds fall to the extreme left. Within the centre of PC3 lies Holocene aurochs from Central Asia. More easterly aurochs from Borly (meridian ~71° east) and Roschinskoe (meridian ~69° east) are pulled closer to this Eastern cluster compared to aurochs from Varfolomeevka (meridian ~48° east). The sample Bor1, which predates Bor2 and Bor3 by 1,000-1,500 years, is also pulled closer to the Eastern cluster. An aurochs recovered from a region in the Former Soviet Union (Tango1) also clusters with samples from Varfolomeevka (Var1, Var2). Several ancient and modern samples that lie on the “Western” axis drift toward the Eastern cluster. The most prominent are modern East Asian breeds, a Neolithic East Asian sample (Shimao), a Central Asian Bronze Age sample (Shi1) and a Holocene Southwest Asian aurochs (Gyu2). Some Eastern affinity may also be interpreted in Bronze Age Southwest Asian domesticates and some Pleistocene Rhine aurochs. Modern breeds from West Africa and hybrids of East Asia also display increased Eastern affinity compared to other breeds.

PC4 depicts an African cline. West African taurines collect at the extreme bottom of this PC while aurochs and modern taurine, indicine and hybrid breeds collect at the extreme top. East African hybrids drift from the main cluster of animals toward West African taurines. Additionally, one modern Southwest Asian breed (South Anatolian Red) occupies this space with East African hybrids. Some ancients display affinity toward the African axis which is most evident in Tsa1, which is an Iron Age Levantine. Bronze Age samples (Has3, Qaz1 and Ace1) from Southwest Asia drift slightly toward the African cluster - displaying more affinity than their Neolithic counterparts. Gyu2, a Holocene Armenia aurochs, is also pulled toward this African cluster.

PC5 depicts an influence seen most prominently in East Asian taurine and ancient domesticates (Shimao and Shi1). However, this PC is not controlled by East Asian aurochs as these animals cluster at the extreme left. Southwest Asian domesticates are pulled slightly to this PC space. PC6 distinguishes South Asian and East Asian indicine and their respective influence on hybrids. Two ancient Central Asian hybrids fall to the extreme bottom of the PC (South Asian). South Asian indicine and their hybrids are pulled toward this axis. East Asian hybrids are pulled toward the East Asian cluster.

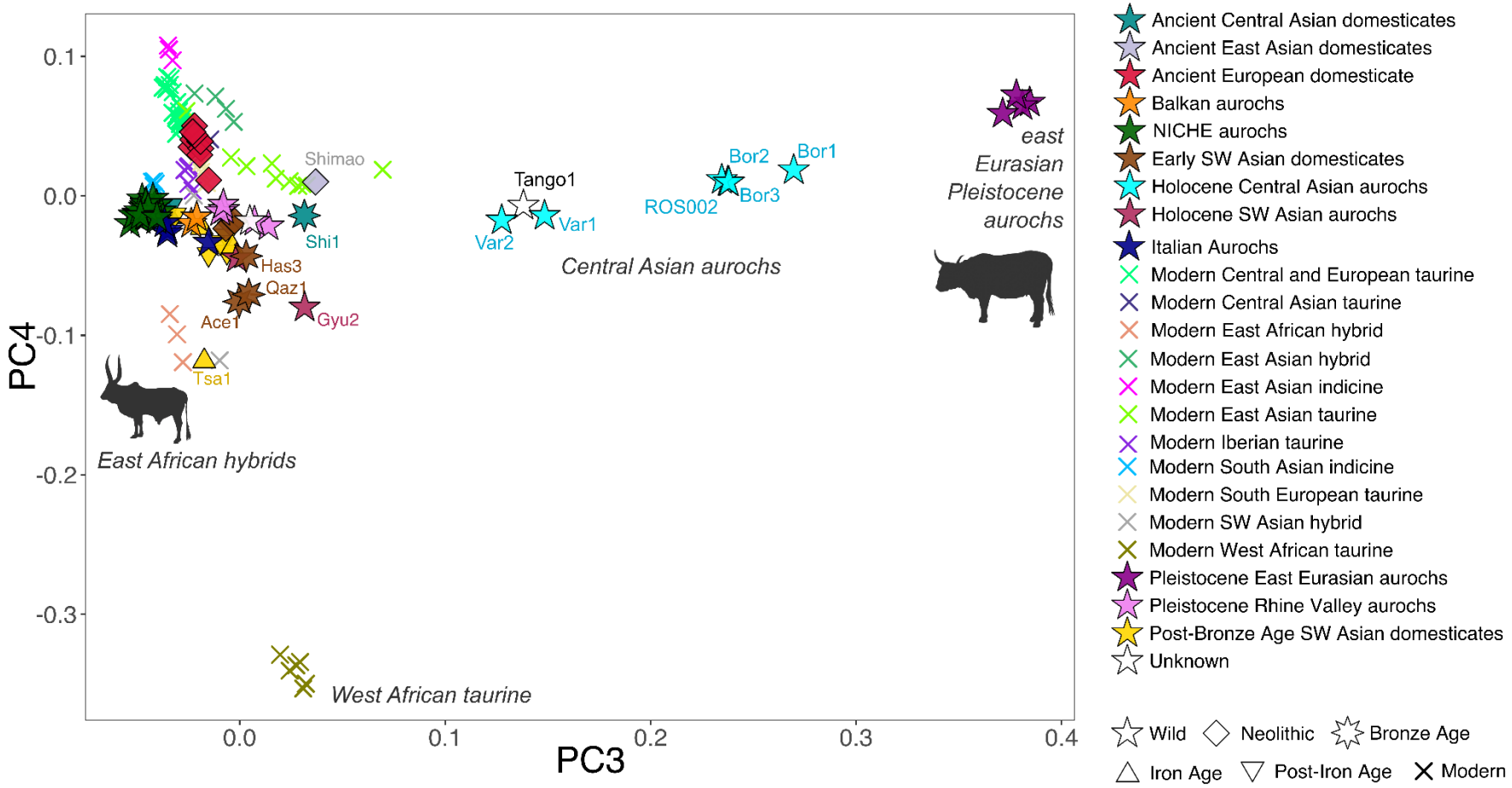


Figure 9.8 - A PCA comprising 147 *Bos* genomes computed using genotype likelihoods with PC3 and PC4 depicted. PC3 depicts a West-East axis of wild diversity while PC4 is controlled by African taurine ancestry.

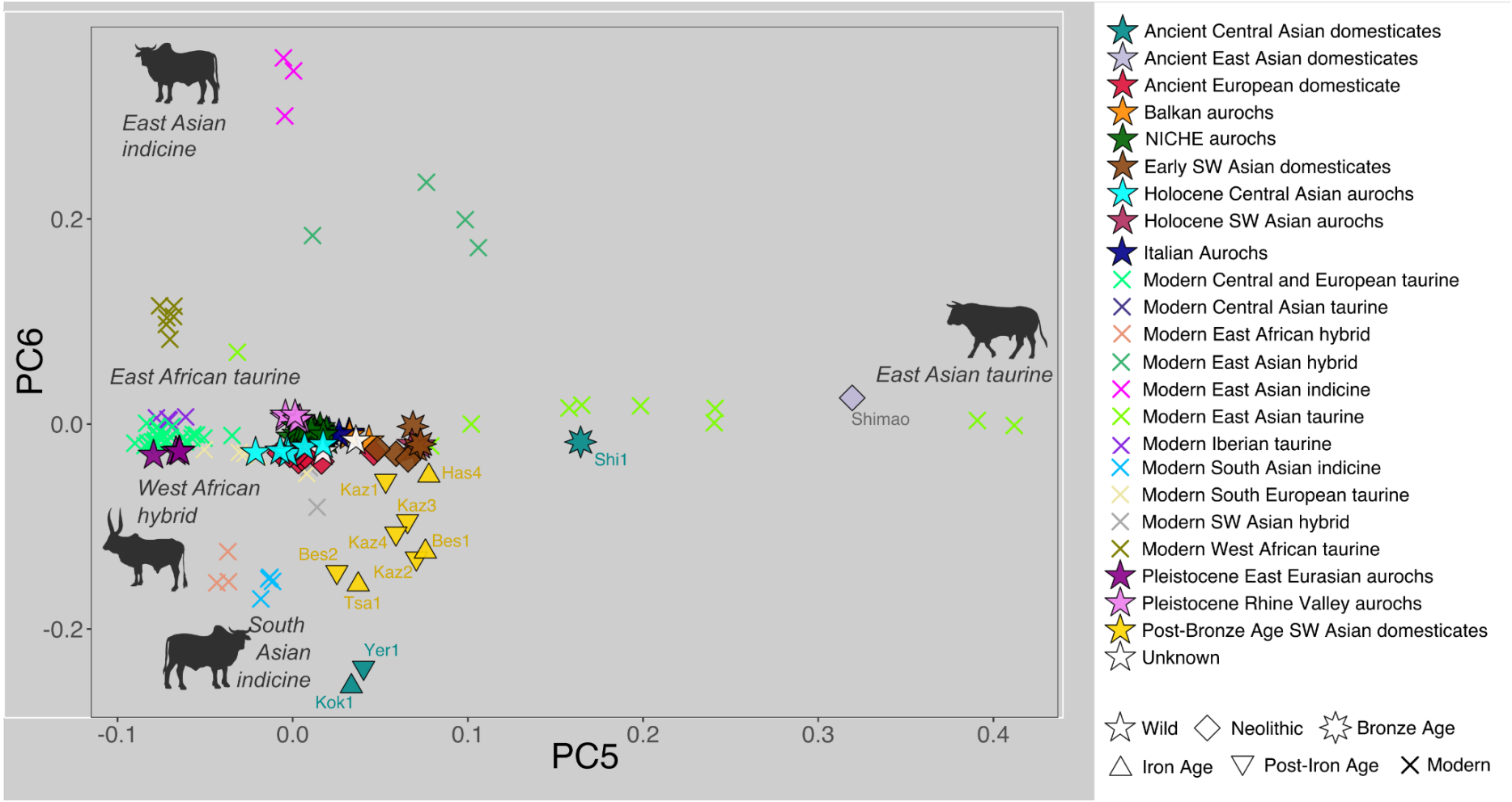


Figure 9.9 - A PCA comprising 147 *Bos* genomes computed using genotype likelihoods with PC5 and PC6 depicted. PC5 is driven by East Asian taurine diversity while PC6 distinguishes South and East Asian indicine ancestry.

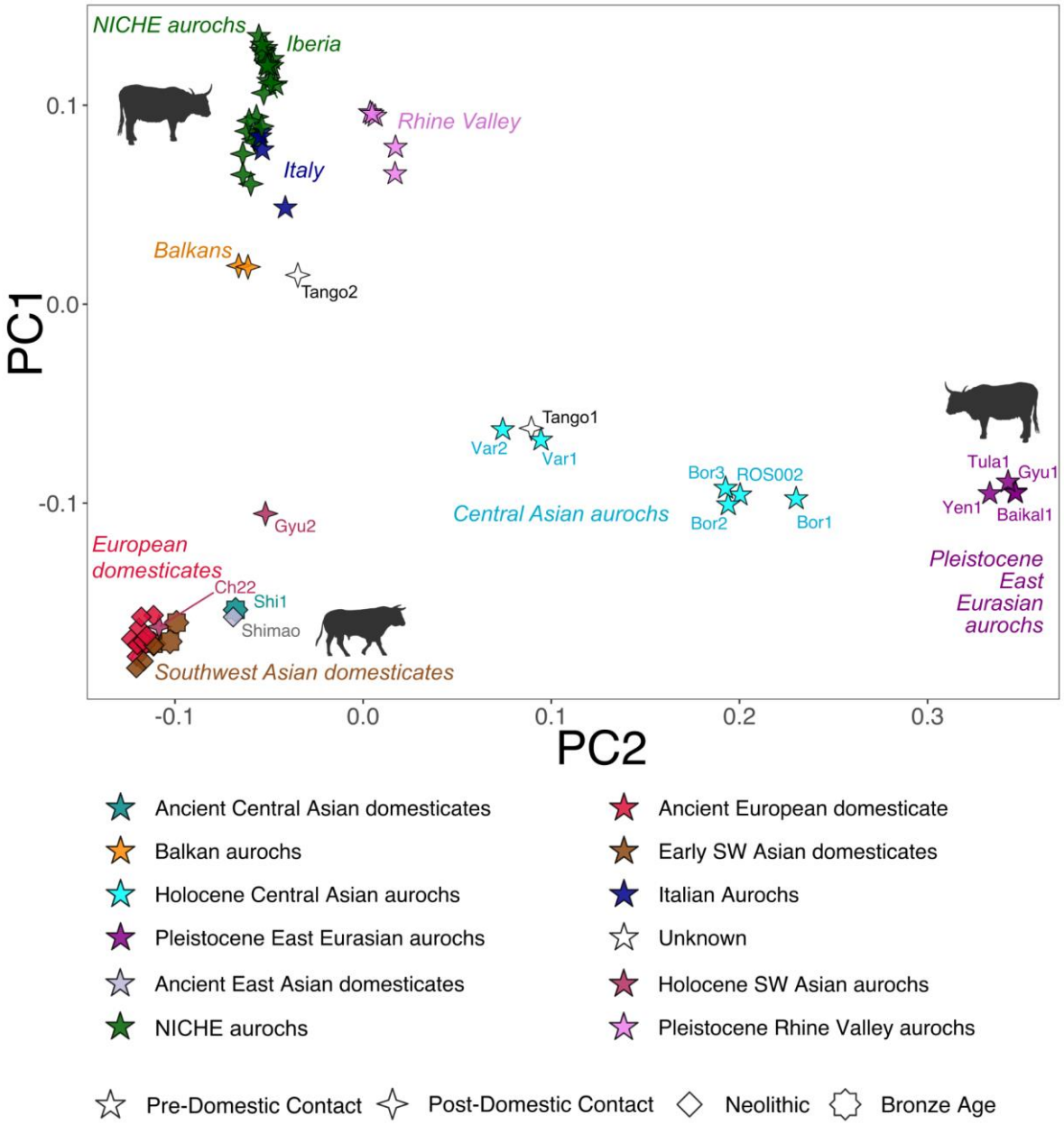


Figure 9.10 - A PCA of ancient *Bos primigenius* variation computed using genotype likelihoods. PC1 is a cline of European wild versus Southwest Asian wild and domestic taurine breeds while PC2 separates Western and Eastern genomes.

A second PCA only ancient sequences with the exclusion of any indicine variation was computed using genotype likelihoods (Figure 9.10-11). PC1 separates Southwest Asian ancestry and European ancestry while PC2 describes variation along a West-East axis. Overall, this depicts three apices of aurochs diversity: European, Southwest Asian and East Eurasian. Structure within the "domestic cluster" emerges in greater focus with ancient European domesticates clustering slightly away from the Southwest Asian domesticates. Within Southwest Asian domestic variation, some flux is evident with Bronze Age samples

drifting slightly to the East. Central and East Asian domesticates similarly drift toward the Eastern cluster, although more affinity is evident. Within this PC space, affinities to each pole of diversity are apparent in Gyu2, Tango1, Tango2 and the Central Asian aurochs, to varying degrees.

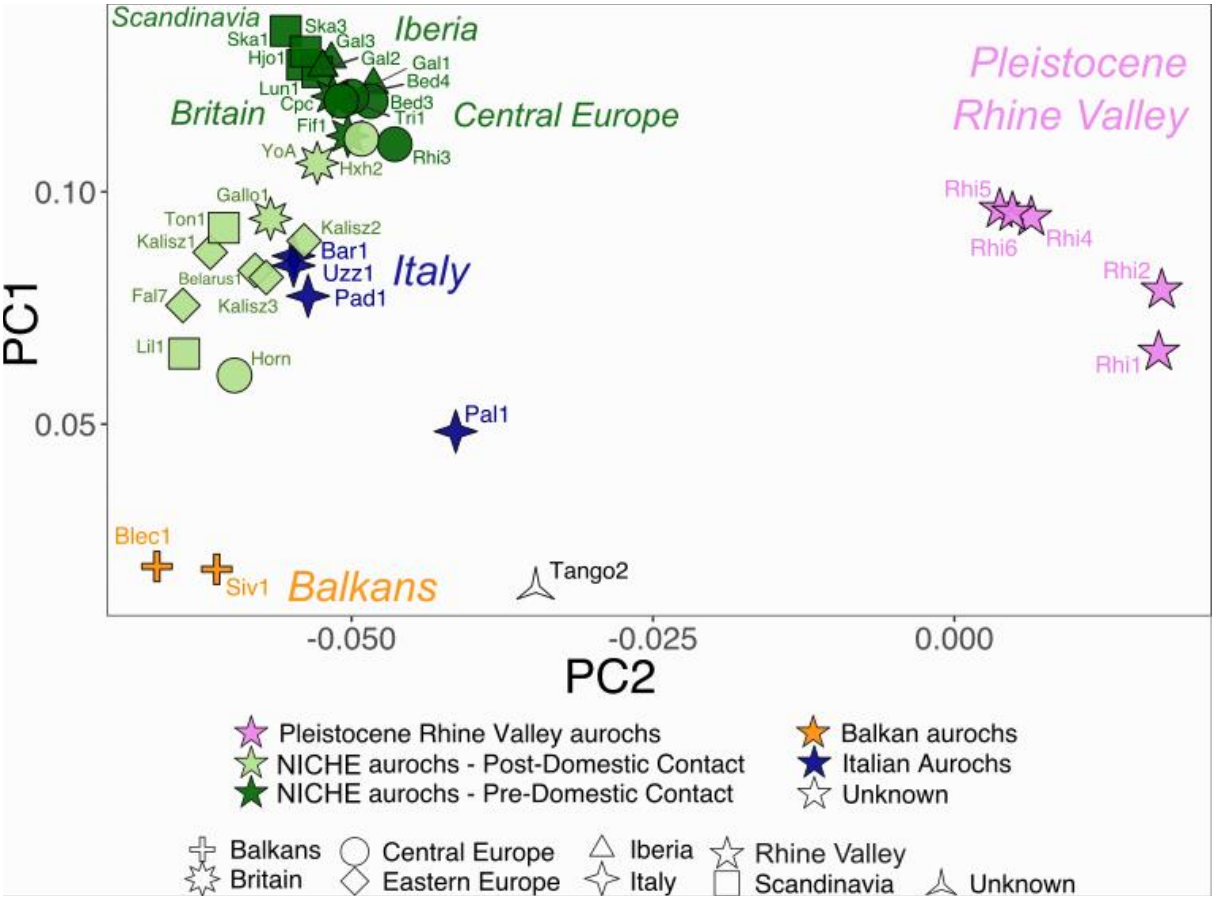


Figure 9.11 - A PCA of ancient European *Bos primigenius* variation computed using genotype likelihoods. PC1 depicts clusters of European aurochs based on their relative affinity to Southwest Asia. PC2, which is controlled by East Eurasian variation, pulls pre-LGM Pleistocene Rhine Valley aurochs from the main cluster. A Late Glacial Italian aurochs (Pal1) and an aurochs of Unknown origin (Tango2) similarly show increased affinity to the Eastern axis.

Structure within the European aurochs pole also emerges in greater detail. While the Pleistocene Rhine Valley aurochs display a high affinity to the cluster of European aurochs, they drift toward the Eastern axis of PC2. Tango2 and the oldest post-LGM European aurochs (Pal1) similarly drift toward the Eastern pole, although this is not as pronounced. Some clusters are apparent within the post-LGM aurochs, perhaps controlled by affinity with Southwest Asia. Pre-domestic NICHE aurochs form a cluster that is furthest from the Southwest Asian pole. Geographic structure is apparent with weak clusters of Iberian, Scandinavian, British and Central European aurochs appearing. Tango2 and two Balkan aurochs (Siv1 and Blec1) drift

closest to the Southwest Asian cluster. Intermediate between these Balkan aurochs and the pre-domestic contact NICHE aurochs are Italian aurochs and a group of post-domestic contact NICHE aurochs. The Italian aurochs exhibit a high level of genetic variability; especially the oldest Pal1 which drifts from the tighter cluster of Bar1, Uzz1 and Pad1. The group of post-domestic contact NICHE aurochs in this space are pulled by the Southwest Asian axis to varying degrees but display a clear separation from the pre-domestication contact NICHE cluster. The exceptions to this are YoA and Hxh2 which occupy a space on the fringes of the pre-domestication contact NICHE cluster.

9.4.3. *D* statistics

D statistics were computed to address specific questions relating to aurochs gene flow. For clarity each test will be discussed separately.

Question 1: How do different aurochs populations relate to each other?
Four populations of aurochs were defined based on PCA and IBS clustering: Southwest Asian aurochs, Pleistocene East Eurasian aurochs, European aurochs and *Bos namadicus*. Southwest Asian aurochs were defined by the earliest domestic Southwestern domestics available (Sub1, Men1, Men2) while East Eurasian aurochs comprise Pleistocene East Eurasian aurochs that display low IBS distance and close clustering on PCA (Tula1, Gyu1, Baikal1, Yen1). European aurochs were defined by pre-Neolithic animals with the least apparent influence from from Southwest Asia (Bed3, Bed4, Cpc, Fif1, Gal1, Gal2, Gal3, Hjo1, Lun1, Rhi3, Ska1, Ska3, Tri1). Finally, as there are no ancient sequences of *Bos namadicus,* four South Asian Indicine breeds (Tharparker, Sahiwel, Hariana, Gir) were selected as the best representatives of a pre-domestication population.

Unequal allele-sharing was detected for Indicine and populations of *Bos primigenius*. Indicine appear to share the most alleles with East Eurasian aurochs, followed by Southwest Asian populations. As detected by clustering methods, a closer relationship of East Eurasian aurochs and European/Southwest Asian populations with respect to indicine is depicted. There is a non-significant difference in allele sharing between East Eurasian aurochs and European/Southwest Asian populations. A closer relationship between European and Southwest Asian populations depicted. Overall, these results suggest that different aurochs populations have unequal sharing with each other.

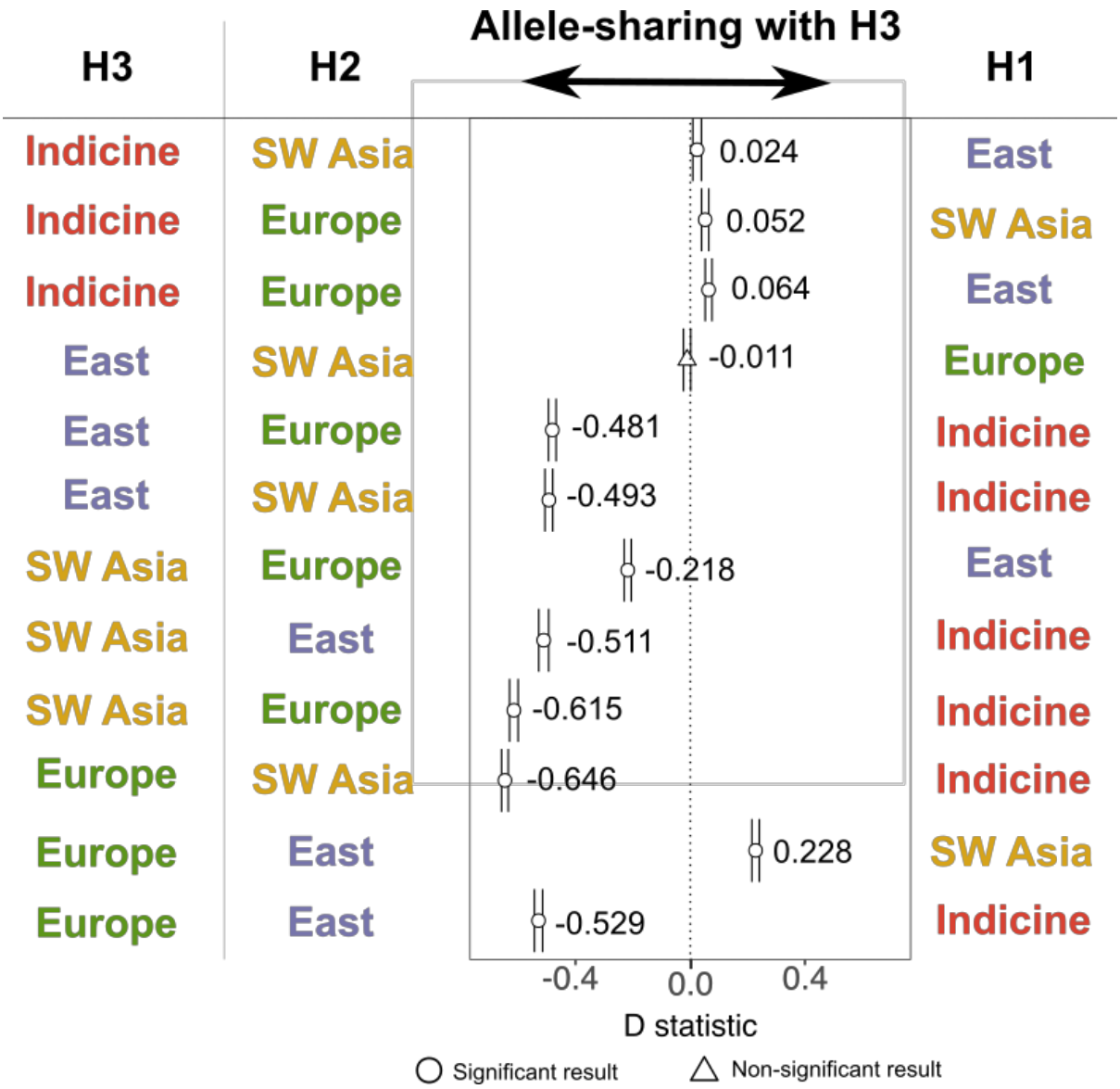


Figure 9.12 - Relative allele-sharing of four populations of *Bos primigenius* variation: Southwest Asian aurochs, Pleistocene East Eurasian aurochs, European aurochs and *Bos namadicus*. Reproduced from Appendix Table 9.2.

Question 2: How have different aurochs populations influenced European aurochs?

To attempt to detect gene flow from different aurochs populations into European aurochs on a temporal basis, the *D* statistic (Test, Bed3, Aurochs population, Outgroup) was computed. H1/Test were all European aurochs and Tango2 due to its clustering with European aurochs on PCA. Bed3 was chosen as an appropriate H2 for several reasons. Firstly its age (~11,600 yrBP) and providence (Central Europe) makes it a likely early representative of the main population that recolonised much of Europe. Secondly, it is the highest coverage pre-domestic context genome at ~22X which will maximise the number of sites that can be used to detect gene flow. H3/Aurochs population were defined as the main axes of aurochs diversity as

detailed in *Question 1* above. Th7 was also used as a H3 test as a Northern African aurochs population.

D statistics reveal significantly less derived allele-sharing between recent Holocene aurochs (Horn-CMLIX, Fal7 and Gallo1) and Pleistocene aurochs (Rhi2 and Rhi4) with indicine animals. In contrast, significant allele-sharing between indicine cattle and several European aurochs (Pal1, Siv1 and Tango2) was detected. However, caution should be used when interpreting these results. As seen in *Question 1*, the Indicine population group shares significantly more alleles with Eastern aurochs and Southwest Asian aurochs compared to European aurochs. As such, the significant Eastern and Southwest Asian influence evident in Pal1, Siv1 and Tango2 may be confounding the allele-sharing results with indicine cattle breeds.

Significant allele-sharing of Southwest Asian aurochs with twelve European aurochs is evident. This allele sharing is most prominent in Balkan aurochs (Siv1 and Blec1) and Tango2. Three Italian aurochs similarly show significant allele-sharing with Southwest Asia, although to a lesser extent as the Balkan aurochs. One Italian aurochs (Pad1) has a non-significant positive D statistic (D=0.023, Z=2.6). None of the pre-domestic contact NICHE aurochs produce a significant result. Meanwhile six post-domestic contact NICHE aurochs produced significant positive D statistics while six produced non-significant D statistics (Hxh2, YoA, Gallo1, Fal7, Horn genomes). All Rhine Valley aurochs displayed significantly less derived allele-sharing with Southwest Asian aurochs.

Eastern aurochs displayed significantly negative allele-sharing with one Rhine Valley aurochs and several recent NICHE aurochs (Gallo1, Fal7, Horn-10537). Significantly positive *D* statistics were recorded for several genomes including Italy (Pal1), the Balkans (Siv1, Blec1), Poland (Kalisz3) and most prominently Tango2. Allele-sharing with Th7 produced generally insignificant results except for slightly negative values for Gallo1 and four Rhine samples.

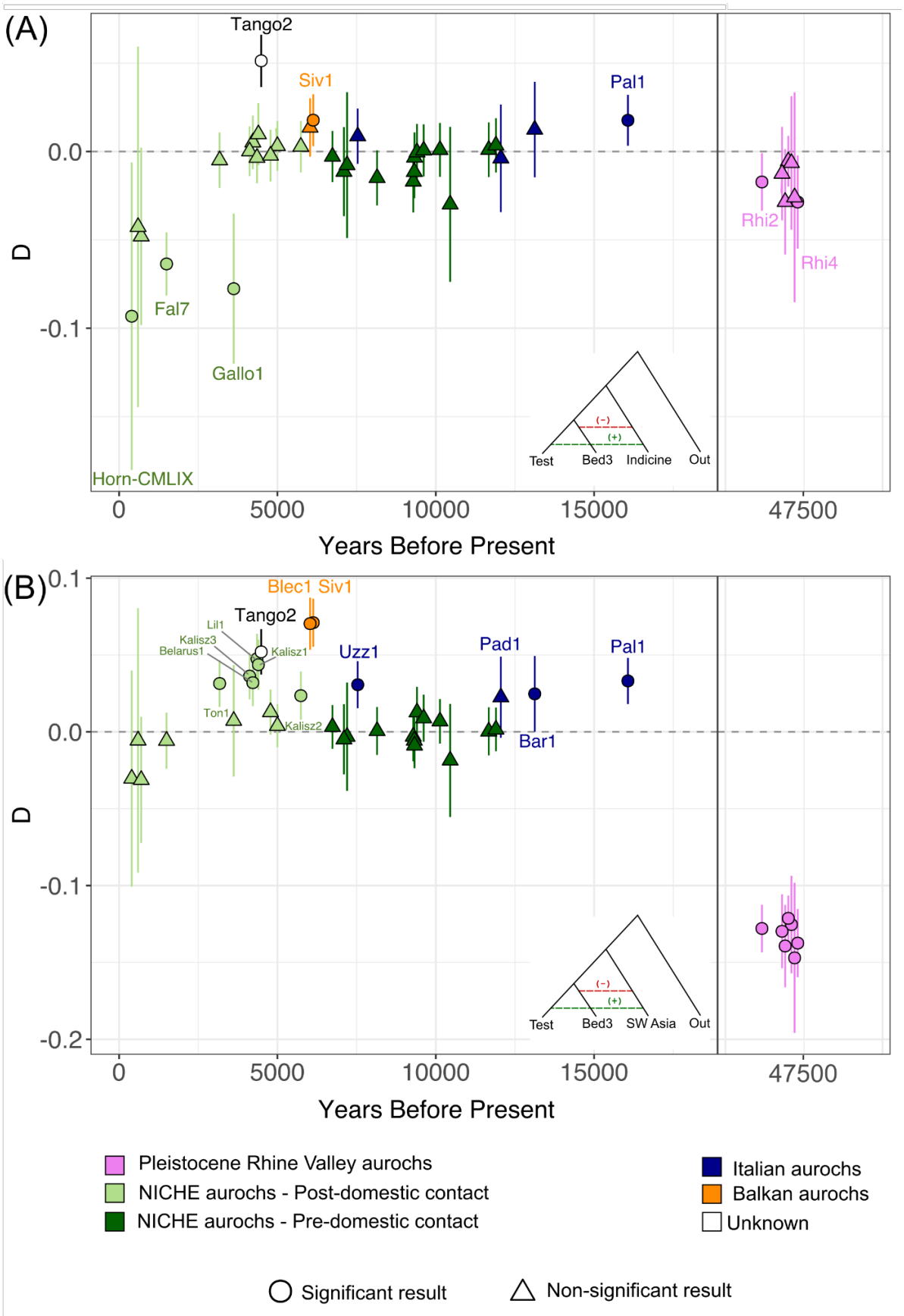


Figure 9.13 - Relative allele-sharing of (A) indicine and (B) Neolithic Southwest Asian domesticates with European aurochs. Reproduced from Appendix Table 9.3-4.

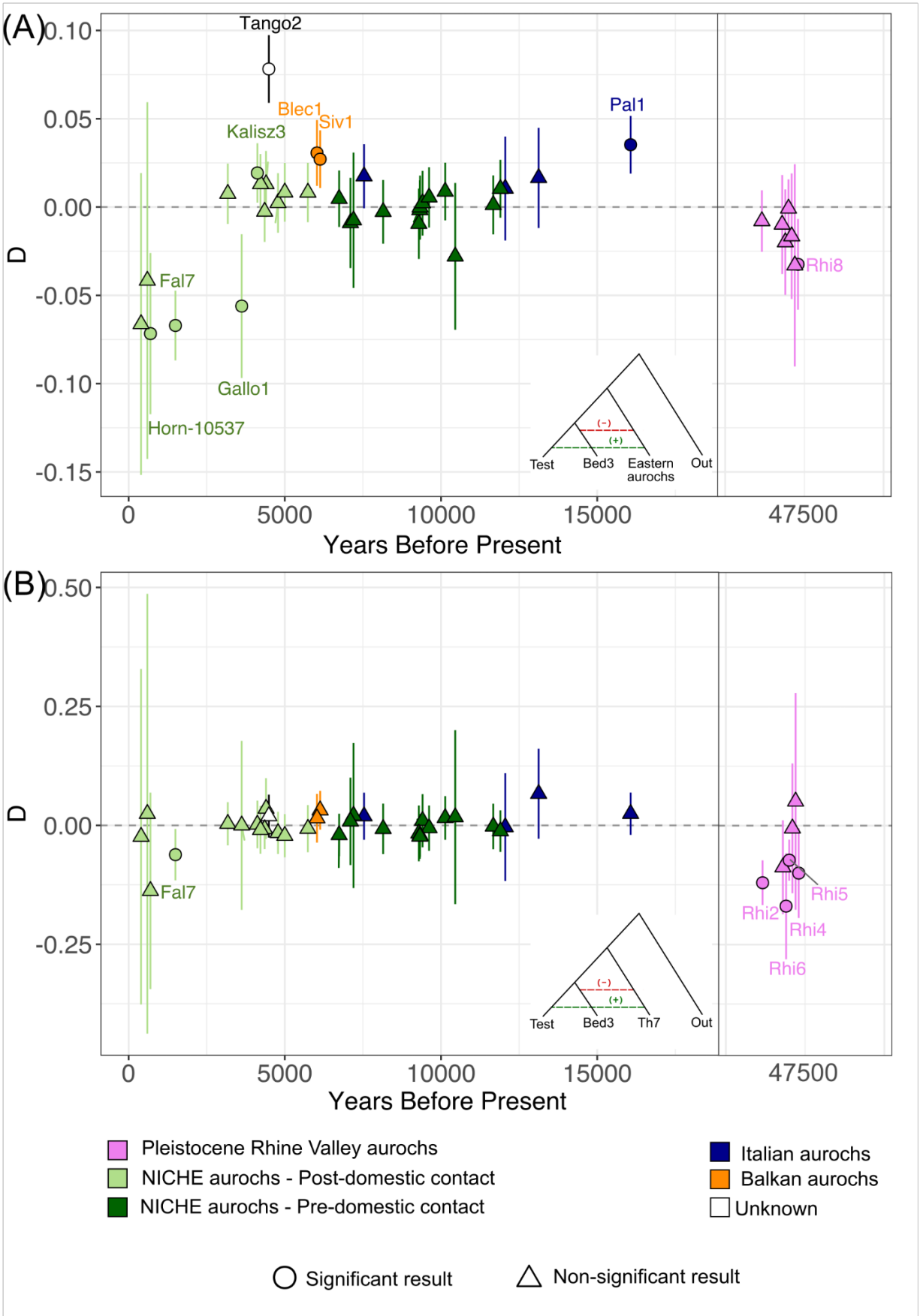


Figure 9.14 - Relative allele-sharing of (A) Eastern aurochs and (B) Th7 with European aurochs. Reproduced from Appendix Table 9.5-6.

Question 3: How has Southwest Asian aurochs influenced European aurochs populations?
To specifically address the question of Southwest Asian aurochs influence on European aurochs since the LGM, the *D statistic* (Test, Pleistocene Rhine Valley, Southwest Asia, Outgroup) was computed. H2/Pleistocene Rhine Valley comprised all available Pleistocene Rhine Valley aurochs as a proxy for pre-LGM European populations. Southwest Asia comprised Sub1, Men1 and Men2.

The results depict a clear influence of Southwest Asia in Europe with respect to the Pleistocene Rhine Valley. This influence existed prior to the arrival of the Neolithic and is apparent shortly after the LGM in Italy (Pal1, ~16,000 years old). A cline of Southwest Asian ancestry is depicted which appears to be controlled by geographical proximity to Southwest Asia. Balkan aurochs display the highest allele-sharing with Southwest Asia. Following this, Italian aurochs, Tango2 and most post-domestic contact NICHE aurochs display the highest allele-sharing. Post-domestic contact aurochs from Poland and Belarus (Kalisz1, Kalisz2, Kalisz3, Belarus1) show the most elevated allele-sharing of all NICHE aurochs. Of the Scandinavian samples, the post-domestic contact aurochs (Lil1, Ton1) display more allele-sharing than pre-domestic contact aurochs (Ska1, Ska3, Lun1, Lun2). The lowest allele-sharing is seen in early Holocene Scotland (Fif1) and the drinking horn samples.

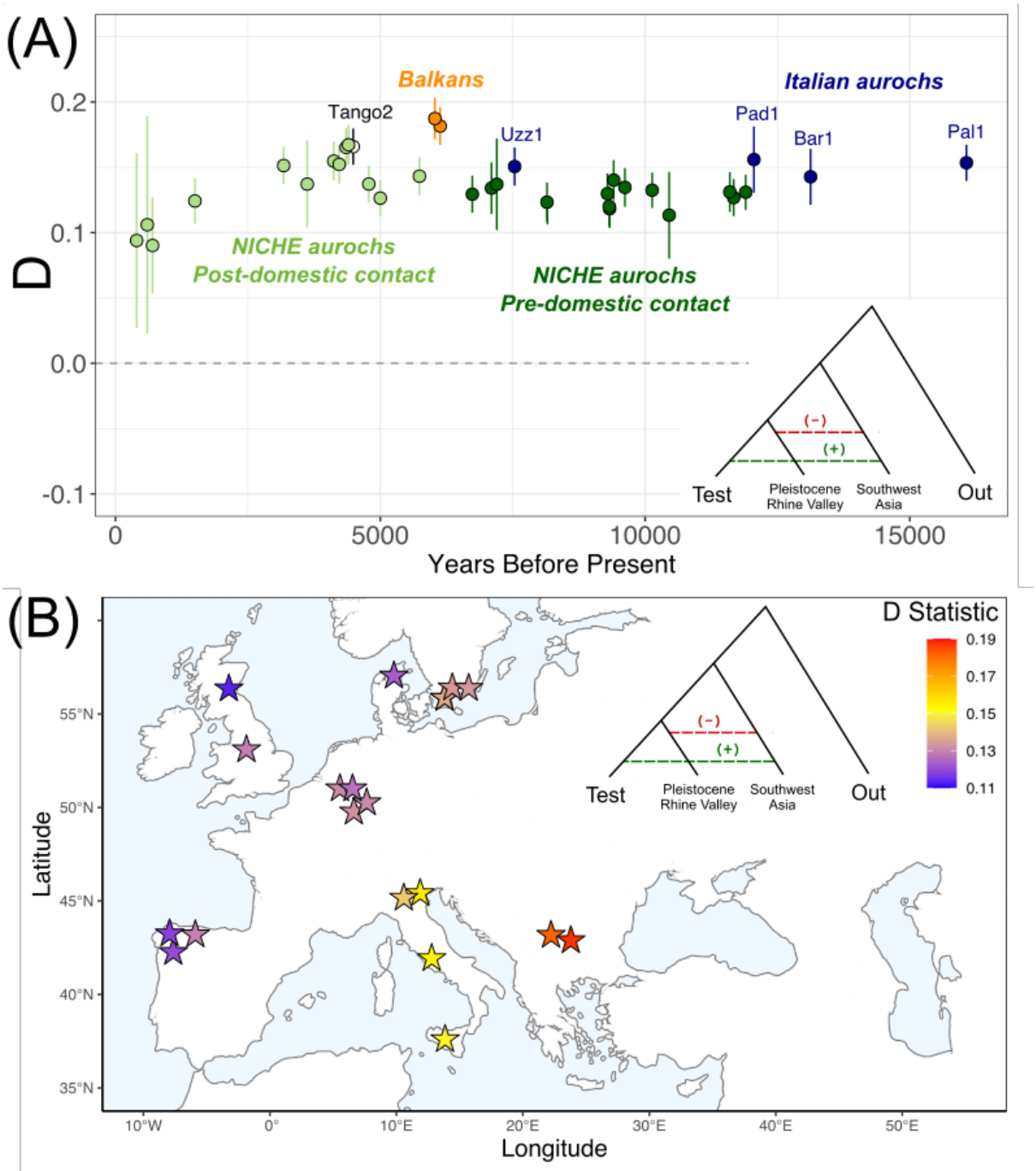


Figure 9.15 - (A) Excess allele-sharing in post-LGM European aurochs with Southwest Asian genomes with Pleistocene Rhine Valley genomes as a reference. (B) Geographical variation of *D* statistics of Balkan, Italian and pre-domestic contact NICHE animals. Reproduced from Appendix Table 9.7.

Question 4: How have aurochs populations influenced ancient domestic cattle?

To attempt to detect post-domestication wild introgression in ancient domesticates, the *D* statistic (Test, Neolithic Anatolia, Aurochs population, Outgroup) was computed. H2/Neolithic Anatolia is represented by Sub1, Men1 and Men2 as the best available representation of the

original population of domesticated Southwest Asian cattle. H3/Aurochs populations are defined as in *Question 1*. As Gyu2 and Ch22 derive most of their ancestry from Southwest Asian aurochs, they are included in this analysis as a “Test” genome.

European aurochs input is detected in Neolithic Southeast European domesticates to varying extents. Notably, several earlier European domesticates do not display increased allele-sharing. These samples comprise domesticates from Blagotin and Bubanj in Serbia (Bla1 and Bub1) and Boeotia in Greece (Sar38). Sar38 and several later domesticates (the Chinese Neolithic Shimao, Iron Age Levantine Tsa3 and Iranian Has4) display significantly negative allele sharing with European aurochs. None of the Group 5 Horn samples designated as aurochs show significantly positive allele sharing with European aurochs, with one sample displaying a negative D statistic (Horn-617). Southwest Asian aurochs Gyu2 and Ch22 display significant allele sharing with European aurochs.

Eastern aurochs influence is most prominent in Caucasian aurochs Gyu2 and Central and East Asian domesticates, where significant allele-sharing is detected. While not statistically significant, a positive *D* value was computed for Iranian Bronze Age samples Qaz1 and Has3 (D=0.015, Z=2.98; D=0.015, Z=2.91). This contrasts with Anatolian Bronze Age sample Ace1 and Iranian Neolithics Sac3 and Far1 where neutral or negative *D* statistics are computed. Significantly negative *D* statistics are also computed for several ancient European domesticates and Has4, an Iron Age Iranian domesticate.

Significant allele-sharing with indicine is detected in some later Southwest Asian domesticates - Has1, Has4, Tsa3, Has5. Significantly negative allele-sharing is detected in all European ancient domesticates, again possibly confounded by a European wild introgression. Shimao, a Chinese Neolithic sample, also displays negative allele-sharing with indicine. Th7 does not compute any significantly positive D statistic but does result in a negative D statistic for Has4, Shimao and Horn-11570.

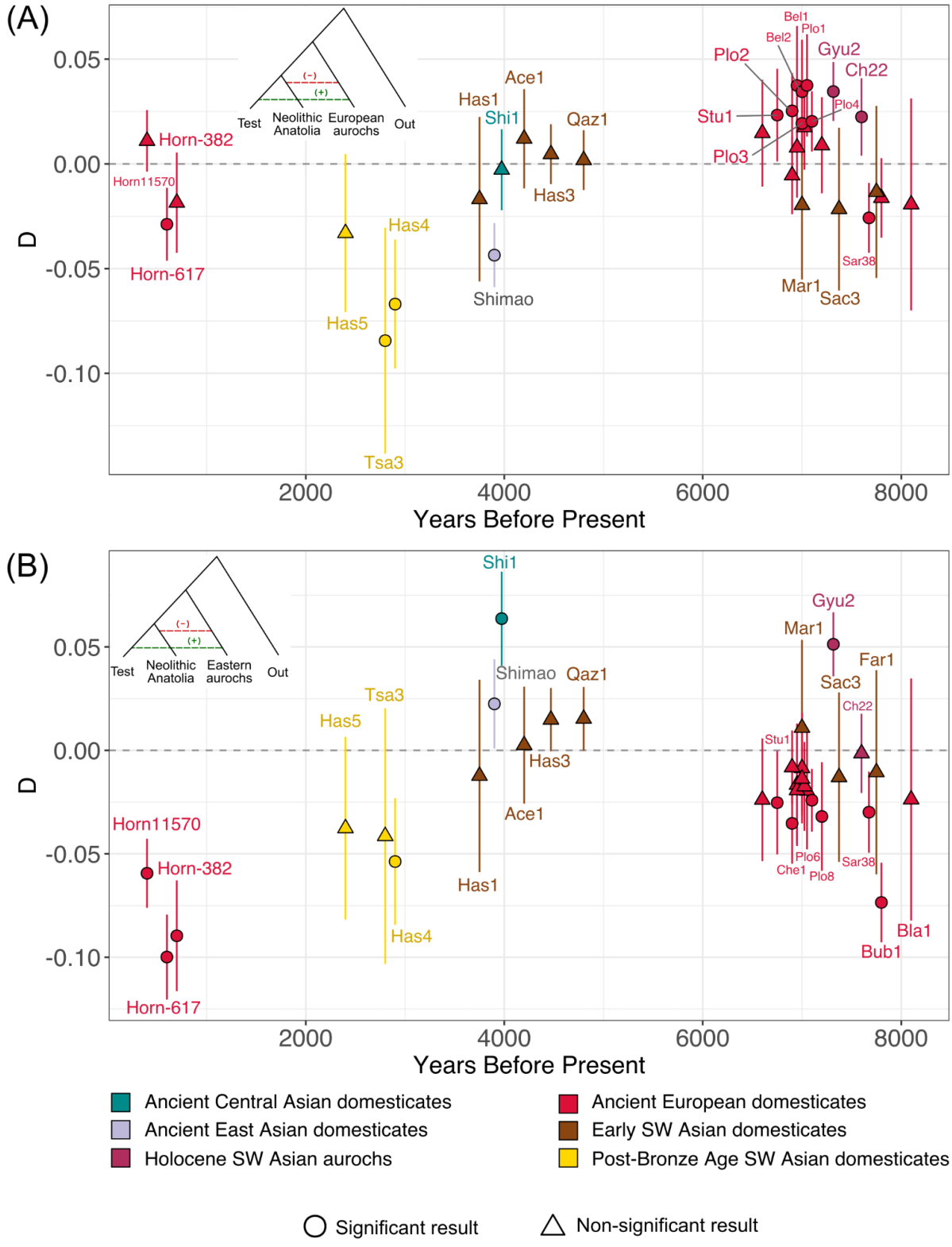


Figure 9.16 - Relative allele-sharing of (A) European aurochs and (B) Eastern aurochs with ancient domesticates. Reproduced from Appendix Table 9.8-9.

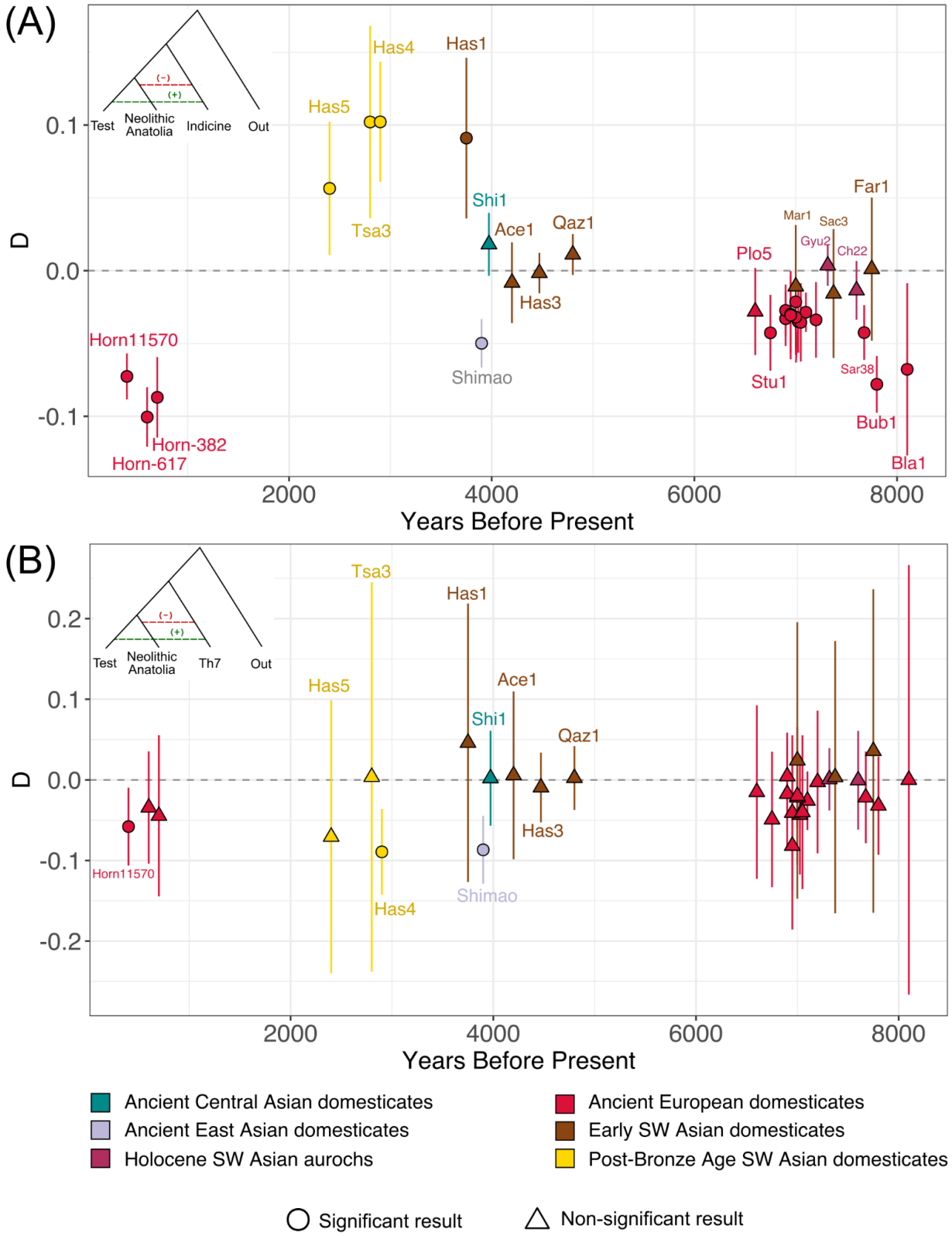


Figure 9.17 - Relative allele-sharing of (A) indicine and (B) Th7 with ancient domesticates. Reproduced from Appendix Table 9.10-11.

Question 5: How have aurochs populations influenced modern taurine breeds?
To detect excess wild allele-sharing in modern breeds, the *D* statistic (Test Breed, Neolithic Anatolia, Aurochs population, Outgroup) was computed. H2/Neolithic Anatolia is represented by Sub1, Men1 and Men2 as the best available representation of the original population of domesticated Southwest Asian cattle. H3/Aurochs populations are defined as in *Question 1*.

As expected, all hybrid breeds (modern West African hybrids, modern East Asian hybrids, modern SW Asian hybrids) share excess alleles with indicine. This majority component of Indicine ancestry will therefore confound other *D* statistics for these hybrids. As such, all hybrid breeds compute a significantly negative *D* statistic with *Bos primigenius* aurochs populations.

All East African taurine breeds have significant allele-sharing with indicine populations, although consistently lower than in West African hybrids. However, unlike the hybrids, three East African taurines (Muturu, Baoule, Somba) show a significantly positive *D* statistic with the African *Bos primigenius* Th7. While significantly negative allele-sharing with European aurochs is detected for all East African taurines, three breeds (Muturu, Baoule, Somba) produce a non-significant positive *D* statistic with Eastern aurochs.

The patterns for Indicine and Eastern aurochs allele-sharing with Central and East Asian taurines depict varied aurochs contributions among breeds. Significantly positive allele-sharing with Indicine is detected for seven Asian breeds (Mongolian, Goursounsi, Bohai, Tibetan, Yakut, Yanbian and Kazakh) with *D* statistics that range from 0.036 to 0.244. While this indicine introgression likely confounds other tests, four Asian breeds (Kuchinoshima, Tibetan, Mishima, Hanwoo) still computed for allele-sharing a significantly positive D statistic with Eastern aurochs. All Asian taurines have negative allele-sharing with European aurochs and African aurochs.

European breeds reveal clear trends based on geographic location. Most evidently, Southern European taurines all exhibit a positive *D statistic* with indicine - a trend that is not observed in any Central and Northern European or Iberian breed. Similarly, all Central and Northern European or Iberian breeds except Hereford and Charolais display derived allele-sharing with European aurochs (eight of which are non-statistically significant). It is noteworthy that a reference bias may be affecting Hereford as the *Bos taurus* reference genome was assembled from a Hereford breed. In contrast, all Southern European breeds exhibit negative allele-

sharing with European aurochs. All European breeds reveal negative allele-sharing with African and Eastern aurochs with respect to Neolithic Anatolia.

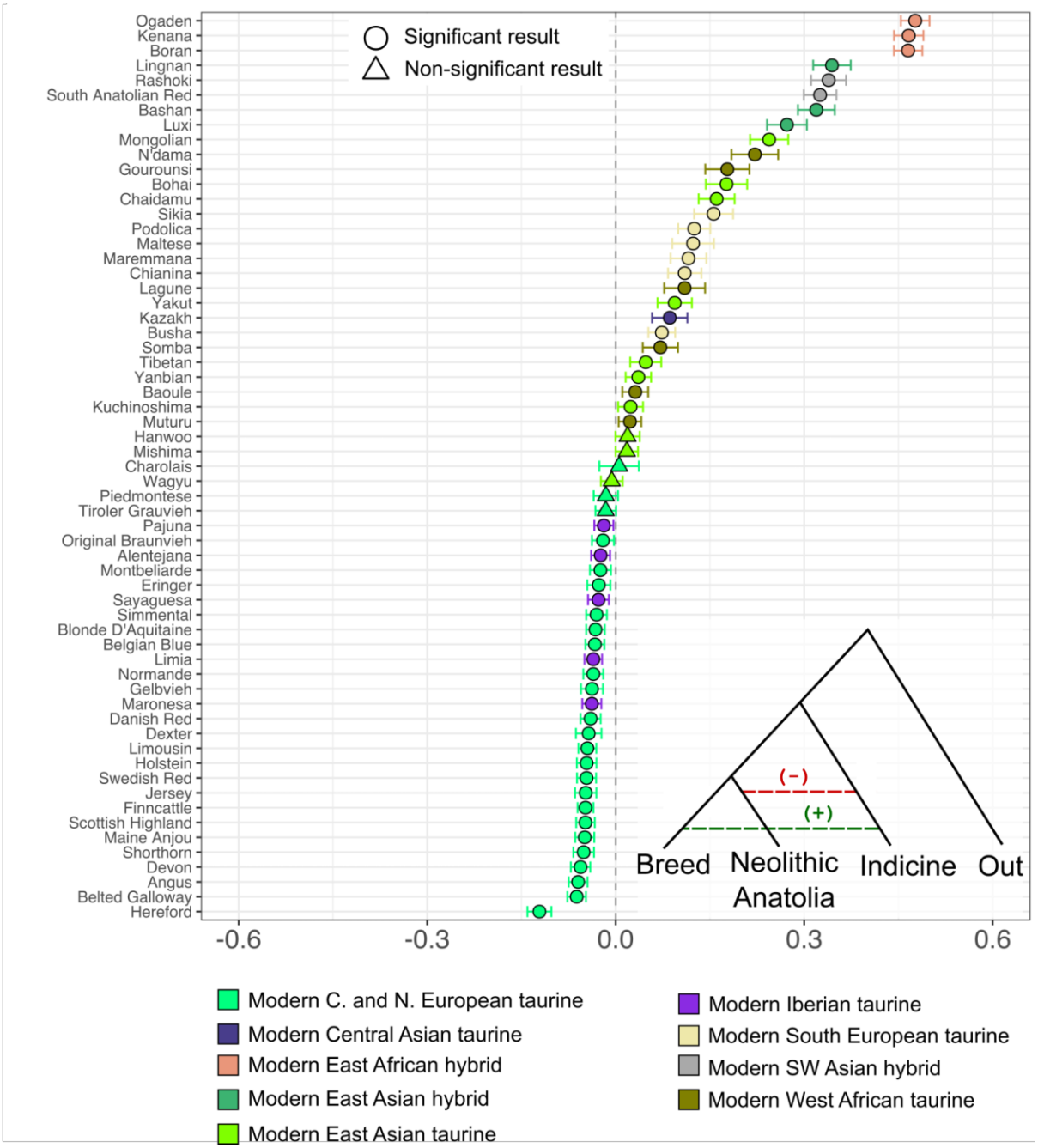


Figure 9.18 - Excess allele-sharing of modern breeds with indicine cattle. Reproduced from Appendix Table 9.12.

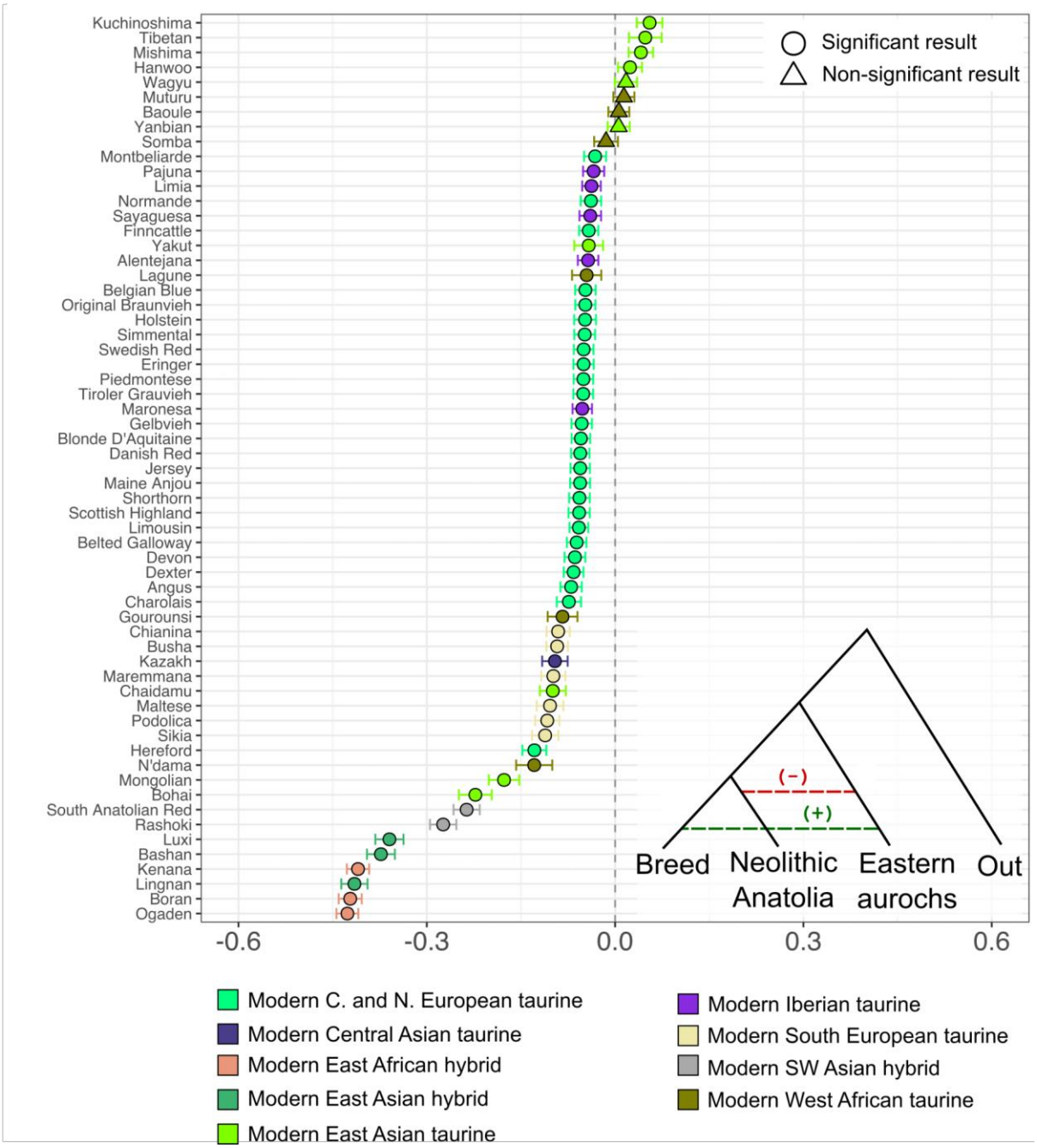


Figure 9.19 - Excess allele-sharing of modern breeds with Eastern aurochs. Reproduced from Appendix Table 9.13.

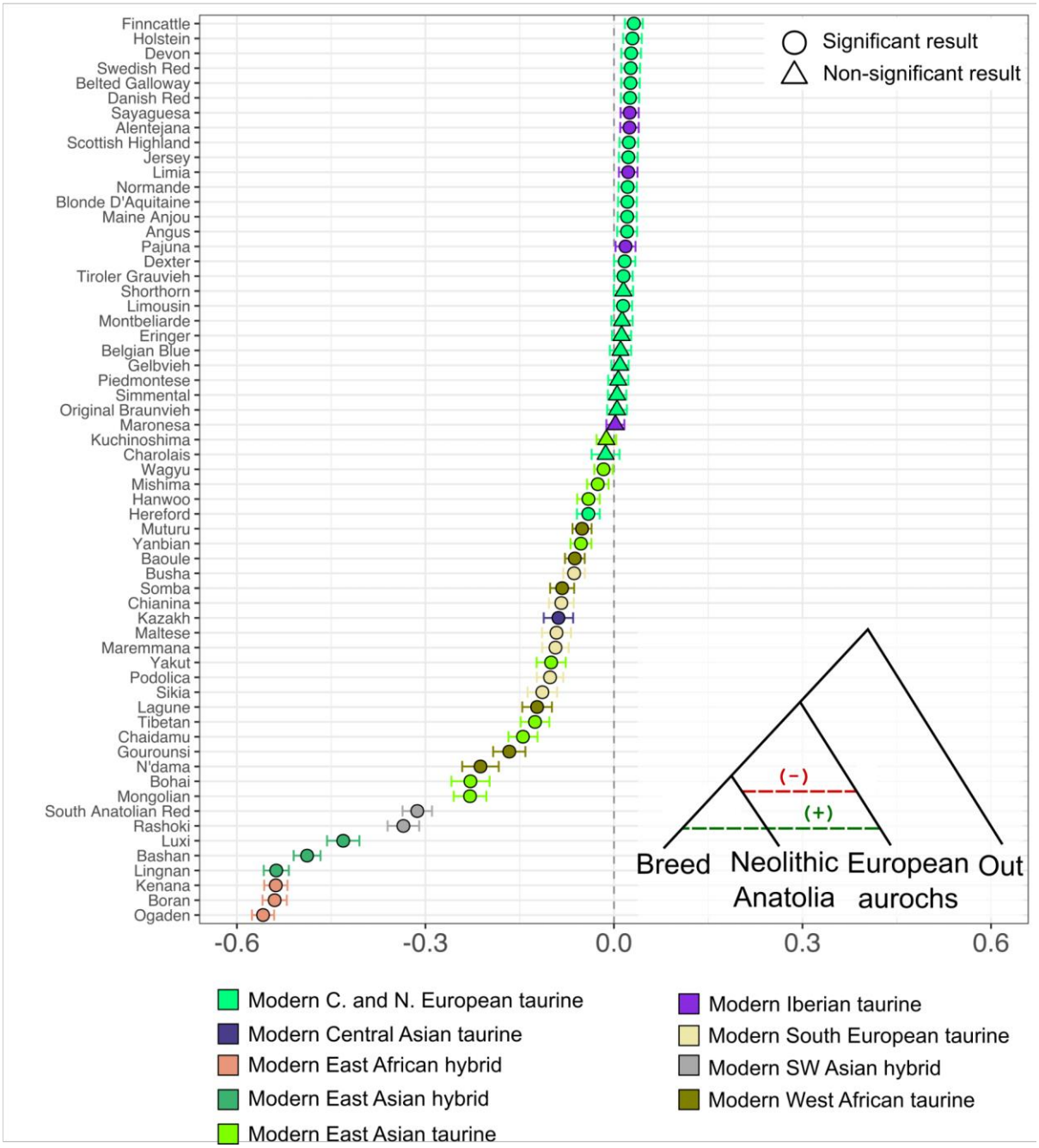


Figure 9.20 - Excess allele-sharing of modern breeds with European aurochs. Reproduced from Appendix Table 9.14.

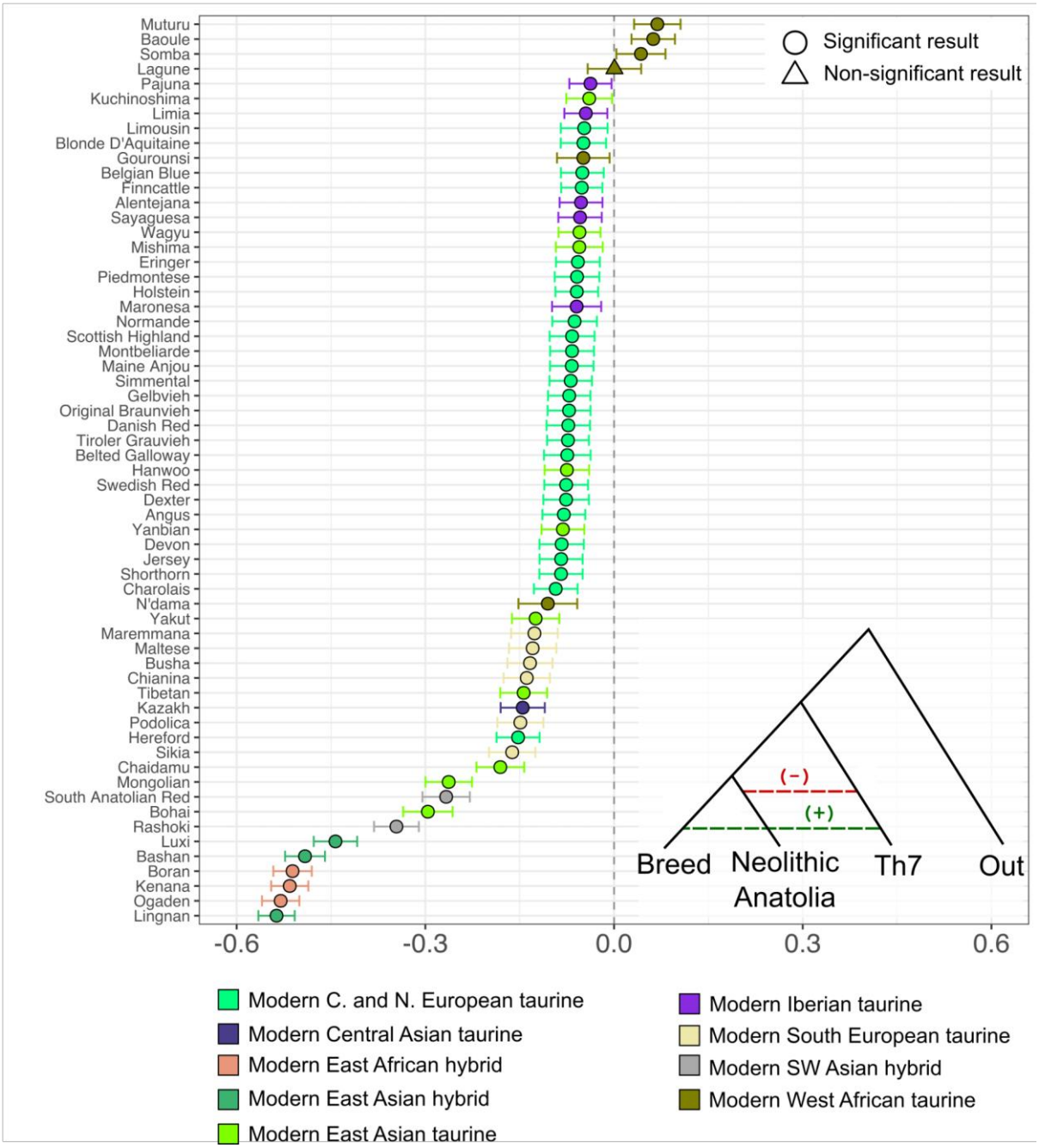


Figure 9.21 - Excess allele-sharing of modern breeds with Th7. Reproduced from Appendix Table 9.15.

9.4.4. ADMIXTURE Analysis

ADMIXTURE was run on all samples above 1X coverage using a range of K values (1-10). The CV error indicated that K1-K4 resulted in the most consistent results (Appendix Table 9.16).

- K2 described two ancestral components that describe Eurasian *Bos primigenius* versus *Bos namadicus*. The exception is that East Eurasian aurochs ancestry is modelled as ~20% indicine, ~80% taurine.
- K3 describes three ancestral components that divide domestic taurine ancestry, indicine ancestry and European aurochs. East Eurasian aurochs are now modelled as 90% European aurochs and 10% indicine.
- K4 describes four ancestral components that have been previously noted in PC1, PC2 and PC3: European aurochs, East Eurasian aurochs, domestic taurine and domestic indicine variation.
- K5 describes European aurochs, East Eurasian aurochs and indicine components as before but now splits taurine variation into two groups: modern European and Asian taurine breeds. Ancient domesticates and taurine breeds of Africa and Southwest Asia are modelled as a composite of these components.
- K6 describes variation as above but includes an ancestral component of African taurine. Some ancient domesticates and taurine breeds of Southwest Asia are modelled as a composite of these three taurine components. Neolithic Anatolian breeds are now modelled as the same ancestral component as Asian taurine breeds.

In the four population model (Figure 9.22-24), the implied components are: European aurochs (green), East Eurasian aurochs (purple), Southwest Asian aurochs (yellow) and Indicine variation (red). The results are depicted in three plots for clarity.

Pre-domestic contact NICHE aurochs are modelled as deriving all ancestry from one European ancestral component. Pre-domestic contact Italy similarly derives all ancestry from this component except for Pal1, a ~16,000 year old Italian aurochs which has ~2% of its genome modelled as Southwest Asian. In a post-domestic contact context, some NICHE aurochs are modelled with a minority of Southwest Asian ancestry (Lil1, Fal7, Horn-10537, Horn-10543). Southwest Asian ancestry in European aurochs peaks with two Balkan aurochs that are modelled with ~25%. Pleistocene Rhine Valley aurochs are modelled with a majority European component with a minority Eastern component (6-14%). One aurochs of unknown providence (Tango2) displays an ancestral profile most similar to Balkan aurochs, with ~69% European ancestry and ~22% Southwest Asian ancestry. However, unlike Balkan aurochs, it also has a ~9% Eastern component.

An Eastern component appears most prominently in Pleistocene East Eurasian aurochs and Holocene Central Asian aurochs. The four Pleistocene East Eurasian aurochs are modelled

as deriving 100% of their ancestry from this ancestral population. In Central Asia, it appears as the majority component in all hybrids, although some aurochs also present Southwest Asian and European ancestry (Var1, Var2, Bor2, Bor3). Var1 and Var2 display a larger component of European ancestry (~27-30%) compared to other Central Asian samples which range from 0-11%. The ancestral profile of Var2 is very similar to Tango1, although the European and Eastern components of Tango1 are slightly higher (33% and 45% compared to 30% and 43%) and its Southwest Asian component is slightly lower (22% compared to 26%).

The two ancient Eastern domesticates (Shimao and Shi1) are modelled with a majority Southwest Asian component with a minority Eastern (13-16%) component. A similar profile is observed in all but one Central and East Asian taurine breeds where a minority Eastern component ranges from ~2-30%. The Chinese breed Bohai is the exception, which is modelled as mostly taurine with some indicine (~23%) influence. Indicine ancestry also appears in other Central and East Asian breeds to varying degrees (0-23%).

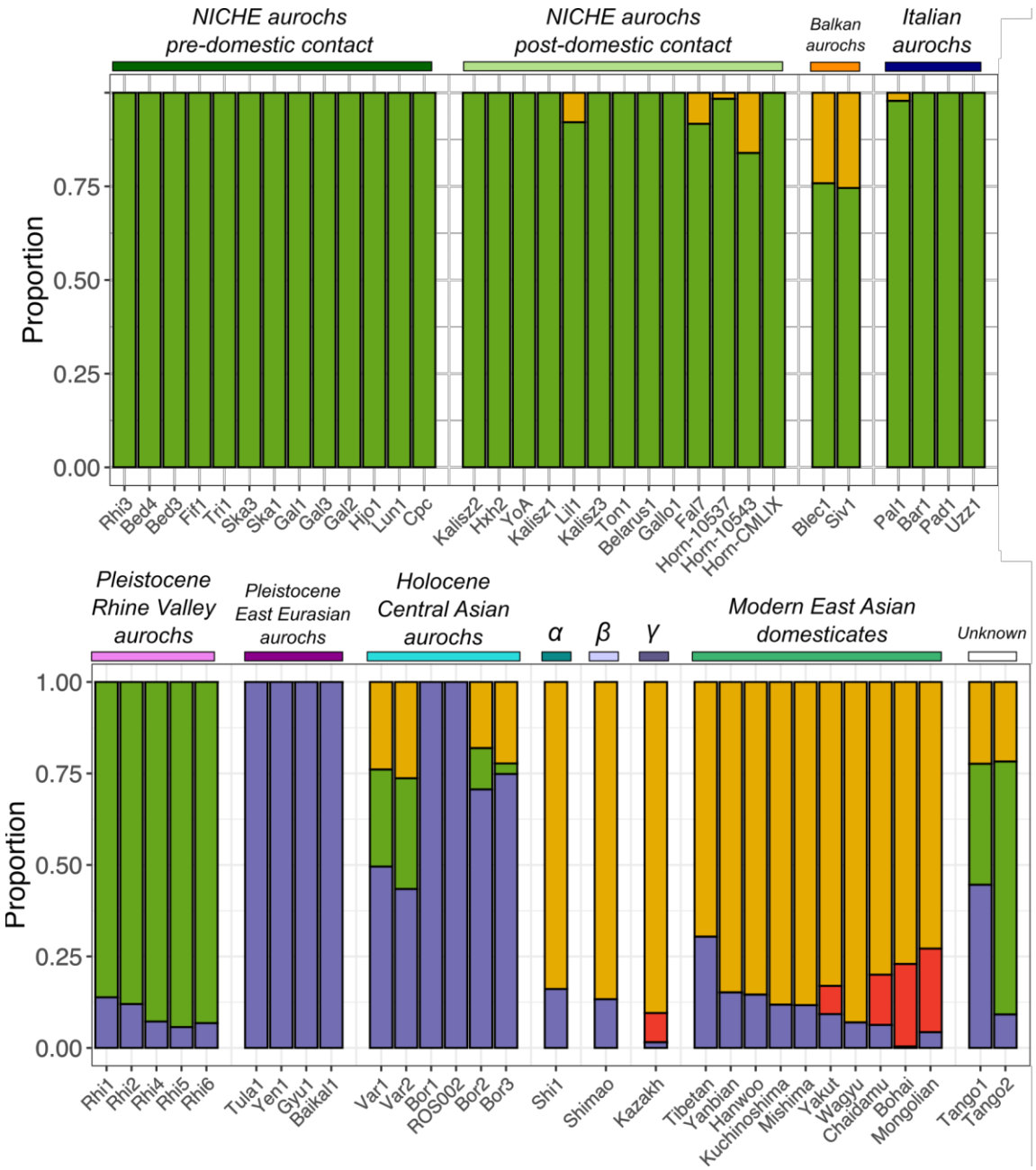


Figure 9.22 - Results I. from the unsupervised ADMIXTURE analysis of ancient and modern *Bos* with K=4 ancestral components. α = Ancient Central Asian domesticates, β = Ancient East Asian domesticate, γ = Modern Central Asian taurine. Results from Appendix Table 9.17.

The ancestry of Neolithic Southwest Asian domesticates (Sub1, Men1 and Men2) is modelled completely by a Southwest Asian component. Two Neolithic Southwest Asian aurochs (Gyu2, Ch22) are also dominated by Southwest Asian ancestry, with minority components of European (~23% and ~8%) and Eastern ancestry (~12% and ~1%) also appearing. Some flux is observed in the Bronze Age (Ace1, Qaz1 Has3) where the Eastern component ancestry appears as a minority component (2-3%). In Post-Bronze Age Southwest Asia and Central Asia, a substantial intrusion of indicine ancestry is observed. This indicine ancestry is modelled

with a large range of values in ancestral profiles, constituting both minority and majority components (~5-83%). The minority Eastern component that is first observed in Bronze Age Southwest Asian domesticates persists in some Post-Bronze Age Southwest Asia samples (Has3, Kaz4, Kaz1), although to a lower degree (~1-2%). In a single sample (Kaz3), a minority European ancestral component also appears.

The profiles of Ancient European domesticates are similarly dominated by a Southwest Asian ancestry, with sporadic minor components of European ancestry (~1-10%). Similarly, Modern Central and Northern European taurine and Iberian taurine are modelled with mostly Southwest Asian ancestry, with a minority component of European ancestry appearing in all Iberian breeds (~5-8%) and most Central and Northern breeds (~0-12%). A similar trend is observed in Southern European taurine breeds substituting the minority European component with an Indicine component (~7-13%).

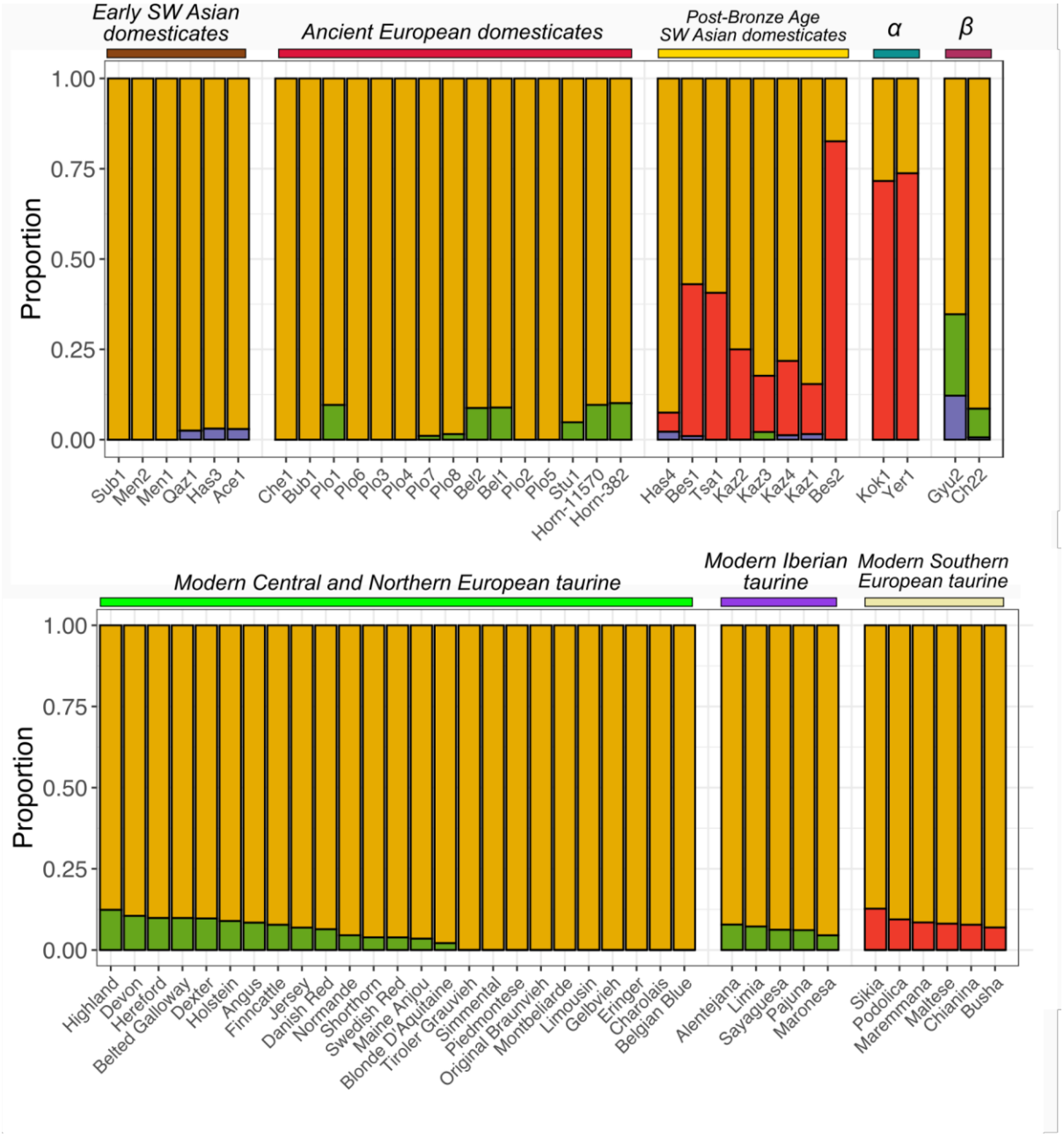


Figure 9.23 - Results II. from unsupervised ADMIXTURE analysis of ancient and modern *Bos* with K=4 ancestral components. α = Ancient Central Asian domesticates, β = Holocene SW Asian aurochs. Results from Appendix Table 9.17.

Modern South and East Asian indicine breeds are modelled as deriving from a single ancestral indicine population. Hybrid breeds of East Africa, East Asia and Southwest Asia are modelled as a composite of Southwest Asian and Indicine ancestries, with East Africa displaying the highest proportion of indicine and Southwest Asia displaying the least. Three of the four East Asian hybrid breeds are also modelled with a minority Eastern component (~1-5%). Finally, modern African taurine breeds are depicted with a majority Southwest Asian component with

3-5% Eastern ancestry. Four breeds (N'dama, Gourounsi, Lagune, Somba) also display an intrusion of indicine ancestry (~7-20%).

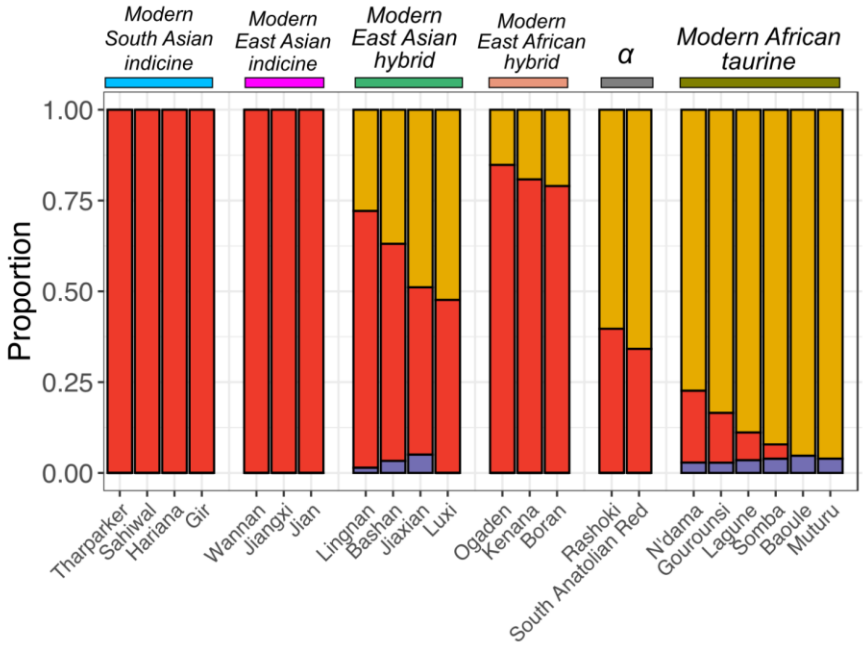


Figure 9.24 - Results III. from unsupervised ADMIXTURE analysis of ancient and modern *Bos* with K=4 ancestral components. α = Southwest Asian taurine. Results from Appendix Table 9.17.

9.4.5. Runs of Homozygosity

The two pre-LGM aurochs displayed no signs of ROH greater than 1Mb. After the LGM, extents of ROH are variable. The Late Glacial Siberian East Eurasian aurochs (Yen1 and Baikal1) display more elevated inbreeding compared to the Caucasian East Eurasian aurochs, Gyu1. Low levels of homozygosity runs are observed in some Holocene aurochs from Europe and Central Asia (Gal1, Var1, YoA). The highest estimated ROH is found in the Balkan aurochs Siv1, and the Danish aurochs Lil1. In modern breeds, the percentage of ROH is generally higher although some breeds depict no ROH. A genome of the endangered Japanese breed, Kuchinoshima, displays extreme inbreeding, with up to 55% of its genome in ROH (Figure 9.26).

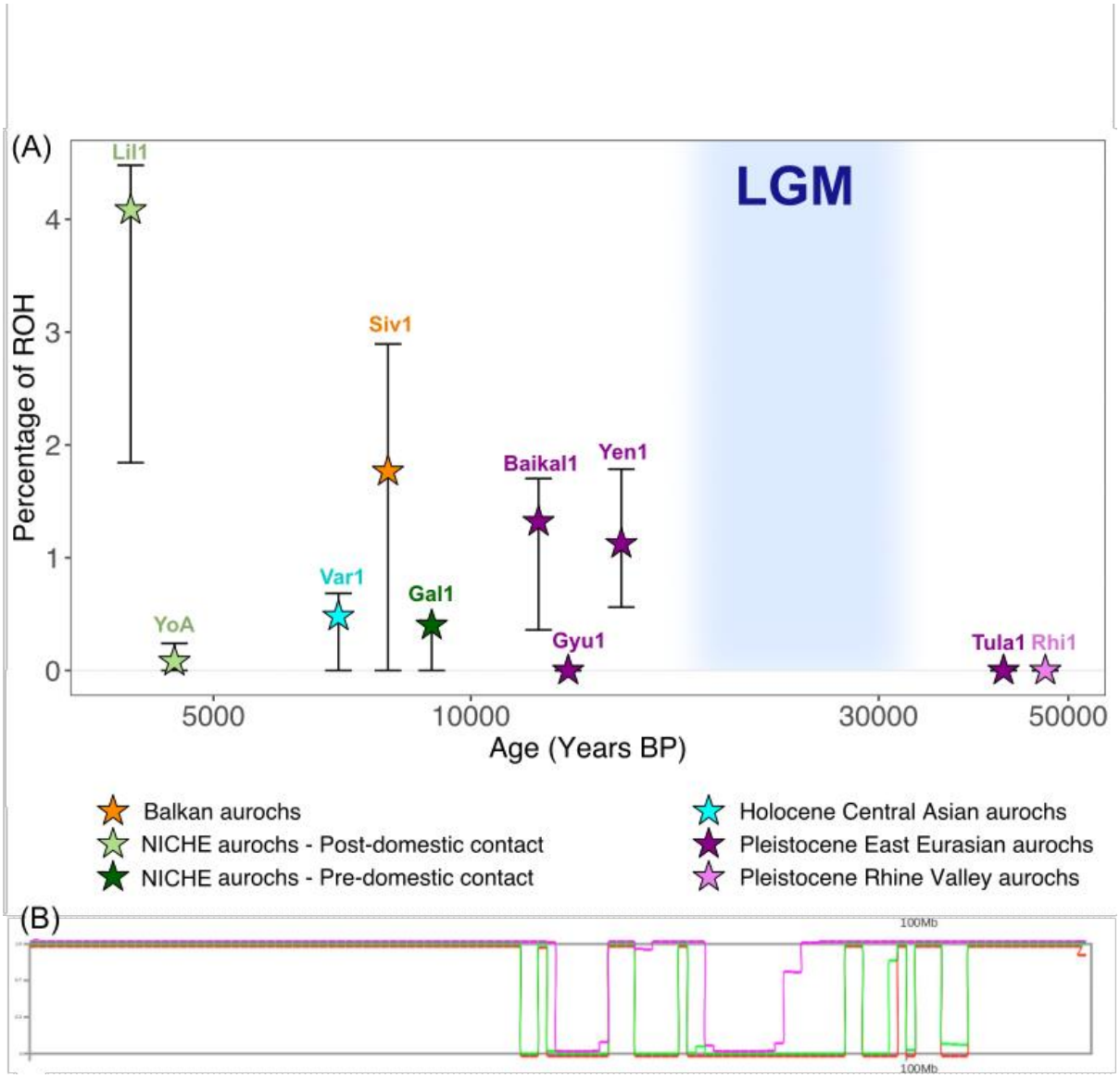


Figure 9.25 - Percentage of ROH in ancient genomes. (A) The point estimate is depicted by a star and the lower and upper confidence intervals derived from the low point and high point of the MCMC chains. **(B)** A visualisation of the ROHs on chromosome 5 of the Danish aurochs Lil1.

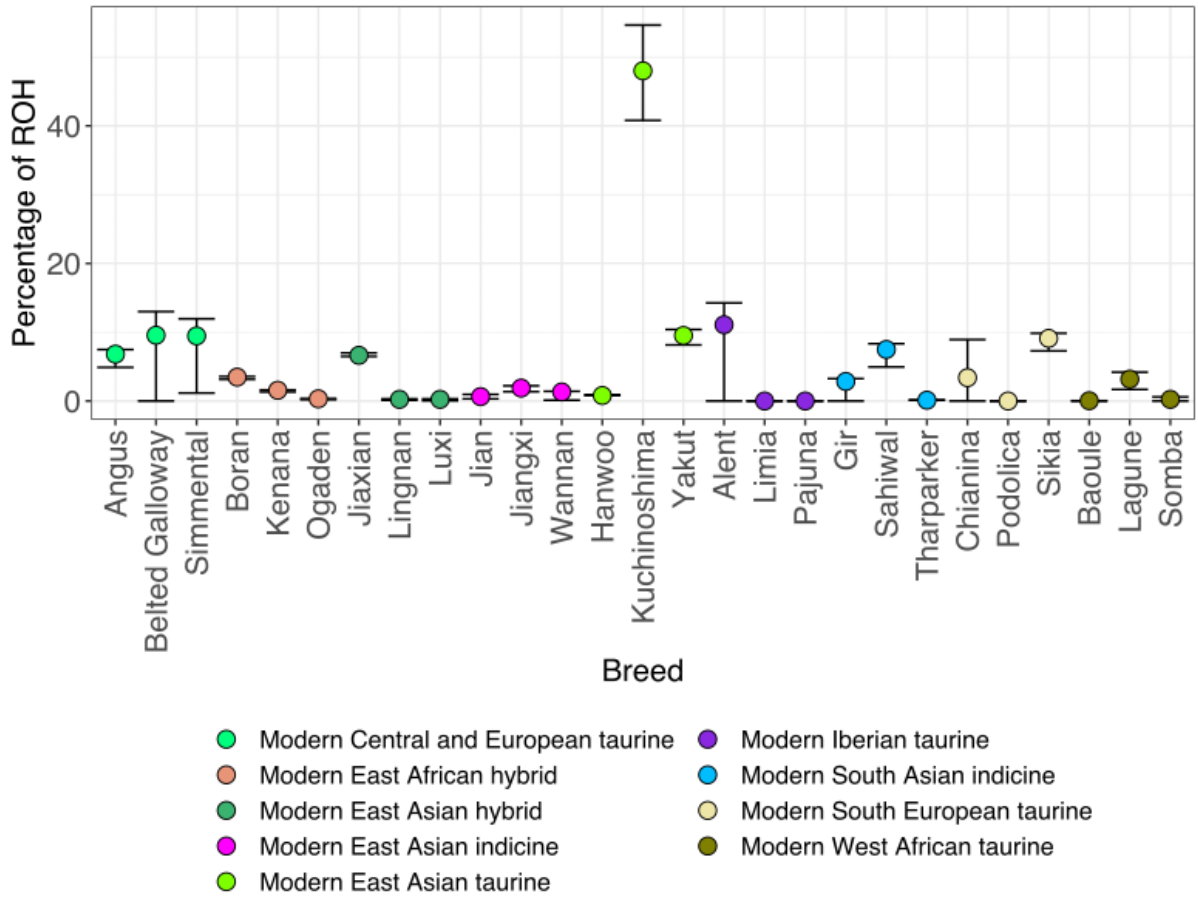


Figure 9.26 - Percentage of ROH in modern breeds represented by one genome. The point estimate is depicted by a circle and the lower and upper confidence intervals derived from the low point and high point of the MCMC chains.

9.4.6. MSMC

The PSMC' curves estimating the demographic history of 12 ancient and modern *Bos* genomes are displayed in Figure 9.27. Genetic separation of lineages may be inferred when PSMC' curve trajectories diverge. The oldest divergence in *Bos* is between *Bos primigenius* and *Bos namadicus*, beginning approximately 300,000-400,000 years ago. At this point the *Ne* of the indicine lineages begins to recover from a downward trend which continues until a peak around 100,000 years ago. It is notable that the Mediaeval Iraqi domesticate Bes2 has a deflated recovery, likely due to its admixture obscuring demographic history. After this peak, *Ne* collapses for indicine lineages and continues to fall until the last time series around 10,000 years ago. *Bos primigenius* population sizes continued to fall after diverging from *Bos namadicus*. Within the *Bos primigenius* lineage another divergence is observed around 150,000 years ago when Eastern lineages (Baikal1, Gyu1, Yen1) continue to decrease in *Ne* while the Western lineages begin to recover steadily. Eastern *Ne* continues to fall until around 100,000 years ago when it begins to climb rapidly, eventually surpassing Western lineages.

All Western lineages depict a sharp collapse in *Ne* starting around 35,000 years ago. These lineages continue to fall in *Ne* through the Last Glacial Maximum.

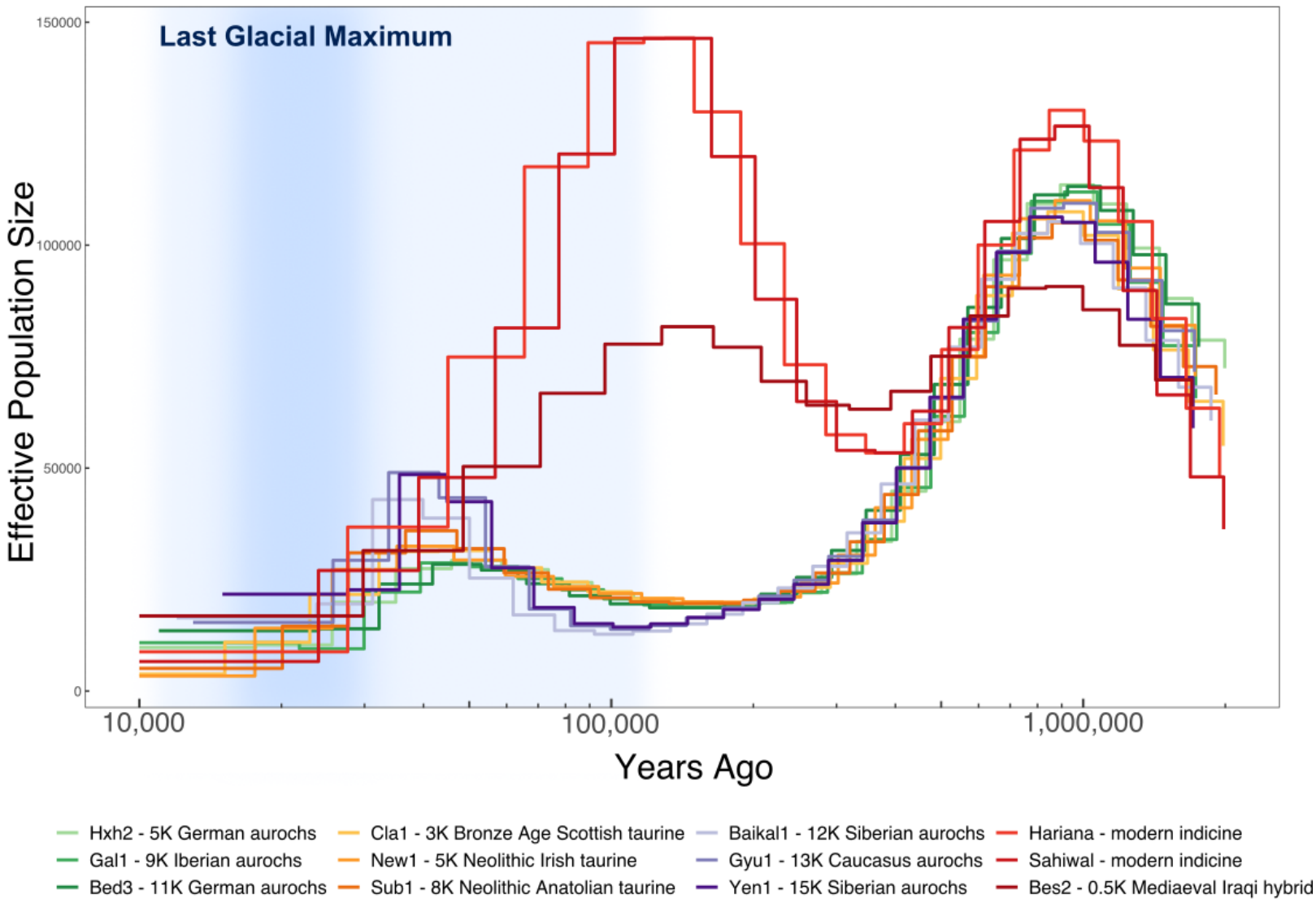


Figure 9.27 - PSMC' curves of 12 ancient and modern lineages of *Bos*.

9.5. Discussion

9.5.1. Relationship between *Bos primigenius and Bos namadicus*

Since the taurine and indicine mitochondrial bifurcation was first described in the early 1990s, the deep divergence between these subspecies has been fundamental to understanding *Bos* genetic variation (Loftus et al. 1994). Modelling demographic history using MSMC similarly depicts a deep separation that began between their progenitor species, *Bos primigenius and Bos namadicus*, at least 300,000 years ago. A clear separation of *Bos primigenius* versus *Bos namadicus* is also depicted with PCA, IBS and model-based ancestry estimates, suggesting that a degree of genetic separation of these subspecies has been long lasting and persistent. However, while the deep divergence of *Bos primigenius* and *Bos namadicus* is evident, this does not exclude the possibility that gene flow between different populations of the subspecies may have occurred. Indeed, unequal allele-sharing was revealed by *D* statistics between modern indicine cattle and different ancient populations of aurochs. Modern indicine breeds displayed the greatest allele-sharing with Pleistocene East Eurasian aurochs, followed by early Southwest Asian domesticates. While the Hindu Kush-Himalayan region was likely a strong physical barrier to gene flow between *Bos primigenius* and *Bos namadicus*, this suggests that gene flow occurred posterior to subspecies splitting. The Iranian Plateau, west of the Hindu Kush, was host to aurochs through the end of the Pleistocene and Holocene (Nasab, Shirvani, and Rigaud 2019) and may have been a point of limited gene flow between populations.

While these results are intriguing, it is important to note that they may be confounded by multiple biases in this analysis. In light of East Eurasian allele sharing, a *Bos taurus* genome assembly may have inflated affinity between more divergent lineages (*i.e.* Indicine and East Eurasia). Such a reference bias effect would be less pronounced in European aurochs, due to their closer genetic identity to *Bos taurus*. For the Southwest Asian connection, the use of modern indicine cattle breeds to represent *Bos namadicus* is potentially problematic. Previous analyses have shown a minor contribution of Southwest Asian taurine ancestry within modern indicine breeds (Kumar et al. 2003; Verdugo et al. 2019). Thus, modern breed formation rather than actual gene flow between pre-Neolithic populations of aurochs may be producing the higher affinity between Southwest Asia and indicine. To date, no ancient sequencing data has been published on the wild ancestors of indicine cattle, likely due to poor aDNA preservation in this region. This has been a noted issue in human aDNA study, where, in one example, only one of 61 South Asian samples yielded sufficient DNA to make genetic inferences (Shinde et al. 2019). As the last known *Bos namadicus* fossil is dated to approximately 3,800 thousand

years ago in Northern India, whole genome sequencing of wild, predomestic representatives is unlikely (Chen et al. 2010). However, Southern Indian breeds display less allele sharing (Kumar et al. 2003; Strucken et al. 2021) and WGS of these breeds may better delineate the natural histories of this subspecies (Sharma et al. 2015).

9.5.2. Pleistocene *Bos primigenius* structure

Populations of Eurasian aurochs appear to have enjoyed a relatively stable period of growth from the onset of the Last Glacial Period according to MSMC analysis. No ROH were detected in pre-LGM Rhi1 and Tula1 aurochs which further suggests that this was a period of large census size for Eurasian aurochs. However, several lines of evidence suggest that non-indicine *Bos primigenius* populations were highly structured by the end of the Pleistocene due to long term isolation by distance. PCA, model-based ancestry estimates and allele-sharing patterns most prominently differentiate Late Pleistocene/Early Holocene aurochs autosomes along an East-West axis (Figure 9.7; 9.12; 9.22-24). This pattern is further emphasised in uniparental markers, where Y chromosomal and mitochondrial lineages diverged in East and West Eurasian aurochs prior to the LGM (Chapters 7 and 8). The MSMC curves similarly depicts a pre-LGM divergence between East Eurasian (Yen1, Gyu1, Baikal1) and West Eurasian (Gal1, Bed3, Sub1) genomes. This divergence occurred around 150,000 years ago, preceding the divergence of East Eurasian (K,C) and Western Eurasian (PQT) mitochondrial haplotypes, approximately 77,000 years ago (median node age = 76,607 yrBP, HPD 95% = 64,961 - 89,840 yrBP).

Long term genetic separation between East and West is reflected in the respective autosomal affinities of three samples dated to pre-LGM in this dataset: Rhi1 (47,000+ yrBP), Rhi2 (~46,000 yrBP) and Tula1 (~41,500 yrBP) (Figure 9.7; Figure 9.10). In ADMIXTURE analysis, these samples reiterate the East-West division of *Bos primigenius* in spite of their relatively close proximity and age (Figure 9.28 below). A physical barrier to prevent gene flow between these populations is not obvious. It is possible that the reduced gene flow in aurochs in northern Eurasia may be due to habitat shrinkage with the onset of the glacial period, reflecting the lower tolerance for colder conditions of *Bos primigenius* than cold-adapted *Bison priscus.* The latter may have prevailed in Eastern Europe, creating a buffer zone between eastern and western populations (Kosintsev and Bachura 2013). This pattern of pre-LGM genetic isolation contrasts with the substantial gene-flow throughout Eurasia within other Pleistocene megafauna, such as horses and wolves (Orlando 2020; Bergström et al. 2022). However, despite this genetic structure in pre-LGM aurochs populations, limited gene flow may be observed. The Rhine samples appear to have a greater affinity to Pleistocene East Eurasia

compared to Holocene European aurochs, as observed in PCA, IBS distance and the presence of a minority Eastern ancestral component in ADMIXTURE (6-14%; Figure 9.28 below). Similarly, *D* statistics suggest that the oldest post-LGM aurochs (Pal1 - ~16,000 yrBP, Italy) shares significantly more alleles with Eastern aurochs compared to early Holocene aurochs. Long term genetic isolation through the LGM increased the genetic distance between East and West populations.

9.5.3. Decline and refugia of the Last Glacial Maximum

During the Last Glacial Period, but prior to the LGM, the *Ne* of Eastern and Western aurochs recovered steadily, especially for Eastern aurochs. From around 30,000-40,000 years ago, however, *Ne* fell sharply for all modelled populations of aurochs, reaching a low-point during the LGM. This period was characterised by decline and eventual extinction of dozens of megafaunal species, such as the straight-tusked elephant (de Carvalho et al. 2020; Barnosky et al. 2004). However, it is possible that the effect of the LGM was less pronounced in East Eurasia, as glacial ice reached further south in western Europe compared to eastern Europe and western Asia (Patton et al. 2017). Despite this, the effect of this cooling period still likely reduced warm-adapted species ranges as evidenced by the southward retreat of East Asian temperate forests due to encroaching tundra (Qian and Ricklefs 2001). The range of East Eurasian aurochs also receded - most notably in Central Asia, where no aurochs remains have been recovered during the LGM (van Vuure 2005). As discussed in Chapter 7, East Eurasian aurochs populations may have retreated to East Asian and Caucasian refugia based on the dispersal of K and C haplotypes in the Late Glacial. However, East Eurasian aurochs display a notable autosomal genetic continuity from pre-LGM until the Late Glacial despite potential range fracturing. Pleistocene East Eurasian aurochs cluster together over wide geographic (~meridian 38° east to ~meridian 110° east) and temporal ranges (~41,500 yrBP to ~12,000 yrBP). This is suggestive of lasting gene flow throughout East Eurasia during the LGM or that the vast area of East Eurasia was rapidly recolonised post-LGM from a single refugial area. The latter framework of recolonisation is observed in the collared lemming, which may have recolonised Holocene Eurasia from an Eastern refuge (Palkopoulou et al. 2016; Lord et al. 2022). In extant Eurasian brown bear populations, it remains equivocal whether northern Eurasian brown bears expanded from a single small refuge area or if they survived in multiple fractured areas (Davison et al. 2011).

Glacial refugia are better characterised in Europe, where the Weichselian ice sheet expanded across northern Europe and much of central Europe was covered in a permafrost tundra (Svendsen et al. 2004). Along with many warm-adapted species, European aurochs retreated

south to the forest and steppe refugia in France-Iberia, Italy and the Balkans during the LGM (Leonardi et al. 2022). Like East Eurasia, continuity of ancestry is observed in Europe through the LGM, although with more evidence of genetic flux. In particular, allele-sharing with Southwest Asian aurochs populations greatly increased in post-LGM European aurochs compared to Pleistocene Rhine Valley aurochs. The cline of such ancestry is correlated to the distance from Southwest Asia - peaking in the Balkans and steadily decreasing toward Iberia and Britain (Figure 9.15). This suggests there was a diffusion of Southwest Asian ancestry westward through Europe after ~46,000 yrBP (the approximate age of Rhi2). The Late Glacial Italian aurochs Pal1 also modelled a small Southwest Asian component suggesting that this introgression began pre-Holocene. This intrusion of Southwest Asian ancestry into Europe may be connected to the replacement of the divergent Rhine mitochondrial haplotypes (N and G) by P after the LGM. TMRCA of the European P haplotype and the Southwest Asian haplotype (T and Q) is estimated just prior to the LGM (median node age = 45,183 yrBP, 95% HPD = 35,415-55,345 yrBP).

9.5.4. Postglacial expansion in Europe

The climatic amelioration of the Late Glacial saw many warm-adapted species, that had retreated to southern refugia, expand and recolonise northern territories (Hewitt 1999). For example, in brown bears, Europe was mostly recolonised from Iberia as the Alps acted as a barrier to gene flow from Italy (Davison et al. 2011; Bray et al. 2013). In Europe, recolonisation routes have been characterised for several terrestrial mammals using archaeological remains, extant genetic variation and aDNA. While tracking P mitochondrial lineages did not resolve refugial structure, autosomal variation-based NJ trees and PCA identified putative clusters corresponding to each southern European refugial peninsula (Iberia, Italy, Balkans; Figure 9.1 and Figure 9.11). These geographic clusters appear to be principally controlled by the extent of influence from Southwest Asia, with early Holocene Iberian animals displaying the least affinity toward Southwest Asia in PCA space (Figure 9.15B). Early Scandinavian, German and British aurochs also cluster with these, which strongly implicates Iberia as the major source of northwestern European recolonisation. More recent animals of Germany, Britain and Scandinavia (Hxh2, Fal7, YoA, Gallo1, Ton1, Lil1) drift from this Iberian cluster, with some occupying a PCA space with Italian aurochs. However, as these aurochs existed at a time contemporaneous with domestic cattle, this is probably influenced by domestic introgression. Domestic introgression into the wild is explicitly evident in the case of the British aurochs YoA which harbours a domestic haplotype (T3) imported to Europe by early farmers.

Eastern European aurochs cluster with Iberian aurochs in IBS trees but are similarly pulled by an apparent Southwest Asian affinity in PCA. These are post-domestic animals and domestic introgression is explicitly evidenced by the presence of a domestic Q haplotype in Kalisz2, a ~4,500 year old Polish sample. Without a representative of pre-domestic contact context, it is difficult to discern if Balkan and/or Italian populations have influenced these Eastern European genomes. Taken together, Iberia was the likely major point of postglacial recolonisation for Central Europe, Britain and Scandinavia with Italy and the Balkans possibly influencing Eastern Europe.

9.5.5. Holocene migration and hybridisation

Central Eurasia may have been an important barrier to gene flow throughout the Last Glacial Period giving a strong population structure along an East-West axis. In particular, Central Asia was unoccupied by aurochs during the LGM (van Vuure 2005). However, postglacial climatic amelioration facilitated the expansion of European aurochs into Central and East Asia. This has been evidenced by detection of the European P mitotype in modern East Asian breeds, which was likely recruited by early cattle farmers in Asia (Mannen et al. 2020). Mitochondrial and Y chromosomal sampling of aurochs confirm European ancestry in Central Asia but also reveals Southwest Asian and East Eurasian haplotypes (Chapter 7 and 8). A hybrid zone also emerges in PCA analysis, where Central Asian aurochs occupy a space intermediate between the three axes of aurochs diversity (Europe, Southwest Asia, East Eurasia). Similarly ADMIXTURE analysis models most Central Asian aurochs as a composite of the three *Bos primigenius* populations (Figure 9.28). Two animals, Bor1 and ROS002, are modelled with completely East Eurasian ancestry but still occupy a hybrid position in PCA (Figure 9.10). Also, ROS002 harbours the European P haplotype. In addition to Central Asia, the Southern Caucasus also emerges as a hybrid zone. Gyu2, a ~7,000 year old Armenian aurochs is similarly modelled as a composite of three distinct ancestries.

While these animals are composites of three ancestral populations, it is noteworthy that the relative contribution of each is not equal and appears to be influenced by both geography and time (Figure 9.26). The most westerly Central Asian aurochs in this dataset (Var1, Var2) are modelled with the largest amount of European ancestry and those from 1,500km further east, Borly and Roschinskoe, are modelled with much less. The proportion of ancestry can fluctuate within the same site. The Caucasian Gyu2 is also modelled with a large component of European ancestry. While this ancestry may have diffused from Southeast European aurochs moving across the Bosphorus straits to reach the Caucasus, the ancestral profile of Ch22 contravenes this. Ch22, an Anatolian aurochs contemporaneous to Gyu2, has a lower

proportion of European ancestry despite its close proximity to Europe. This suggests that European populations may have migrated from northeast of the Black Sea through the Pontic Steppe to influence the ancestry of the Holocene Caucasus. The influx into Gy2 (~7,000 yrBP) occurs in the Holocene as the Caucasus transforms from being modelled as completely East Eurasian in the Late Glacial (Gyu1, ~13,000 yrBP) (Figure 9.28). The turnover of ancestry may also point to an extirpation that occurred between ~13,000 and ~7,000 years ago. The Younger Dryas was a sudden cold period that occurred just prior to the Holocene and led to an arid desert or semi-desert steppe in parts of South Caucasus (Messager et al. 2013). Southwest Asian ancestry appears to have contributed more equally to Central Asian aurochs in this dataset compared to the varied European contribution. Human movement of domestic Southwest Asian cattle is not implicated in this as domestic cattle do not appear in the Pontic-Caspian Steppe area or west Kazakhstan until ~5,500 yrBP and ~4,500 yrBP (Wilkin et al. 2021; Hermes et al. 2019; Outram et al. 2012).

Two aurochs of unknown origin harbouring a domestic T3 allele are included in this dataset, Tango1 and Tango2. Tango1 displays an ancestral profile that is remarkably similar to that of Varfolomeevka. With roughly equal proportions of ancestry from Europe, Southwest Asia and the East and sampled from somewhere in the proximity to the former Soviet Union, Tango1 may come from the Pontic-Caspian Steppe. Tango2 derives most of its ancestry from European aurochs based on PCA clustering and ADMIXTURE modelling, suggesting it is of more western origin.

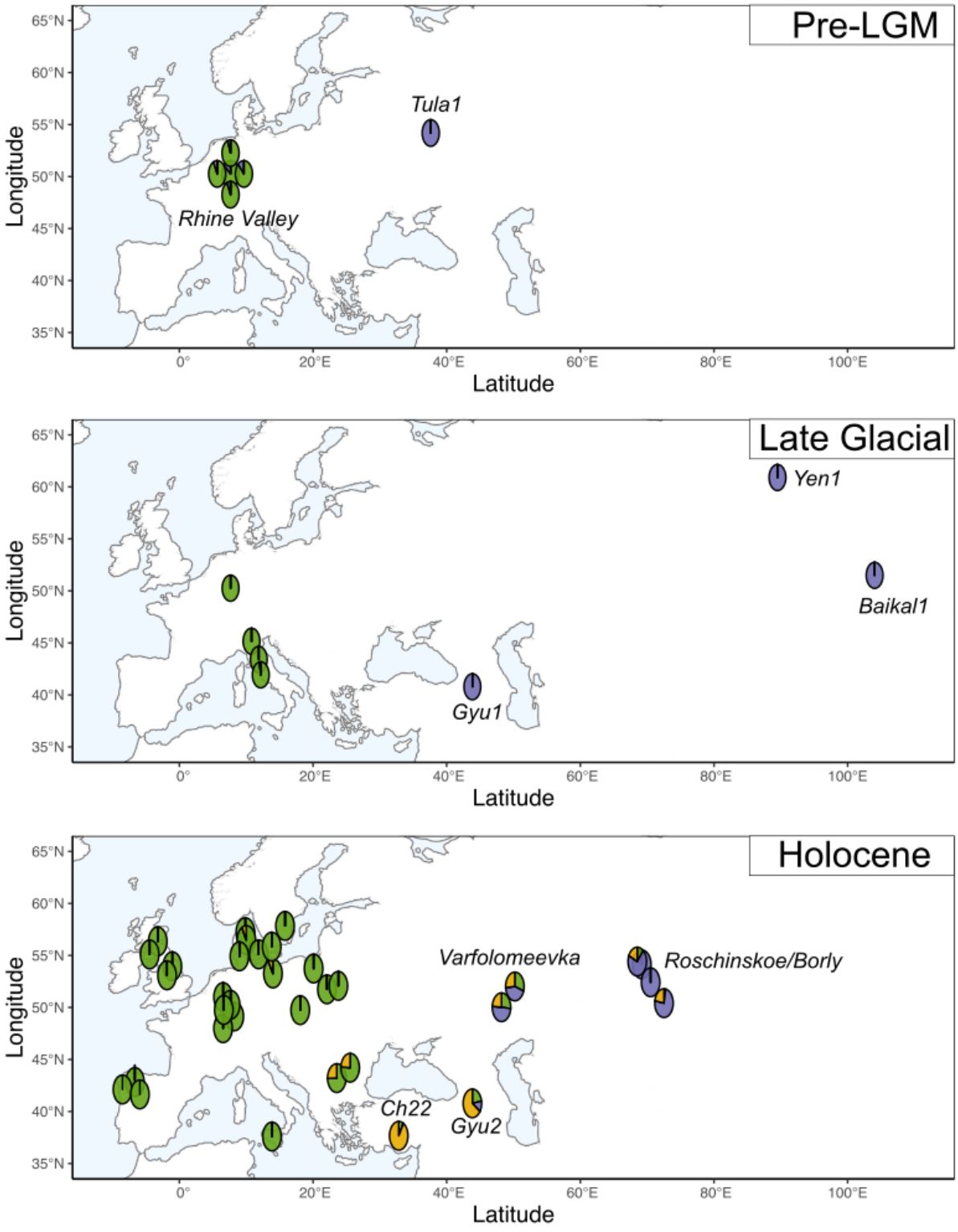


Figure 9.28 - Turnover of ancestries across Eurasia, as modelled by ADMIXTURE K=4.

9.5.6. *Bos taurus* domestication in Southwest Asia

A traditional view has been held that taurine cattle were first domesticated around 10,500 years ago in the Upper and Middle Euphrates Valley and subsequently moved with people as part of a Neolithic package. This has been emphasised by genetic studies that describe a small starting stock of females and a restricted Northern Fertile Crescent ancestry, diffusing from Southwest Asia (Verdugo et al. 2019). However, this view of a single cattle domestication origin in Northern Fertile Crescent has recently been critiqued based on inconsistent appearances of traditional markers of domestication, such as decreased sexual dimorphism (Arbuckle and Kassebaum 2021). The authors argue that cattle were instead domesticated from multiple wild Neolithic Southwest Asian populations, similar to early goat management (Daly et al. 2018). Within this dataset, the ability to dissect early Southwest Asian domestic ancestry is greatly impeded by a lack of sequencing data of pre-Neolithic aurochs variation. However, the Southern Caucasian wild Gyu2 and Anatolian wild Ch22 depict clearly different ancestral profiles in *D* statistics compared to Anatolian (Sub1, Men1, Men2) and Iranian (Sac3, Far1, Mar1) Neolithic domesticates, which appear much more uniform in ancestry. Similarly, as discussed in Chapter 7, the expansion of the Southwest Asian haplotypes (T and Q) are consistent with a very limited starting stock of domestic taurines. Taken together, the current data reflects a strong case for a restricted source of domestic cattle ancestry. While this does not preclude multiple centres of aurochs management in Southwest Asia, a framework of a single *Bos taurus* domestication event is best supported by current genetic data.

Some amount of genetic flux is observed in Southwest Asian Bronze Age cattle where a minority Eastern component (~2-3%) appears in the ADMIXTURE profiles of three Anatolian and Iranian genomes (Figure 9.23). This is also observed in increased non-significant allele-sharing with Eastern aurochs. Such patterns are not observed in Neolithic Anatolia or Iran, suggesting that a post-domestication recruitment of Eastern ancestry may have occurred. Based on the ancestral profile of aurochs Gyu2, Eastern ancestry had reached the South Caucasus by the time but also includes significant wild European ancestry, which is not detected in these Bronze Age domesticates. European ancestry decreases significantly in Central Asian hybrid aurochs northeast of the Caspian Sea, where they are modelled as hybrids mostly of Eastern and Southwest Asian ancestries. Therefore, it is possible that Eastern ancestry diffused south toward the Iranian Plateau where it was recruited by the Bronze Age. Modern Southwest Asian breeds are hybrids of taurine and South Asian indicine ancestry from introgression first occurring in the Near East around 4,000 yrBP, around the 4.2k dry climatic event (Verdugo et al. 2019). After this major change, the minority component of

Eastern ancestry is retained in Iranian and Iraqi Iron Age samples and in Mediaeval Caucasus genomes.

9.5.7. Local incorporation of *Bos primigenius* into domestic stock

African Cattle

While the ancestry of modern African taurine breeds is thought to be mostly derived from Southwest Asia, local aurochs incorporation gives a distinctive ancestral pattern (Decker et al. 2014); (Pitt et al. 2019). This is reflected in IBS clustering and allele sharing patterns of West African breeds with Th7, an ~9,000 year old Moroccan aurochs (Figure 9.3). An early Levantine domesticate (Tsa3) also displayed a higher affinity toward African taurines than their ancient Anatolian counterparts. This suggests that back-flow of African ancestry into early Levantine domesticates occurred or that Levantine aurochs (that had a North African affinity) were incorporated. A connection between West African taurine breeds and East Eurasian aurochs is hinted at in PC3 and ADMIXTURE analysis, where they are modelled at 3 - 5%. However, it may be a case that ancestral alleles shared between East Eurasian and African aurochs ancestry is modelled as Eastern due to the lack of African representatives. Low coverage restricted ADMIXTURE modelling of Th7 but IBS analysis indicated this aurochs as an outgroup of European and Southwest Asian aurochs. A substantial imprint of indicine introgression is evident on the African continent and has been previously described (Freeman et al. 2004; Mwai et al. 2015). This has been enhanced by centuries of trade between Swahili and Arab civilisation and an 1890 epizootic of rinderpest which devastated African cattle populations.

European Cattle

The earliest farmers arrived into the Balkans via the Bosphorus Strait around 9,000 yrBP (van Andel and Runnels 1995). PCA and ADMIXTURE analysis depict a persistence of Southwest Asian ancestry as the Neolithic moves in Southeast Europe. However, most ancient Europeans cluster slightly away from the early Southwest Asian domesticates in the IBS analysis and PCA. *D* statistics reveal that this is likely caused by incorporation of wild European variation, as has previously been reported (Verdugo et al. 2019). However, this wild recruitment was not uniform across Neolithic Southeast Europeans. In particular, the earliest Southeast European Neolithic animals (Bla1, Bub1, Sar38) and a later Bulgarian (Che1) do not display excess allele sharing with European aurochs, suggesting that significant incorporation of wild variation may have been staged. Incorporation of wild ancestry is apparent in many European breeds today. In ADMIXTURE analysis, this component peaks in British and Irish breeds (Scottish Highland, Devon, Belted Galloway, Dexter), as previously described (Park et al. 2015). As discussed in

Chapter 8, the early farming management of incorporation of European aurochs bull on the island of Britain may have directly resulted in a high proportion of European ancestry. Iberian breeds also display a consistent European wild signal, as previously detected (Upadhyay et al. 2017).

Central and East Asian Cattle

A continuity of ancestry is detected in Central and East Asian cattle, driven by ancient incorporation of local wild East Eurasian variation. Shi1, a ~4,000 year old domestic cow from western Kazakhstan, close to the Altai Mountains, has an ancestral profile that comprises 16% ancestral component of Eastern aurochs. A commonly held view is that pastoralism was brought to Kazakhstan by Yamnaya herders from the Pontic Steppe between 5,500 to 4,700 years ago (Damgaard et al. 2018; Anthony 2010). This is supported by ancient human genetic evidence which depicts a clear connection between herders in Chalcolithic/Bronze Age Pontic Steppe and the first pastoralists in Kazakhstan (Damgaard et al. 2018; Anthony 2010). Notably, aurochs from the Pontic Steppe (Var1, Var2) had considerably more European ancestry than Shi1. This suggests that the recruitment of wild variation in Shi1 did not occur in the Pontic Steppe but further west, where European ancestry decreases in the wild (*e.g.* Bor1, ROS002). Alternatively, the lack of substantial European ancestry in Shi1 may be due to an alternative dispersal route to Kazakhstan from Iran. The arrival of pastoralism in western Kazakhstan has also been proposed to arrive via the Inner Asian Mountains Corridor (IAMC) west of the Caspian Sea (Hermes et al. 2019). This movement of agricultural technologies is assumed to have been by cultural diffusion rather than large-scale human migration. As Iranian Bronze Age cattle (Has3, Qaz1) are modelled with a minority component of East Eurasian ancestry, it is possible that Shi1 is derived from domestic cattle populations that moved from the Iranian Plateau along the IAMC. This framework is challenged somewhat when considering that Post-Bronze Age Uzbekistani hybrids Yer1 and Kok1 are not modelled with Eastern ancestry but Yer1 harbours a European P haplotype. Therefore, without direct sampling of Bronze Age cattle from the Pontic Steppe and regions west of the Caspian Sea, the ancestry of Shi1 supports both putative dispersal routes of farming to the Altai Mountains.

Cattle first appeared in Northern China during the late Neolithic, around 4,000 to 5,000 years ago (Cai et al. 2014). From the Altai Mountains, both the Hexi corridor, which cuts a path between the Mongolian and Tibetan Plateaus, and a more northern route through the Eurasian steppe have been proposed as routes of dispersal of cattle into East Asia (Brunson et al. 2020). Similar to Shi1, the Neolithic Chinese domesticate Shimao displays a large component of Eastern ancestry (~13%), emphasising an ancestral link between Central and East Asian

cattle. This Eastern component persists in almost all modern Central and East Asian cattle, with many breeds (Yanbian, Hanwoo, Kuchinoshima) showing levels of wild Eastern ancestry similar to that of Shimao and Shi1. Tibetan breeds display higher levels of Eastern ancestry (~30%), perhaps signalling further wild introgressive events. Indicine cattle were introduced to southern China from South Asia situated further west approximately 2,500 years ago (Higham 1996). It is apparent that modern East Asian hybrids derive their indicine ancestry from these East Asian indicine cattle based on affinities in PC6. The taurine component of these hybrids is also derived from East Asian taurines based on the persistence of an Eastern minority component modelled at 1-5%.

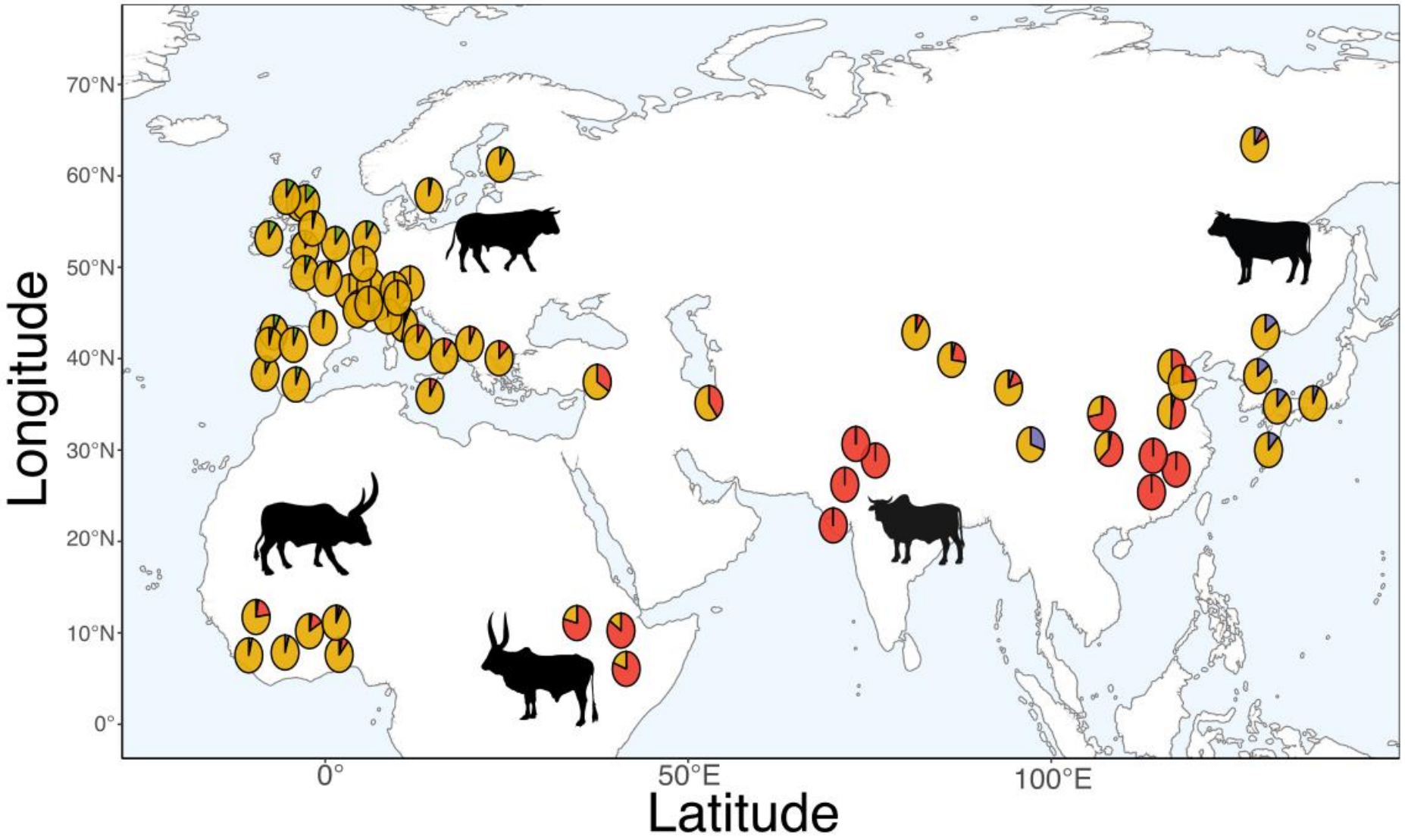


Figure 9.28 - Ancestral components of modern breeds as modelled by ADMIXTURE, K=4.

9.5.8. The last stand for aurochs

The ultimate fate of the aurochs presents a unique paradigm as, despite their extinction in the wild 400 years ago, approximately 1.5 billion domestic forms exist today. While the relative impact of climatic changes and anthropomorphic effects on Pleistocene megafaunal extinctions in Northern Eurasian are still debated (Stuart 1991; Barnosky et al. 2004; (Sandom et al. 2014), it is thought that the extinction of the aurochs was mostly precipitated by hunting and habitat destruction, with climate change having a minor impact. In Eurasia, megafaunal extinctions were generally staggered and more gradual than Australia or America, perhaps due to the long-term occupancy of *Homo* here (Sandom et al. 2014). The sudden post-LGM climate change in Eurasia caused population decline and range reduction for cold-adapted species, such the woolly mammoth (Palkopoulou et al. 2013) which may have left these species vulnerable to expanding human hunting populations. While a reduction of *Ne* was observed in aurochs during the LGM, the changing climate of the Holocene resulted in a sizable niche expansion for aurochs across Eurasia and perhaps a recovery of maternal effective population size in Europe (Chapter 7). It is unlikely that climate change at the Holocene is significantly associated with the extinction of aurochs; multiple anthropogenic effects associated with agricultural revolution are more likely drivers.

Populations of European humans exploded after the invention of mixed agriculture around 8,000 years ago and led to dramatic alteration of habitat (Gignoux, Henn, and Mountain 2011). Some European species benefited from this transformation of landscape, such as the pink-footed goose (Pujolar et al. 2017), but this likely created a strain on aurochs populations. Through clearing of dense forest to make way for agricultural land, (Kalis, Merkt, and Wunderlich 2003; Woodbridge et al. 2014). populations of aurochs became neighbours and competitors with domestic cattle (van Vuure 2005). After domestication, hunting of aurochs also continued (Turvey et al. 2013); Pöllath et al. 2018; Marom and Bar-Oz 2013), creating a mounting threat to aurochs as human populations continued to grow. Gene flow between domestic and wild is evident with explicit evidence of domestic input by the existence of domestic mitochondrial haplotypes in Tango1 (T3), Tango2 (T3), Kalisz2 (Q) and YoA (T3). While pastoralists may have used the wild intermittently to restock their domestic herds, dwindling wild populations may have also forced aurochs to interbreed with domestic cows. For example, the genome of Lil1 reveals clear signs of both inbreeding and input from domestic lineages. However, this is not evident in the Neolithic British aurochs YoA, which displays no excess of ROH and retains a strong wild ancestry profile despite its mitochondrial introgression.

Through the ages, management of domesticated cattle intensified, while aurochs populations dwindled (Manning et al. 2015; van Vuure 2005). The last stand of aurochs occurred in the Polish Middle Ages, where the last of the aurochs were pushed into the forests of Jaktorów. These extremely small populations were vulnerable to disease from nearby domesticates and likely inbred (van Vuure 2005). Notably, the most recent aurochsen in this dataset (Gallo1, Fal7, Horns) all display unusual allele-sharing with groups of aurochs compared to other post-domestic contact NICHE aurochs, possibly as a result of extreme inbreeding. In the 16th and 17th Century, opportunistic hunters, foresters and farmers took advantage of the poor Royal protection of the last aurochs. The very last aurochs cow is said to have died in 1627, however as the genome of the last aurochs bull (Horn-11570) has a clearly domestic ancestral profile, the aurochs may have actually died out earlier than this. The last confirmed aurochs bull is therefore presented here in Horn-CMLIX, which had an inscribed age of 1604.

10. Conclusions

10.1. Part I: Retrieval of DNA from ancient materials

Part I of this thesis explored the sampling and sequencing of ancient material for aDNA analysis, including the characterisation of DNA damage in unconventional ancient substrates.

10.1.1. Preparation, Screening and Sequencing of Ancient and Historic Materials

Chapter 3 presents the screening data of 285 historical and ancient samples, offering a large dataset on aDNAn retrieval. Technological advances have transformed the field of aDNA since the first ancient fragments of DNA were sequenced using a museum specimen almost 40 years ago (Higuchi et al. 1984). Despite these, substantial obstacles that remain are low endogenous DNA content and *post-mortem* damage (Pääbo et al. 2004). These were explored by employing a pre-extraction bleach digestion step to maximise the retrieval of endogenous DNA. While this improved endogenous DNA retrieval in some samples, a delicate balance between maximising endogenous content and reduction of complexity was necessary when deep-sequencing ancient samples.

This dataset also corroborated previous findings that the petrous bone is an excellent reservoir of endogenous aDNA (Gamba et al. 2014). However, there were multiple instances of failure to retrieve endogenous aDNA in petrous bone samples. For example, petrous samples that were recovered from arid and high-temperature climates yielded consistently poor endogenous content, as has been widely reported (Rizzi et al. 2012). However, multiple petrous samples from Sweden also performed poorly despite a relatively cool and stable mean temperature. This emphasises how *post-mortem* DNA decay is still not easily predicted, and likely influenced by a range of taphonomic factors (Keighley et al. 2021). For samples with very low levels of endogenous DNA, several bioinformatic approaches were leveraged for taxonomic assignment. In total, 223 of the 285 samples (~78%) screened were assigned genus-level taxonomy.

10.1.2. Exploration of aDNA preservation in mummified remains

Mummified remains have long attracted interest as a potential source of aDNA (Pääbo 1985). However, mummification is a rare process that requires an anhydrous environment to rapidly dehydrate and preserve tissue prior to decomposition. Sequencing data of MUM2, a ~1,600 year old Iranian sheep that was naturally mummified in a salt mine, was interrogated in Chapter 4. The extent of aDNA preservation in MUM2 was singular in comparison to other ancient

samples from the region. DNA fragmentation and cytosine deamination, hallmarks aDNA damage, were substantially reduced in MUM2. The rapid dehydration of the tissue by the salt cave environment may have slowed these processes as aDNA strand fragmentation and cytosine deamination processes are thought to be driven by hydrolytic reactions. Metagenomic analyses of MUM2 also revealed a unique microbial profile. While mammalian tissues are typically rapidly colonised by saprophytic microbial communities post-mortem (Metcalf et al. 2016), the microbial profile was predominated by halophilic archaea and bacteria. The highly saline taphonomic environment may have had a profound influence on decomposition by restricting growth of saprophytic microbial communities, thus further arresting tissue decay.

MUM2 is also the first published ancient sheep nuclear genome. Consistent genetic clustering of this sheep with modern Southwest Asian breeds suggested a continuity of ancestry in the region since at least the Sasanian period. Genotyping this animal revealed derived alleles associated with the fat-tail phenotype and the presence of an ancestral “hairy” allele (Dong et al. 2020). The unique preservation of MUM2 soft tissue also allowed for hair fibre imaging and analysis in which scanning electron Microscope imaging of the mostly-unpigmented mummified hair fibres revealed mosaic scales with fine lines on the scale surface, consistent with a hairy/medium-wool phenotype (Rast-Eicher 2016).

10.1.3. Preservation, metagenomics and population genomic analysis of ancient Levantine goat coprolites

There is a growing appreciation of the potential that aDNA analysis of coprolites may offer to make genetic inferences about the diet, health and ancestry of depositing species (Shillito et al. 2020). In a pilot study, aDNA was extracted from goat coprolites recovered from a rockshelter in the Negev desert in the Levant. These coprolites were deposited in layers that dated to the late Neolithic, Bronze Age and the recent era (300 yrBP) (Rosen et al. 2005; Rosen 2013). In several samples from each horizon, aDNA was successfully extracted, amplified and sequenced. DNA concentration was low, perhaps due to the limited volume of material available. However, an unexpectedly high level of endogenous DNA was recovered. This success was in spite of typically very low endogenous DNA levels in modern faeces (Perry et al. 2010) and that mineralised ancient samples from this region generally performed poorly in aDNA studies. This is the first aDNA sequencing data generated for ungulate coprolites and the first aDNA shotgun sequencing of autosomal coprolite aDNA. The success in sequencing was attributed to the desiccation of coprolites shortly after deposition, protecting them from invasion of saprophytic decomposers. The arid climate and rockshelter cover further protected these coprolites from DNA leaching.

The metagenomic landscape of these coprolites was also interrogated, producing a detectable ancient ungulate gut microbiome and signals of past diets. Low coverage genomes also provided insight into the prehistory of Levantine animal husbandry by presenting a local time series of goat management in the region. Mitochondrial lineages revealed a continuity of maternal goat ancestry from the Neolithic, through the Bronze Age until present-day in the southern Levant. This was also reflected in autosomal variation, which clustered two Neolithic and one Bronze Age coprolite samples in genetic analyses with Bronze Age Levantine goats. These coprolites displayed an ancestral profile distinct from other Neolithic Levantine goats, suggesting a possible early influence from Eastern Fertile Crescent domestic herds that is detected in Bronze Age Levant (Daly et al. 2018). However, interrogation of damage of the coprolite aDNA revealed patterns of high fragmentation and unusual nucleotide misincorporations which cautioned against firm conclusions.

10.1.4. Next generation sequencing of modern Chromium (III) tanned leather using ancient DNA techniques

Despite long-standing interest in leather DNA, very few studies have reported successful DNA amplification from modern or ancient sources. Lack of success in leather DNA amplification has been attributed to the harsh tanning process which is thought to damage and degrade the DNA by extreme pH oscillations and chemical modification of the collagen network (Vuissoz et al. 2007). In Chapter 6, successful extraction and sequencing of modern DNA from chromium (III) tanned leather was achieved by employing common ancient DNA techniques designed for low-concentration, fragmented and damaged DNA. It was observed that the DNA had been severely damaged by the tanning process, resulting in fragmentation and nucleotide misincorporation. This was possibly caused by chromium (III) interactions with the DNA. Furthermore, an unusual pattern of genomic preservation was observed resulting in an overrepresentation of constitutive heterochromatin and mitochondrial sequences. This enrichment may have been due to protective effects by heterochromatin packing and the double membrane of the mitochondrial genome, respectively. These patterns had previously been detected in ancient parchment DNA (Teasdale et al. 2017). Both leather tanning and parchment production include an alkaline treatment of animal skin. As highly alkaline conditions cause DNA denaturation (Wang, Lim, and Son 2014), protection may be conferred by tight binding to basic histone proteins in heterochromatin. Leather DNA research has potential impact in commercial tracing applications.

10.1.5. Ethical considerations of aDNA analysis

In recent years, the use of aDNA analysis has been instrumental to understanding the complex domestication process (Frantz et al. 2020). However, there is a growing awareness of ethical considerations in the field of aDNA (Alpaslan-Roodenberg et al. 2021). Researchers have a duty of care to the human or animal remains they study. This is especially relevant to work on ancient human populations, where inferences made using aDNA can have profound effects on present-day communities. However, consideration in aDNA should also extend to the physical destruction of irreplaceable archaeological material (Pálsdóttir et al. 2019). These considerations extend to detailed sample curation, high-quality sequencing data creation and ultimately publication of both positive and negative results. Of the 285 samples screened here, 108 yielded endogenous DNA content below 1% which greatly restricts feasible genetic analysis. Therefore, it is advisable to initially sample a small proportion of the material available from any new site. Negative results should be published to inform future sampling strategies and avoid needless destruction of limited archaeological material. Less destructive sampling methods have also proven successful for ancient substrates other than teeth and bone and may offer alternative sampling strategies for select time periods (Teasdale et al. 2017). Using comparable modern materials can inform methods to best extract aDNA and characterise damage prior to sampling and destruction of scarce ancient material. Also, 3D scanning of archaeological models prior to sampling now offers a digital archive of the samples that is usable downstream morphometric analysis.

10.2. Part II: The demography of the extinct aurochs

Part II of this thesis conducts a large-scale genomic survey of the extinct aurochs. While the deep split of taurine versus indicine has been well-characterised, little is known of the structure of Eurasian aurochs in the Late Pleistocene. This is addressed through analysis of 153 ancient *Bos* genomes.

10.2.1. Genetic structure of Eurasian aurochs in the Late Pleistocene

PCA, model-based ancestry estimates and allele-sharing patterns differentiate Late Pleistocene/Early Holocene aurochs along an East-West axis. Reconstruction of the population history of these aurochs using MSMC analysis depicts a clear split in Eastern and Western ancestry occurring prior to the LGM; population structure is observed in three pre-LGM aurochs that were radiocarbon dated to at least 41,000 years ago. This divergence was also reflected in uniparental markers. A new Eastern mitochondrial haplotype was described (K), first discovered in an Armenian aurochs, while the previously described C haplotype was

dispersed across Siberia. Western mitochondrial groups include the PQT haplogroup, which was widely dispersed across Europe and Southwest Asia in the Holocene, while the R haplotype found in Northern Africa and the E haplotype found in Eastern Europe. Bayesian dating of Eastern versus Western lineages occurred approximately 77,000 years ago (median node age = 76,607 yrBP, HPD 95% = 64,961 - 89,840 yrBP). Meanwhile, in Y chromosomes, a new Y4 clade is well-defined in Eastern aurochs which also splits from Western aurochs (Y2 and Y1) pre-LGM when population structure may have been driven by declining climatic conditions creating a barrier to gene-flow in Eastern Europe.

In Western aurochs, a continuity of Y chromosomal ancestry is observed through the LGM. In contrast, a dramatic turnover of European maternal ancestry is observed, where the divergent pre-LGM G and N haplotypes are replaced with the P haplotype. This turnover co-occurs with an influx of Southwest Asian autosomal ancestry into Europe. As TMRCA of the European P haplotype and the Southwest Asian haplotypes (T and Q) is estimated just prior to the LGM (median node age = 45,183 yrBP, 95% HPD = 35,415-55,345 yrBP), this diffusion of ancestry may have occurred just prior to the LGM. The maternal and autosomal variation of European and Southwest Asian aurochs display clear geographic structure by the Holocene transition. The Y chromosomal geographic structuring across Europe and Southwest Asia is comparatively weak. In Eastern aurochs, there is coincident geographical structure emerged in uniparental markers. The Siberian Y4a is harboured in mtDNA C aurochs, while a distinct Caucasian Y4b emerged in K aurochs. Despite this distinction between Eastern aurochs, autosomal variation suggests that they were highly connected through the Late Pleistocene. Geographic structure also emerged in post-LGM European autosomal variation and is perhaps related to Southwest Asian migration and LGM-driven genetic isolation between different southern *refugia*.

10.2.2. Aurochs response to postglacial Eurasia and domestication

Through the LGM, populations of Eastern and Western aurochs declined in *Ne* as estimated by MSMC. However, the postglacial retreat of tundra conditions offered a niche expansion across Eurasia. In Europe, aurochs likely expanded from an Iberian refugia to recolonise much of Northwestern Europe and is reflected in a recovery of European maternal effective population size. From their respective refugia, European, Southwest Asian and Eastern aurochs co-expanded into Central Asia, a region that was uninhabited by aurochs during the LGM. Shortly after the Holocene transition, human management of aurochs began in Southwest Asia. A restricted autosomal profile is observed in Neolithic domesticates which were subsequently moved into Europe and Central Asia by migrating farmers. The expansion

of domestic maternal lineages of these animals is consistent with a small starting stock of related aurochs cows. Similarly, a distinctive Neolithic domestic Y chromosome emerges. The dispersal of domestic cattle across Europe and Asia introduced new neighbours to local populations of aurochs and domestic influence on the wild population is evident by multiple instances of domestic maternal haplotype introgression.

10.2.3 Legacy of local aurochs populations in modern domestic cattle

Most of the ancestry of modern taurine cattle is derived from the Southwest Asian aurochs domesticated in the Fertile Crescent. However, incorporation of local wild variation was widespread in early cattle domestication and remains a defining aspect of breeds today. Incorporation of wild ancestry into domesticates occurs early in Neolithic Southeast Europe and persists in European breeds. Wild introgression is thought to have been mostly male-mediated due to the retention of the original domesticate maternal lineages (Park et al. 2015; Orlando 2015). Multiple instances of ancient wild Y recruitment and turnover are observed. A local wild bull recruitment is likely in Neolithic Britain and the influence of European aurochs peaks in British and Irish breeds. Some maternal incorporation is observed with the establishment of the divergentR haplotype in Italian breeds. In Central Asia, incorporation of Eastern wild ancestry is detected in the Early Bronze Age. A similar level of Eastern ancestry is then observed in Neolithic Chinese cattle, suggesting a connection between these populations. The modern East Asian taurine Y chromosome may be a direct ancestor of the aurochs bulls that were domesticated in the Fertile Crescent. However, maternal wild ancestry dated to Bronze Age Central Asia is harboured in East Asian breeds.

10.2.4. Future directions

This large genomic survey of *Bos primigenius* has illuminated aspects of evolution, domestication and eventual extinction of this megafaunal Pleistocene animal. While not possible within the time constraints of this thesis, this dataset may be used to interrogate selection signatures that occurred along the first steps of cattle domestication, as has been shown in other domesticates (Librado et al. 2017; Daly et al. 2018; Frantz et al. 2019). Additionally, it has previously been suggested that local wild restocking of cattle may have been a purposeful strategy of early farmers, possibly to acquire new advantageous traits that would adapt the imported Anatolian cattle to their new surroundings (Park et al. 2015; Orlando 2015). Leveraging the high quality genomes generated in this study, candidate genes of adaptive introgression may be searched for in modern cattle which still retain this recruited local wild ancestry (Racimo et al. 2015).

This dataset represents a valuable resource for the future of cattle research. For over a decade the 1000 Bulls Genome Project has been curated to determine genetic changes that are related to animal health and production in cattle (Hayes and Daetwyler 2019). One of the long term goals of this project is to produce a global cattle reference to impute the thousands of cattle genotyped on low-cost SNP chips to phased, diploid genomes. This wild dataset contains several genomes that may improve quality of imputation, especially in rare breeds that may harbour wild variation uncharacterised in the modern reference dataset. Imputation has also proven successful in a range of low coverage ancient human genomes (Gamba et al. 2014; Martiniano et al. 2017; Cassidy et al. 2020). This dataset may be valuable if such methods are implemented in ancient cattle genomics, which would enhance cost efficiency.

Further graphical and temporal sampling may enhance the understanding of *Bos primigenius*. As discussed, the large geographic range of aurochs in the Late Pleistocene and Early Holocene means that there may be many more hybrid zones that have not yet been described. It has also been acknowledged that South Asian and African aurochs remain a deficit, but poor aDNA preservation of these regions will likely impede future sequencing efforts. Finally, an intriguing question that remains unanswered is the origin of the divergent N and G haplotypes in the Pleistocene Upper Rhine Valley aurochs. The tentative connection between these aurochs and VEM212, the oldest aurochs yet sequenced, necessitates further sampling of pre-LGM Europe to disentangle the fascinating and complex history of the aurochs.

11. References

Aali, Abolfazl, and Thomas Stöllner. 2015. *The Archaeology of the Salt Miners: Interdisciplinary Research 2010-2014*. Bochum: Deutsches Bergbau-Museum.

Aali, Abolfazl, Thomas Stöllner, Aydin Abar, and Frank Rühli. 2012. “The Salt Men of Iran: The Salt Mine of Douzlākh, Chehrābād” 42 (1): 61–81.

Achilli, Alessandro, Silvia Bonfiglio, Anna Olivieri, Arianna Malusà, Maria Pala, Baharak Hooshiar Kashani, Ugo A. Perego, et al. 2009. “The Multifaceted Origin of Taurine Cattle Reflected by the Mitochondrial Genome.” *PloS One* 4 (6): e5753.

Achilli, Alessandro, Anna Olivieri, Marco Pellecchia, Cristina Uboldi, Licia Colli, Nadia Al-Zahery, Matteo Accetturo, et al. 2008. “Mitochondrial Genomes of Extinct Aurochs Survive in Domestic Cattle.” *Current Biology: CB* 18 (4): R157–58.

Agranat-Tamir, Lily, Shamam Waldman, Mario A. S. Martin, David Gokhman, Nadav Mishol, Tzilla Eshel, Olivia Cheronet, et al. 2020. “The Genomic History of the Bronze Age Southern Levant.” *Cell* 181 (5): 1146–57.e11.

Alexander, David H., and Kenneth Lange. 2011. “Enhancements to the ADMIXTURE Algorithm for Individual Ancestry Estimation.” *BMC Bioinformatics* 12 (1): 1–6.

Alexander, David H., John Novembre, and Kenneth Lange. 2009. “Fast Model-Based Estimation of Ancestry in Unrelated Individuals.” *Genome Research* 19 (9): 1655–64.

Allain, Daniel, Bérengère Pena-Arnaud, Didier Foulquie, Yves Bourdillon, and Dominique François. 2014. “Introgression of Wool-Shedding Genes into the Romane Breed Sheep.” In *10. World Congress of Genetics Applied to Livestock Production*. American Society of Animal Science. https://hal.inrae.fr/hal-02739709/document.

Allentoft, Morten E., Matthew Collins, David Harker, James Haile, Charlotte L. Oskam, Marie L. Hale, Paula F. Campos, et al. 2012. “The Half-Life of DNA in Bone: Measuring Decay Kinetics in 158 Dated Fossils.” *Proceedings. Biological Sciences / The Royal Society* 279 (1748): 4724–33.

Allentoft, Morten E., Martin Sikora, Alba Refoyo-Martínez, Evan K. Irving-Pease, Anders Fischer, William Barrie, Andrés Ingason, et al. 2022. “Population Genomics of Stone Age Eurasia.” *bioRxiv*. https://doi.org/10.1101/2022.05.04.490594.

Alpaslan-Roodenberg, Songül, David Anthony, Hiba Babiker, Eszter Bánffy, Thomas Booth, Patricia Capone, Arati Deshpande-Mukherjee, et al. 2021. “Ethics of DNA Research on Human Remains: Five Globally Applicable Guidelines.” *Nature* 599 (7883): 41–46.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. “Basic Local Alignment Search Tool.” *Journal of Molecular Biology* 215 (3): 403–10.

Alves, Paulo C., José Melo-Ferreira, Hélder Freitas, and Pierre Boursot. 2008. “The Ubiquitous Mountain Hare Mitochondria: Multiple Introgressive Hybridization in Hares, Genus Lepus.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363 (1505): 2831–39.

Andel, Tjeerd H. van, and Curtis N. Runnels. 1995. “The Earliest Farmers in Europe.” *Antiquity* 69 (264): 481–500.

Anderson, S., M. H. de Bruijn, A. R. Coulson, I. C. Eperon, F. Sanger, and I. G. Young. 1982. “Complete Sequence of Bovine Mitochondrial DNA. Conserved Features of the Mammalian Mitochondrial Genome.” *Journal of Molecular Biology* 156 (4): 683–717.

Andrews, Simon. 2010. “FastQC.” Https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. 2010.

Anthony, D. W. 2010. “The Horse, the Wheel, and Language.” *The Horse, the Wheel, and Language*. https://www.degruyter.com/document/doi/10.1515/9781400831104/html?lang=de.

Arbuckle, Benjamin S., and Emily L. Hammer. 2019. “The Rise of Pastoralism in the Ancient Near East.” *Journal of Archaeological Research* 27 (3): 391–449.

Arbuckle, Benjamin S., Sarah Whitcher Kansa, Eric Kansa, David Orton, Canan Çakırlar, Lionel Gourichon, Levent Atici, et al. 2014. “Data Sharing Reveals Complexity in the

Westward Spread of Domestic Animals across Neolithic Turkey." *PloS One* 9 (6): e99845.
Arbuckle, Benjamin S., and Theo M. Kassebaum. 2021. "Management and Domestication of Cattle (Bos Taurus) in Neolithic Southwest Asia." *Animal Frontiers : The Review Magazine of Animal Agriculture* 11 (3): 10–19.
Avise, John C. 2000. *Phylogeography: The History and Formation of Species*. Harvard University Press.
Azaza, Mohamed, and Lídia Colominas. 2020. "The Roman Introduction and Exportation of Animals into Tunisia: Linking Archaeozoology with Textual and Iconographic Evidence." *Journal of Archaeological Science: Reports*. https://doi.org/10.1016/j.jasrep.2019.102076.
Bach, Bryan H., Ashling B. Quigley, Kaitlyn M. Gaynor, Alex McInturff, Kristin L. Charles, Janelle Dorcy, and Justin S. Brashares. 2022. "Identifying Individual Ungulates from Fecal DNA: A Comparison of Field Collection Methods to Maximize Efficiency, Ease, and Success." *Mammalian Biology = Zeitschrift Fur Saugetierkunde*, February, 1–12.
Bailey, J. F., M. B. Richards, V. A. Macaulay, I. B. Colson, I. T. James, D. G. Bradley, R. E. Hedges, and B. C. Sykes. 1996. "Ancient DNA Suggests a Recent Expansion of European Cattle from a Diverse Wild Progenitor Species." *Proceedings. Biological Sciences / The Royal Society* 263 (1376): 1467–73.
Baker, Polydora, and Fay Worley. 2019. *Animal Bones and Archaeology: Recovery to Archive*.
Ballard, J. William O., and Michael C. Whitlock. 2004. "The Incomplete Natural History of Mitochondria." *Molecular Ecology* 13 (4): 729–44.
Banks, William E., Marie-Hélène Moncel, Jean-Paul Raynal, Marlon E. Cobos, Daniel Romero-Alvarez, Marie-Noëlle Woillez, Jean-Philippe Faivre, et al. 2021. "An Ecological Niche Shift for Neanderthal Populations in Western Europe 70,000 Years Ago." *Scientific Reports* 11 (1): 5346.
Barnosky, Anthony D. 1986. "'Big Game' Extinction Caused by Late Pleistocene Climatic Change: Irish Elk (Megaloceros Giganteus) in Ireland." *Quaternary Research* 25 (1): 128–35.
Barnosky, Anthony D., Paul L. Koch, Robert S. Feranec, Scott L. Wing, and Alan B. Shabel. 2004. "Assessing the Causes of Late Pleistocene Extinctions on the Continents." *Science* 306 (5693): 70–75.
Batchelor, Christine L., Martin Margold, Mario Krapp, Della K. Murton, April S. Dalton, Philip L. Gibbard, Chris R. Stokes, Julian B. Murton, and Andrea Manica. 2019. "The Configuration of Northern Hemisphere Ice Sheets through the Quaternary." *Nature Communications* 10 (1): 3713.
Batini, Chiara, Pille Hallast, Daniel Zadik, Pierpaolo Maisano Delser, Andrea Benazzo, Silvia Ghirotto, Eduardo Arroyo-Pardo, et al. 2015. "Large-Scale Recent Expansion of European Patrilineages Shown by Population Resequencing." *Nature Communications* 6 (May): 7152.
Beckmann, J. S., and M. Soller. 1983. "Restriction Fragment Length Polymorphisms in Genetic Improvement: Methodologies, Mapping and Costs." *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 67 (1): 35–43.
Beja-Pereira, Albano, David Caramelli, Carles Lalueza-Fox, Cristiano Vernesi, Nuno Ferrand, Antonella Casoli, Felix Goyache, et al. 2006. "The Origin of European Cattle: Evidence from Modern and Ancient DNA." *Proceedings of the National Academy of Sciences of the United States of America* 103 (21): 8113–18.
Bergström, Anders, David W. G. Stanton, Ulrike H. Taron, Laurent Frantz, Mikkel-Holger S. Sinding, Erik Ersmark, Saskia Pfrengle, et al. 2022. "Grey Wolf Genomic History Reveals a Dual Ancestry of Dogs." *Nature* 607 (7918): 313–20.
Bertels, Frederic, Olin K. Silander, Mikhail Pachkov, Paul B. Rainey, and Erik van Nimwegen. 2014. "Automated Reconstruction of Whole-Genome Phylogenies from Short-Sequence Reads." *Molecular Biology and Evolution* 31 (5): 1077–88.

Boast, Alexander P., Laura S. Weyrich, Jamie R. Wood, Jessica L. Metcalf, Rob Knight, and Alan Cooper. 2018. “Coprolites Reveal Ecological Interactions Lost with the Extinction of New Zealand Birds.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (7): 1546–51.

Bocquet-Appel, Jean-Pierre. 2011. “When the World’s Population Took Off: The Springboard of the Neolithic Demographic Transition.” *Science*. https://doi.org/10.1126/science.1208880.

Boessenkool, Sanne, Kristian Hanghøj, Heidi M. Nistelberger, Clio Der Sarkissian, Agata T. Gondek, Ludovic Orlando, James H. Barrett, and Bastiaan Star. 2017. “Combining Bleach and Mild Predigestion Improves Ancient DNA Recovery from Bones.” *Molecular Ecology Resources* 17 (4): 742–51.

Bollongino, Ruth, Joachim Burger, Adam Powell, Marjan Mashkour, Jean-Denis Vigne, and Mark G. Thomas. 2012. “Modern Taurine Cattle Descended from Small Number of near-Eastern Founders.” *Molecular Biology and Evolution* 29 (9): 2101–4.

Bollongino, Ruth, Julia Elsner, Jean-Denis Vigne, and Joachim Burger. 2008. “Y-SNPs Do Not Indicate Hybridisation between European Aurochs and Domestic Cattle.” *PloS One* 3 (10): e3418.

Bonfiglio, Silvia, Alessandro Achilli, Anna Olivieri, Riccardo Negrini, Licia Colli, Luigi Liotta, Paolo Ajmone-Marsan, Antonio Torroni, and Luca Ferretti. 2010. “The Enigmatic Origin of Bovine mtDNA Haplogroup R: Sporadic Interbreeding or an Independent Event of Bos Primigenius Domestication in Italy?” *PloS One* 5 (12): e15760.

Bonfiglio, Silvia, Catarina Ginja, Anna De Gaetano, Alessandro Achilli, Anna Olivieri, Licia Colli, Kassahun Tesfaye, et al. 2012. “Origin and Spread of Bos Taurus: New Clues from Mitochondrial Genomes Belonging to Haplogroup T1.” *PloS One* 7 (6): e38601.

Borry, Maxime, Bryan Cordova, Angela Perri, Marsha Wibowo, Tanvi Prasad Honap, Jada Ko, Jie Yu, et al. 2020. “CoproID Predicts the Source of Coprolites and Paleofeces Using Microbiome Composition and Host DNA Content.” *PeerJ* 8 (April): e9001.

Bos, Kirsten I., Verena J. Schuenemann, G. Brian Golding, Hernán A. Burbano, Nicholas Waglechner, Brian K. Coombes, Joseph B. McPhee, et al. 2011. “A Draft Genome of Yersinia Pestis from Victims of the Black Death.” *Nature* 478 (7370): 506–10.

Bouckaert, Remco R., and Alexei J. Drummond. 2017. “bModelTest: Bayesian Phylogenetic Site Model Averaging and Model Comparison.” *BMC Evolutionary Biology* 17 (1): 42.

Bovine HapMap Consortium, Richard A. Gibbs, Jeremy F. Taylor, Curtis P. Van Tassell, William Barendse, Kellye A. Eversole, Clare A. Gill, et al. 2009. “Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds.” *Science* 324 (5926): 528–32.

Brace, Selina, Yoan Diekmann, Thomas J. Booth, Lucy van Dorp, Zuzana Faltyskova, Nadin Rohland, Swapan Mallick, et al. 2019. “Ancient Genomes Indicate Population Replacement in Early Neolithic Britain.” *Nature Ecology & Evolution* 3 (5): 765–71.

Bradley, Daniel G., Ronan T. Loftus, Patrick Cunningham, and David E. MacHugh. 1998. “Genetics and Domestic Cattle Origins.” *Evolutionary Anthropology: Issues, News, and Reviews*. https://doi.org/10.1002/(sici)1520-6505(1998)6:3<79::aid-evan2>3.0.co;2-r.

Bradley, D. G., D. E. MacHugh, P. Cunningham, and R. T. Loftus. 1996. “Mitochondrial Diversity and the Origins of African and European Cattle.” *Proceedings of the National Academy of Sciences of the United States of America* 93 (10): 5131–35.

Bray, Sarah C. E., Jeremy J. Austin, Jessica L. Metcalf, Kjartan Østbye, Eivind Østbye, Stein-Erik Lauritzen, Kim Aaris-Sørensen, Cristina Valdiosera, Christina J. Adler, and Alan Cooper. 2013. “Ancient DNA Identifies postglacial Recolonisation, Not Recent Bottlenecks, as the Primary Driver of Contemporary mtDNA Phylogeography and Diversity in Scandinavian Brown Bears.” *Diversity and Distributions*. https://doi.org/10.1111/j.1472-4642.2012.00923.x.

Breda, M., S. E. Collinge, S. A. Parfitt, and A. M. Lister. 2010. “Metric Analysis of Ungulate Mammals in the Early Middle Pleistocene of Britain, in Relation to Taxonomy and Biostratigraphy.” *Quaternary International*. https://doi.org/10.1016/j.quaint.2010.05.010.

Briggs, Adrian W., Jeffrey M. Good, Richard E. Green, Johannes Krause, Tomislav Maricic, Udo Stenzel, Carles Lalueza-Fox, et al. 2009. “Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes.” *Science* 325 (5938): 318–21.
Briggs, Adrian W., Udo Stenzel, Philip L. F. Johnson, Richard E. Green, Janet Kelso, Kay Prüfer, Matthias Meyer, et al. 2007. “Patterns of Damage in Genomic DNA Sequences from a Neandertal.” *Proceedings of the National Academy of Sciences of the United States of America* 104 (37): 14616–21.
Briggs, Adrian W., Udo Stenzel, Matthias Meyer, Johannes Krause, Martin Kircher, and Svante Pääbo. 2010. “Removal of Deaminated Cytosines and Detection of in Vivo Methylation in Ancient DNA.” *Nucleic Acids Research* 38 (6): e87.
Broushaki, Farnaz, Mark G. Thomas, Vivian Link, Saioa López, Lucy van Dorp, Karola Kirsanow, Zuzana Hofmanová, et al. 2016. “Early Neolithic Genomes from the Eastern Fertile Crescent.” *Science* 353 (6298): 499–503.
Browning, Brian L., and Sharon R. Browning. 2016. “Genotype Imputation with Millions of Reference Samples.” *American Journal of Human Genetics* 98 (1): 116–26.
Browning, Brian L., Xiaowen Tian, Ying Zhou, and Sharon R. Browning. 2021. “Fast Two-Stage Phasing of Large-Scale Sequence Data.” *American Journal of Human Genetics* 108 (10): 1880–90.
Browning, Brian L., Ying Zhou, and Sharon R. Browning. n.d. “A One Penny Imputed Genome from next Generation Reference Panels.” https://doi.org/10.1101/357806.
Brown, W. M., and J. Vinograd. 1974. “Restriction Endonuclease Cleavage Maps of Animal Mitochondrial DNAs.” *Proceedings of the National Academy of Sciences of the United States of America* 71 (11): 4617–21.
Brunson, Katherine, Ren Lele, Zhao Xin, Dong Xiaoling, Wang Hui, Zhou Jing, and Rowan Flad. 2020. “Zooarchaeology, Ancient mtDNA, and Radiocarbon Dating Provide New Evidence for the Emergence of Domestic Cattle and Caprines in the Tao River Valley of Gansu Province, Northwest China.” *Journal of Archaeological Science: Reports*. https://doi.org/10.1016/j.jasrep.2020.102262.
Brunson, Katherine, Xin Zhao, Nu He, Xiangming Dai, Antonia Rodrigues, and Dongya Yang. 2016. “New Insights into the Origins of Oracle Bone Divination: Ancient DNA from Late Neolithic Chinese Bovines.” *Journal of Archaeological Science* 74 (October): 35–44.
Burney, David A., and Timothy F. Flannery. 2005. “Fifty Millennia of Catastrophic Extinctions after Human Contact.” *Trends in Ecology & Evolution* 20 (7): 395–401.
Cahill, James A., Ian Stirling, Logan Kistler, Rauf Salamzade, Erik Ersmark, Tara L. Fulton, Mathias Stiller, Richard E. Green, and Beth Shapiro. 2015. “Genomic Evidence of Geographically Widespread Effect of Gene Flow from Polar Bears into Brown Bears.” *Molecular Ecology* 24 (6): 1205.
Cai, Dawei, Yang Sun, Zhuowei Tang, Songmei Hu, Wenying Li, Xingbo Zhao, Hai Xiang, and Hui Zhou. 2014. “The Origins of Chinese Domestic Cattle as Revealed by Ancient DNA Analysis.” *Journal of Archaeological Science* 41 (January): 423–34.
Cai, Dawei, Naifan Zhang, Siqi Zhu, Quanjia Chen, Lixin Wang, Xin Zhao, Xiaolin Ma, Thomas C. A. Royle, Hui Zhou, and Dongya Y. Yang. 2018. “Ancient DNA Reveals Evidence of Abundant Aurochs (Bos Primigenius) in Neolithic Northeast China.” *Journal of Archaeological Science*. https://doi.org/10.1016/j.jas.2018.08.003.
Cann, R. L., M. Stoneking, and A. C. Wilson. 1987. “Mitochondrial DNA and Human Evolution.” *Nature* 325 (6099): 31–36.
Carolino, N., and L. T. Gama. 2008. “Indicators of Genetic Erosion in an Endangered Population: The Alentejana Cattle Breed in Portugal1.” *Journal of Animal Science* 86 (1): 47–56.
Carvalho, Carlos Neto de, Silvério Figueiredo, Fernando Muniz, João Belo, Pedro P. Cunha, Andrea Baucon, Luis M. Cáceres, and Joaquín Rodriguez-Vidal. 2020. “Tracking the Last Elephants in Europe during the Würm Pleniglacial: The Importance of the Late Pleistocene Aeolianite Record in SW Iberia.” *Ichnos* 27 (3): 352–60.
Carvalho, Carlos Neto de, Fernando Muñiz, Luis M. Cáceres, Zain Belaústegui, Joaquín

Rodríguez-Vidal, João Belo, Noel Moreira, et al. 2022. “Aurochs Roamed along the SW Coast of Andalusia (Spain) during Late Pleistocene.” *Scientific Reports* 12 (1): 9911.
Cassidy, Lara M., Ros Ó. Maoldúin, Thomas Kador, Ann Lynch, Carleton Jones, Peter C. Woodman, Eileen Murphy, et al. 2020. “A Dynastic Elite in Monumental Neolithic Society.” *Nature* 582 (7812): 384–88.
Cassidy, Lara M., Rui Martiniano, Eileen M. Murphy, Matthew D. Teasdale, James Mallory, Barrie Hartwell, and Daniel G. Bradley. 2016. “Neolithic and Bronze Age Migration to Ireland and Establishment of the Insular Atlantic Genome.” *Proceedings of the National Academy of Sciences of the United States of America* 113 (2): 368–73.
Cassidy, Lara M., Matthew D. Teasdale, Seán Carolan, Ruth Enright, Raymond Werner, Daniel G. Bradley, Emma K. Finlay, and Valeria Mattiangeli. 2017. “Capturing Goats: Documenting Two Hundred Years of Mitochondrial DNA Diversity among Goat Populations from Britain and Ireland.” *Biology Letters* 13 (3). https://doi.org/10.1098/rsbl.2016.0876.
Ceballos, Francisco C., Peter K. Joshi, David W. Clark, Michèle Ramsay, and James F. Wilson. 2018. “Runs of Homozygosity: Windows into Population History and Trait Architecture.” *Nature Reviews. Genetics* 19 (4): 220–34.
Chang, Christopher C., Carson C. Chow, Laurent Cam Tellier, Shashaank Vattikuti, Shaun M. Purcell, and James J. Lee. 2015. “Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets.” *GigaScience* 4 (1). https://doi.org/10.1186/s13742-015-0047-8.
Chang, Ti-Cheng, Yang Yang, Ernest F. Retzel, and Wan-Sheng Liu. 2013. “Male-Specific Region of the Bovine Y Chromosome Is Gene Rich with a High Transcriptomic Activity in Testis Development.” *Proceedings of the National Academy of Sciences of the United States of America* 110 (30): 12373–78.
Chaolong Wang Xiaowei Zhan Liming Liang Gonçalo R. Abecasis Xihong Lin. 2015. “Improved Ancestry Estimation for Both Genotyping and Sequencing Data Using Projection Procrustes Analysis and Genotype Imputation.” *American Journal of Human Genetics* 96 (6): 926–37.
Chen, Ningbo, Yudong Cai, Qiuming Chen, Ran Li, Kun Wang, Yongzhen Huang, Songmei Hu, et al. 2018. “Whole-Genome Resequencing Reveals World-Wide Ancestry and Adaptive Introgression Events of Domesticated Cattle in East Asia.” *Nature Communications* 9 (1): 2337.
Chen, Shan-Yuan, Zi-Yuan Duan, Tao Sha, Jinggong Xiangyu, Shi-Fang Wu, and Ya-Ping Zhang. 2006. “Origin, Genetic Diversity, and Population Structure of Chinese Domestic Sheep.” *Gene* 376 (2): 216–23.
Chen, Shanyuan, Bang-Zhong Lin, Mumtaz Baig, Bikash Mitra, Ricardo J. Lopes, António M. Santos, David A. Magee, et al. 2010. “Zebu Cattle Are an Exclusive Legacy of the South Asia Neolithic.” *Molecular Biology and Evolution* 27 (1): 1–6.
Chessa, Bernardo, Filipe Pereira, Frederick Arnaud, Antonio Amorim, Félix Goyache, Ingrid Mainland, Rowland R. Kao, et al. 2009. “Revealing the History of Sheep Domestication Using Retrovirus Integrations.” *Science* 324 (5926): 532–36.
Chua, Physilia Y. S., Alex Crampton-Platt, Youri Lammers, Inger G. Alsos, Sanne Boessenkool, and Kristine Bohmann. 2021. “Metagenomics: A Viable Tool for Reconstructing Herbivore Diet.” *Molecular Ecology Resources* 21 (7): 2249–63.
Chunnian, Liang, Xiaoyun Wu, Xuezhi Ding, Hongbo Wang, Xian Guo, Min Chu, Pengjia Bao, and Ping Yan. 2016. “Characterization of the Complete Mitochondrial Genome Sequence of Wild Yak (Bos Mutus).” *Mitochondrial DNA. Part A, DNA Mapping, Sequencing, and Analysis* 27 (6): 4266–67.
Clayton, Jonathan B., Pajau Vangay, Hu Huang, Tonya Ward, Benjamin M. Hillmann, Gabriel A. Al-Ghalith, Dominic A. Travis, et al. 2016. “Captivity Humanizes the Primate Microbiome.” *Proceedings of the National Academy of Sciences of the United States of America* 113 (37): 10376–81.
Coissac, Eric, Peter M. Hollingsworth, Sébastien Lavergne, and Pierre Taberlet. 2016. “From

Barcodes to Genomes: Extending the Concept of DNA Barcoding." *Molecular Ecology* 25 (7): 1423–28.
Colli, Licia, Hovirag Lancioni, Irene Cardinali, Anna Olivieri, Marco Rosario Capodiferro, Marco Pellecchia, Marcin Rzepus, et al. 2015. "Whole Mitochondrial Genomes Unveil the Impact of Domestication on Goat Matrilineal Variability." *BMC Genomics* 16 (December): 1115.
Colominas, Lídia, Angela Schlumbaum, and Maria Saña. 2014. "The Impact of the Roman Empire on Animal Husbandry Practices: Study of the Changes in Cattle Morphology in the North-East of the Iberian Peninsula through Osteometric and Ancient DNA Analyses." *Archaeological and Anthropological Sciences*. https://doi.org/10.1007/s12520-013-0116-9.
Cooper, A. 1997. "Reply to Stoneking: Ancient DNA--How Do You Really Know When You Have It?" *American Journal of Human Genetics*.
Cortez, Diego, Ray Marin, Deborah Toledo-Flores, Laure Froidevaux, Angélica Liechti, Paul D. Waters, Frank Grützner, and Henrik Kaessmann. 2014. "Origins and Functional Evolution of Y Chromosomes across Mammals." *Nature* 508 (7497): 488–93.
Costa, Marcio C., Luis G. Arroyo, Emma Allen-Vercoe, Henry R. Stämpfli, Peter T. Kim, Amy Sturgeon, and J. Scott Weese. 2012. "Comparison of the Fecal Microbiota of Healthy Horses and Horses with Colitis by High Throughput Sequencing of the V3-V5 Region of the 16S rRNA Gene." *PloS One* 7 (7): e41484.
Cox, Karen, Niall McKeown, Gloria Antonini, Deborah Harvey, Emanuela Solano, An Van Breusegem, and Arno Thomaes. 2019. "Phylogeographic Structure and Ecological Niche Modelling Reveal Signals of Isolation and Postglacial Colonisation in the European Stag Beetle." *PloS One* 14 (4): e0215860.
Cubric-Curik, Vlatka, Dinko Novosel, Vladimir Brajkovic, Omar Rota Stabelli, Stefan Krebs, Johann Sölkner, Dragica Šalamon, et al. 2021. "Large-scale Mitogenome Sequencing Reveals Consecutive Expansions of Domestic Taurine Cattle and Supports Sporadic Aurochs Introgression." *Evolutionary Applications*, November. https://doi.org/10.1111/eva.13315.
Cummings, Vicki, and James Morris. 2022. "Neolithic Explanations Revisited: Modelling the Arrival and Spread of Domesticated Cattle into Neolithic Britain." *Environmental Archaeology*. https://doi.org/10.1080/14614103.2018.1536498.
Currat, Mathias, Manuel Ruedi, Rémy J. Petit, and Laurent Excoffier. 2008. "The Hidden Side of Invasions: Massive Introgression by Local Genes." *Evolution; International Journal of Organic Evolution* 62 (8): 1908–20.
Cymbron, Teresa, Abigail R. Freeman, M. Isabel Malheiro, Jean-Denis Vigne, and Daniel G. Bradley. 2005. "Microsatellite Diversity Suggests Different Histories for Mediterranean and Northern European Cattle Populations." *Proceedings. Biological Sciences / The Royal Society* 272 (1574): 1837–43.
Dabney, Jesse, and Matthias Meyer. 2019. "Extraction of Highly Degraded DNA from Ancient Bones and Teeth." *Methods in Molecular Biology* 1963: 25–29.
Dabney, Jesse, Matthias Meyer, and Svante Pääbo. 2013. "Ancient DNA Damage." *Cold Spring Harbor Perspectives in Biology* 5 (7). https://doi.org/10.1101/cshperspect.a012567.
Dai, Chuanyin, Na Zhao, Wenjuan Wang, Congtian Lin, Bin Gao, Xiaojun Yang, Zhengwang Zhang, and Fumin Lei. 2011. "Profound Climatic Effects on Two East Asian Black-Throated Tits (Ave: Aegithalidae), Revealed by Ecological Niche Models and Phylogeographic Analysis." *PloS One* 6 (12): e29329.
Daly, Kevin G., Pierpaolo Maisano Delser, Victoria E. Mullin, Amelie Scheu, Valeria Mattiangeli, Matthew D. Teasdale, Andrew J. Hare, et al. 2018. "Ancient Goat Genomes Reveal Mosaic Domestication in the Fertile Crescent." *Science* 361 (6397): 85–88.
Daly, Kevin G., Valeria Mattiangeli, Andrew J. Hare, Hossein Davoudi, Homa Fathi, Sanaz Beizaee Doost, Sarieh Amiri, et al. 2021. "Herded and Hunted Goat Genomes from the Dawn of Domestication in the Zagros Mountains." *Proceedings of the National Academy*

of Sciences of the United States of America 118 (25). https://doi.org/10.1073/pnas.2100901118.
Damgaard, Peter B., Ashot Margaryan, Hannes Schroeder, Ludovic Orlando, Eske Willerslev, and Morten E. Allentoft. 2015. “Improving Access to Endogenous DNA in Ancient Bones and Teeth.” *Scientific Reports* 5 (1): 1–12.
Damgaard, Peter de Barros, Nina Marchi, Simon Rasmussen, Michaël Peyrot, Gabriel Renaud, Thorfinn Korneliussen, J. Víctor Moreno-Mayar, et al. 2018. “137 Ancient Human Genomes from across the Eurasian Steppes.” *Nature* 557 (7705): 369–74.
Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, et al. 2011. “The Variant Call Format and VCFtools.” *Bioinformatics* 27 (15): 2156–58.
Danna, K., and D. Nathans. 1971. “Specific Cleavage of Simian Virus 40 DNA by Restriction Endonuclease of Hemophilus Influenzae.” *Proceedings of the National Academy of Sciences of the United States of America* 68 (12): 2913–17.
Darriba, Diego, Guillermo L. Taboada, Ramón Doallo, and David Posada. 2012. “jModelTest 2: More Models, New Heuristics and Parallel Computing.” *Nature Methods* 9 (8): 772.
Darwin, C. 1883. *On the Origin of Species: By Means of Natural Selection, Or, the Preservation of Favored Races in the Struggle for Life*. Appleton.
Davenport, Emily R., Jon G. Sanders, Se Jin Song, Katherine R. Amato, Andrew G. Clark, and Rob Knight. 2017. “The Human Microbiome in Evolution.” *BMC Biology* 15 (1): 127.
David, Rosalie. 2008. *Egyptian Mummies and Modern Science*. Cambridge University Press.
Davison, John, Simon Y. W. Ho, Sarah C. Bray, Marju Korsten, Egle Tammeleht, Maris Hindrikson, Kjartan Østbye, et al. 2011. “Late-Quaternary Biogeographic Scenarios for the Brown Bear (Ursus Arctos), a Wild Mammal Model Species.” *Quaternary Science Reviews*. https://doi.org/10.1016/j.quascirev.2010.11.023.
Dean, C. J., I. B. Slizovskiy, K. K. Crone, A. X. Pfennig, B. J. Heins, L. S. Caixeta, and N. R. Noyes. 2021. “Investigating the Cow Skin and Teat Canal Microbiomes of the Bovine Udder Using Different Sampling and Sequencing Approaches.” *Journal of Dairy Science* 104 (1): 644–61.
Deb, G. K., R. Khatun, S. M. J. Hossain, S. S. Rahman, M. A. B. Bhuiyan, S. Mobassirin, S. Afrin, et al. 2020. “Complete Mitochondrial Genome Sequence of Bos Frontalis (Gayal) from Bangladesh.” *bioRxiv*. https://doi.org/10.1101/2020.12.31.424938.
Decker, Jared E., Stephanie D. McKay, Megan M. Rolf, Jaewoo Kim, Antonio Molina Alcalá, Tad S. Sonstegard, Olivier Hanotte, et al. 2014. “Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle.” *PLoS Genetics* 10 (3): e1004254.
Demars, Julie, Margarita Cano, Laurence Drouilhet, Florence Plisson-Petit, Philippe Bardou, Stéphane Fabre, Bertrand Servin, et al. 2017. “Genome-Wide Identification of the Mutation Underlying Fleece Variation and Discriminating Ancestral Hairy Species from Modern Woolly Sheep.” *Molecular Biology and Evolution* 34 (7): 1722–29.
Demirbaş, Yasin, İrfan Albayrak, Ayça Özkan Koca, Milomir Stefanović, Felix Knauer, and Franz Suchentrunk. 2019. “Spatial Genetics of Brown Hares (Lepus Europaeus Pallas, 1778) from Turkey: Different Gene Pool Architecture on Either Side of the Bosphorus?” *Mammalian Biology*. https://doi.org/10.1016/j.mambio.2018.09.005.
Demirci, Sevgin, Evren Koban Baştanlar, Nihan Dilşad Dağtaş, Evangelia Pişkin, Atilla Engin, Füsun Özer, Eren Yüncü, Şükrü Anıl Doğan, and İnci Togan. 2013. “Mitochondrial DNA Diversity of Modern, Ancient and Wild Sheep (Ovis Gmelinii Anatolica) from Turkey: New Insights on the Evolutionary History of Sheep.” *PLoS ONE*. https://doi.org/10.1371/journal.pone.0081952.
Diamond, Jared. 2002. “Evolution, Consequences and Future of Plant and Animal Domestication.” *Nature* 418 (6898): 700–707.
Dong, Kunzhe, Min Yang, Jiangang Han, Qing Ma, Jilong Han, Ziyi Song, Cuicheng Luosang, et al. 2020. “Genomic Analysis of Worldwide Sheep Breeds Reveals PDGFD as a Major Target of Fat-Tail Selection in Sheep.” *BMC Genomics* 21 (1): 800.
Douglas, Kory C., Natalie D. Halbert, Claire Kolenda, Christopher Childers, David L. Hunter,

and James N. Derr. 2011. “Complete Mitochondrial DNA Sequence Analysis of Bison Bison and Bison-Cattle Hybrids: Function and Phylogeny.” *Mitochondrion* 11 (1): 166–75.
Drummond, Alexei J., and Andrew Rambaut. 2007. “BEAST: Bayesian Evolutionary Analysis by Sampling Trees.” *BMC Evolutionary Biology*. https://doi.org/10.1186/1471-2148-7-214.
Durand, Eric Y., Nick Patterson, David Reich, and Montgomery Slatkin. 2011. “Testing for Ancient Admixture between Closely Related Populations.” *Molecular Biology and Evolution* 28 (8): 2239–52.
Eastmond, David A., James T. Macgregor, and Ronald S. Slesinski. 2008. “Trivalent Chromium: Assessing the Genotoxic Risk of an Essential Trace Element and Widely Used Human and Animal Nutritional Supplement.” *Critical Reviews in Toxicology* 38 (3): 173–90.
Edgar, Robert C. 2004. “MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput.” *Nucleic Acids Research* 32 (5): 1792–97.
Edwards, Ceiridwen J., Ruth Bollongino, Amelie Scheu, Andrew Chamberlain, Anne Tresset, Jean-Denis Vigne, Jillian F. Baird, et al. 2007. “Mitochondrial DNA Analysis Shows a Near Eastern Neolithic Origin for Domestic Cattle and No Indication of Domestication of European Aurochs.” *Proceedings. Biological Sciences / The Royal Society* 274 (1616): 1377–85.
Edwards, Ceiridwen J., Catarina Ginja, Juha Kantanen, Lucía Pérez-Pardal, Anne Tresset, Frauke Stock, European Cattle Genetic Diversity Consortium, et al. 2011. “Dual Origins of Dairy Cattle Farming--Evidence from a Comprehensive Survey of European Y-Chromosomal Variation.” *PloS One* 6 (1): e15922.
Edwards, Ceiridwen J., Catarina Ginja, Juha Kantanen, Lucía Pérez-Pardal, Anne Tresset, Frauke Stock, Luis T. Gama, et al. 2011. “Dual Origins of Dairy Cattle Farming – Evidence from a Comprehensive Survey of European Y-Chromosomal Variation.” *PLoS ONE*. https://doi.org/10.1371/journal.pone.0015922.
Edwards, C. J., J. Connellan, P. F. Wallace, S. D. E. Park, F. M. McCormick, I. Olsaker, E. Eythórsdóttir, D. E. MacHugh, J. F. Bailey, and D. G. Bradley. 2003. “Feasibility and Utility of Microsatellite Markers in Archaeological Cattle Remains from a Viking Age Settlement in Dublin.” *Animal Genetics* 34 (6): 410–16.
Ehlers, Jürgen, and Philip Gibbard. 2008. “Extent and Chronology of Quaternary Glaciation.” *Episodes* 31 (2): 211–18.
Emmons, Alexandra L., Amy Z. Mundorff, Sarah W. Keenan, Jonathan Davoren, Janna Andronowski, David O. Carter, and Jennifer M. DeBruyn. 2020. “Characterizing the postmortem Human Bone Microbiome from Surface-Decomposed Remains.” *PloS One* 15 (7): e0218636.
Evin, Allowen, Keith Dobney, Renate Schafberg, Joseph Owen, Una Strand Vidarsdottir, Greger Larson, and Thomas Cucchi. 2015. “Phenotype and Animal Domestication: A Study of Dental Variation between Domestic, Wild, Captive, Hybrid and Insular Sus Scrofa.” *BMC Evolutionary Biology* 15 (1): 6.
Fages, Antoine, Andaine Seguin-Orlando, Mietje Germonpré, and Ludovic Orlando. 2020. “Horse Males Became over-Represented in Archaeological Assemblages during the Bronze Age.” *Journal of Archaeological Science, Reports* 31 (102364): 102364.
Fang, Zhijia, Min Zhao, Hong Zhen, Lifeng Chen, Ping Shi, and Zhiwei Huang. 2014. “Genotoxicity of Tri- and Hexavalent Chromium Compounds In Vivo and Their Modes of Action on DNA Damage In Vitro.” *PloS One* 9 (8). https://doi.org/10.1371/journal.pone.0103194.
Fellows Yates, James A., Aida Andrades Valtueña, Åshild J. Vågene, Becky Cribdon, Irina M. Velsko, Maxime Borry, Miriam J. Bravo-Lopez, et al. 2021. “Community-Curated and Standardised Metadata of Published Ancient Metagenomic Samples with AncientMetagenomeDir.” *Scientific Data* 8 (1): 31.
Foote, R. H. 2002. “The History of Artificial Insemination: Selected Notes and notables1.”

Journal of Animal Science. https://doi.org/10.2527/animalsci2002.80e-suppl_21a.
Forster, Peter. 2004. “Ice Ages and the Mitochondrial DNA Chronology of Human Dispersals: A Review.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 359 (1442): 255–64; discussion 264.
Frantz, Laurent A. F., Daniel G. Bradley, Greger Larson, and Ludovic Orlando. 2020. “Animal Domestication in the Era of Ancient Genomics.” *Nature Reviews. Genetics* 21 (8): 449–60.
Frantz, Laurent A. F., James Haile, Audrey T. Lin, Amelie Scheu, Christina Geörg, Norbert Benecke, Michelle Alexander, et al. 2019. “Ancient Pigs Reveal a near-Complete Genomic Turnover Following Their Introduction to Europe.” *Proceedings of the National Academy of Sciences of the United States of America* 116 (35): 17231–38.
Fredsted, Tina, Trine Wincentz, and Palle Villesen. 2006. “Introgression of Mountain Hare (Lepus Timidus) Mitochondrial DNA into Wild Brown Hares (Lepus Europaeus) in Denmark.” *BMC Ecology* 6 (November): 17.
Freeman, A. R., D. G. Bradley, S. Nagda, J. P. Gibson, and O. Hanotte. 2006. “Combination of Multiple Microsatellite Data Sets to Investigate Genetic Diversity and Admixture of Domestic Cattle.” *Animal Genetics* 37 (1): 1–9.
Freeman, A. R., C. M. Meghen, D. E. MacHugh, R. T. Loftus, M. D. Achukwi, A. Bado, B. Sauveroche, and D. G. Bradley. 2004. “Admixture and Diversity in West African Cattle Populations.” *Molecular Ecology* 13 (11): 3477–87.
Fuks, Daniel, and Zachary C. Dunseth. 2020. “Dung in the Dumps: What We Can Learn from Multi-Proxy Studies of Archaeological Dung Pellets.” *Vegetation History and Archaeobotany* 30 (1): 137–53.
Fuller, Dorian Q., George Willcox, and Robin G. Allaby. 2011. “Cultivation and Domestication Had Multiple Origins: Arguments against the Core Area Hypothesis for the Origins of Agriculture in the Near East.” *World Archaeology* 43 (4): 628–52.
Funk, Daniel J., and Kevin E. Omland. 2003. “Species-Level Paraphyly and Polyphyly: Frequency, Causes, and Consequences, with Insights from Animal Mitochondrial DNA.” *Annual Review of Ecology, Evolution, and Systematics*. https://doi.org/10.1146/annurev.ecolsys.34.011802.132421.
Fu, Qiaomei, Alissa Mittnik, Philip L. F. Johnson, Kirsten Bos, Martina Lari, Ruth Bollongino, Chengkai Sun, et al. 2013. “A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes.” *Current Biology: CB* 23 (7): 553–59.
Fu, Qiaomei, Cosimo Posth, Mateja Hajdinjak, Martin Petr, Swapan Mallick, Daniel Fernandes, Anja Furtwängler, et al. 2016. “The Genetic History of Ice Age Europe.” *Nature* 534 (7606): 200–205.
Gamba, Cristina, Eppie R. Jones, Matthew D. Teasdale, Russell L. McLaughlin, Gloria Gonzalez-Fortes, Valeria Mattiangeli, László Domboróczki, et al. 2014. “Genome Flux and Stasis in a Five Millennium Transect of European Prehistory.” *Nature Communications*. https://doi.org/10.1038/ncomms6257.
Gascuel, O. 2006. “Neighbor-Joining Revealed.” *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msl072.
Gautason, Egill, Anna A. Schönherz, Goutam Sahana, and Bernt Guldbrandtsen. 2020. “Relationship of Icelandic Cattle with Northern and Western European Cattle Breeds, Admixture and Population Structure.” *Acta Agriculturae Scandinavica, Section A — Animal Science*. https://doi.org/10.1080/09064702.2019.1699951.
Gautier, Mathieu, Denis Laloë, and Katayoun Moazami-Goudarzi. 2010. “Insights into the Genetic History of French Cattle from Dense SNP Data on 47 Worldwide Breeds.” *PloS One* 5 (9). https://doi.org/10.1371/journal.pone.0013038.
Gelabert, Pere, Susanna Sawyer, Anders Bergström, Ashot Margaryan, Thomas C. Collin, Tengiz Meshveliani, Anna Belfer-Cohen, et al. 2021. “Genome-Scale Sequencing and Analysis of Human, Wolf, and Bison DNA from 25,000-Year-Old Sediment.” *Current Biology: CB* 31 (16): 3564–74.e9.
Gignoux, Christopher R., Brenna M. Henn, and Joanna L. Mountain. 2011. “Rapid, Global

Demographic Expansions after the Origins of Agriculture." *Proceedings of the National Academy of Sciences of the United States of America* 108 (15): 6044–49.
Gilbert, M. Thomas P., Dennis L. Jenkins, Anders Götherstrom, Nuria Naveran, Juan J. Sanchez, Michael Hofreiter, Philip Francis Thomsen, et al. 2008. "DNA from Pre-Clovis Human Coprolites in Oregon, North America." *Science* 320 (5877): 786–89.
Gilbert, M. Thomas P., Lynn P. Tomsho, Snjezana Rendulic, Michael Packard, Daniela I. Drautz, Andrei Sher, Alexei Tikhonov, et al. 2007. "Whole-Genome Shotgun Sequencing of Mitochondria from Ancient Hair Shafts." *Science* 317 (5846): 1927–30.
Gilbert, M. Thomas P., Andrew S. Wilson, Michael Bunce, Anders J. Hansen, Eske Willerslev, Beth Shapiro, Thomas F. G. Higham, et al. 2004. "Ancient Mitochondrial DNA from Hair." *Current Biology: CB* 14 (12): R463–64.
Ginja, Catarina, Luís Telo da Gama, and Maria Cecilia T. Penedo. 2009. "Y Chromosome Haplotype Analysis in Portuguese Cattle Breeds Using SNPs and STRs." *The Journal of Heredity* 100 (2): 148–57.
Götherström, Anders, Cecilia Anderung, Linda Hellborg, Rengert Elburg, Colin Smith, Dan G. Bradley, and Hans Ellegren. 2005. "Cattle Domestication in the Near East Was Followed by Hybridization with Aurochs Bulls in Europe." *Proceedings. Biological Sciences / The Royal Society* 272 (1579): 2345–50.
Gouy, Manolo, Stéphane Guindon, and Olivier Gascuel. 2010. "SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building." *Molecular Biology and Evolution* 27 (2): 221–24.
Green, Eleanor Joan, and Camilla F. Speller. 2017. "Novel Substrates as Sources of Ancient DNA: Prospects and Hurdles." *Genes* 8 (7). https://doi.org/10.3390/genes8070180.
Green, Richard E., Johannes Krause, Adrian W. Briggs, Tomislav Maricic, Udo Stenzel, Martin Kircher, Nick Patterson, et al. 2010. "A Draft Sequence of the Neandertal Genome." *Science* 328 (5979): 710–22.
Greenwood, Paul J. 1980. "Mating Systems, Philopatry and Dispersal in Birds and Mammals." *Animal Behaviour*. https://doi.org/10.1016/s0003-3472(80)80103-5.
Greminger, Maja P., Michael Krützen, Claude Schelling, Aldona Pienkowska-Schelling, and Peter Wandeler. 2010. "The Quest for Y-Chromosomal Markers - Methodological Strategies for Mammalian Non-Model Organisms." *Molecular Ecology Resources* 10 (3): 409–20.
Grigson, Caroline. 1980. "The Craniology and Relationships of Four Species of Bos 5. Bos Iudicus L." *Journal of Archaeological Science* 7 (1): 3–32.
Groves, C. P. 2009. "Systematic Relationships in the Bovini (Artiodactyla, Bovidae)." *Journal of Zoological Systematics and Evolutionary Research*. https://doi.org/10.1111/j.1439-0469.1981.tb00243.x.
Guindon, Stéphane, Jean-François Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk, and Olivier Gascuel. 2010. "New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0." *Systematic Biology* 59 (3): 307–21.
Guindon, Stéphane, and Olivier Gascuel. 2003. "A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood." *Systematic Biology* 52 (5): 696–704.
Günther, Torsten, and Carl Nettelblad. 2019. "The Presence and Impact of Reference Bias on Population Genomic Studies of Prehistoric Human Populations." *PLoS Genetics* 15 (7): e1008302.
Haak, Wolfgang, Iosif Lazaridis, Nick Patterson, Nadin Rohland, Swapan Mallick, Bastien Llamas, Guido Brandt, et al. 2015. "Massive Migration from the Steppe Was a Source for Indo-European Languages in Europe." *Nature* 522 (7555): 207–11.
Hagelberg, E., and J. B. Clegg. 1991. "Isolation and Characterization of DNA from Archaeological Bone." *Proceedings. Biological Sciences / The Royal Society* 244 (1309): 45–50.
Hahn, Christoph, Lutz Bachmann, and Bastien Chevreux. 2013. "Reconstructing

Mitochondrial Genomes Directly from Genomic next-Generation Sequencing Reads—a Baiting and Iterative Mapping Approach." *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkt371.
Haile, James, Richard Holdaway, Karen Oliver, Michael Bunce, M. Thomas P. Gilbert, Rasmus Nielsen, Kasper Munch, Simon Y. W. Ho, Beth Shapiro, and Eske Willerslev. 2007. "Ancient DNA Chronology within Sediment Deposits: Are Paleobiological Reconstructions Possible and Is DNA Leaching a Factor?" *Molecular Biology and Evolution* 24 (4): 982–89.
Hänni, C., V. Laudet, M. Sakka, A. Bègue, and D. Stéhelin. 1990. "Amplification of mitochondrial DNA fragments from ancient human teeth and bones." *Comptes rendus de l'Academie des sciences. Serie III, Sciences de la vie* 310 (9): 365–70.
Hanotte, O., C. L. Tawah, D. G. Bradley, M. Okomo, Y. Verjee, J. Ochieng, and J. E. O. Rege. 2000. "Geographic Distribution and Frequency of a Taurine Bos Taurus and an Indicine Bos Indicus Y Specific Allele amongst Sub-Saharan African Cattle Breeds." *Molecular Ecology*. https://doi.org/10.1046/j.1365-294x.2000.00858.x.
Hansen, Henrik B., Peter B. Damgaard, Ashot Margaryan, Jesper Stenderup, Niels Lynnerup, Eske Willerslev, and Morten E. Allentoft. 2017. "Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum." *PloS One* 12 (1): e0170940.
Harbert, Robert S. 2018. "Algorithms and Strategies in Short-Read Shotgun Metagenomic Reconstruction of Plant Communities." *Applications in Plant Sciences* 6 (3): e1034.
Hassanin, Alexandre, Céline Bonillo, Bui Xuan Nguyen, and Corinne Cruaud. 2010. "Comparisons between Mitochondrial Genomes of Domestic Goat (Capra Hircus) Reveal the Presence of Numts and Multiple Sequencing Errors." *Mitochondrial DNA* 21 (3-4): 68–76.
Hassanin, Alexandre, Frédéric Delsuc, Anne Ropiquet, Catrin Hammer, Bettine Jansen van Vuuren, Conrad Matthee, Manuel Ruiz-Garcia, et al. 2012. "Pattern and Timing of Diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as Revealed by a Comprehensive Analysis of Mitochondrial Genomes." *Comptes Rendus Biologies* 335 (1): 32–50.
Haworth, Sarah E., Kevin S. White, Steeve D. Côté, and Aaron B. A. Shafer. 2019. "Space, Time and Captivity: Quantifying the Factors Influencing the Fecal Microbiome of an Alpine Ungulate." *FEMS Microbiology Ecology* 95 (7). https://doi.org/10.1093/femsec/fiz095.
Hayes, Ben J., and Hans D. Daetwyler. 2019. "1000 Bull Genomes Project to Map Simple and Complex Genetic Traits in Cattle: Applications and Outcomes." *Annual Review of Animal Biosciences* 7 (February): 89–102.
Hebsgaard, Martin B., M. Thomas P. Gilbert, Jette Arneborg, Patricia Heyn, Morten E. Allentoft, Michael Bunce, Kasper Munch, Charles Schweger, and Eske Willerslev. 2009. "'The Farm Beneath the Sand' – an Archaeological Case Study on Ancient 'dirt' DNA." *Antiquity*. https://doi.org/10.1017/s0003598x00098537.
Hellenthal, Garrett, George B. J. Busby, Gavin Band, James F. Wilson, Cristian Capelli, Daniel Falush, and Simon Myers. 2014. "A Genetic Atlas of Human Admixture History." *Science* 343 (6172): 747–51.
Hermes, Taylor R., Michael D. Frachetti, Paula N. Doumani Dupuy, Alexei Mar'yashev, Almut Nebel, and Cheryl A. Makarewicz. 2019. "Early Integration of Pastoralism and Millet Cultivation in Bronze Age Eurasia." *Proceedings. Biological Sciences / The Royal Society* 286 (1910): 20191273.
Hewitt, G. 2000. "The Genetic Legacy of the Quaternary Ice Ages." *Nature* 405 (6789): 907–13.
Hewitt, Godfrey M. 1999. "postglacial Re-Colonization of European Biota." *Biological Journal of the Linnean Society*. https://doi.org/10.1111/j.1095-8312.1999.tb01160.x.
Hiendleder, S., H. Lewalski, and A. Janke. 2008. "Complete Mitochondrial Genomes of Bos Taurus and Bos Indicus Provide New Insights into Intra-Species Variation, Taxonomy and Domestication." *Cytogenetic and Genome Research* 120 (1-2): 150–56.

Higgins, Denice, John Kaidonis, Grant Townsend, Toby Hughes, and Jeremy J. Austin. 2013. "Targeted Sampling of Cementum for Recovery of Nuclear DNA from Human Teeth and the Impact of Common Decontamination Measures." *Investigative Genetics* 4 (1): 1–8.
Higham, Charles. 1996. *The Bronze Age of Southeast Asia*. Cambridge University Press.
Higuchi, R., B. Bowman, M. Freiberger, O. A. Ryder, and A. C. Wilson. 1984. "DNA Sequences from the Quagga, an Extinct Member of the Horse Family." *Nature* 312 (5991): 282–84.
Hofreiter, Michael. 2008. "Long DNA Sequences and Large Data Sets: Investigating the Quaternary via Ancient DNA." *Quaternary Science Reviews* 27 (27): 2586–92.
Hofreiter, Michael, Viviane Jaenicke, David Serre, Arndt von Haeseler, and Svante Pääbo. 2001. "DNA Sequences from Multiple Amplifications Reveal Artifacts Induced by Cytosine Deamination in Ancient DNA." *Nucleic Acids Research* 29 (23): 4793–99.
Horsburgh, K. Ann, Stefan Prost, Anna Gosling, Jo-Ann Stanton, Christy Rand, and Elizabeth A. Matisoo-Smith. 2013. "The Genetic Diversity of the Nguni Breed of African Cattle (Bos Spp.): Complete Mitochondrial Genomes of Haplogroup T1." *PloS One* 8 (8): e71956.
Hübler, Ron, Felix M. Key, Christina Warinner, Kirsten I. Bos, Johannes Krause, and Alexander Herbig. 2019. "HOPS: Automated Detection and Authentication of Pathogen DNA in Archaeological Remains." *Genome Biology* 20 (1): 280.
Hui, Ruoyun, Eugenia D'Atanasio, Lara M. Cassidy, Christiana L. Scheib, and Toomas Kivisild. 2020. "Evaluating Genotype Imputation Pipeline for Ultra-Low Coverage Ancient Genomes." *Scientific Reports* 10 (1): 18542.
Islam, Mohammed Shehadul, Aditya Aryasomayajula, and Ponnambalam Selvaganapathy. 2017. "A Review on Macroscale and Microscale Cell Lysis Methods." *Micromachines*. https://doi.org/10.3390/mi8030083.
Jobling, Mark A., and Chris Tyler-Smith. 2003. "The Human Y Chromosome: An Evolutionary Marker Comes of Age." *Nature Reviews. Genetics* 4 (8): 598–612.
Jones, Eppie R., Gloria Gonzalez-Fortes, Sarah Connell, Veronika Siska, Anders Eriksson, Rui Martiniano, Russell L. McLaughlin, et al. 2015. "Upper Palaeolithic Genomes Reveal Deep Roots of Modern Eurasians." *Nature Communications* 6 (November): 8912.
Jónsson, Hákon, Aurélien Ginolhac, Mikkel Schubert, Philip L. F. Johnson, and Ludovic Orlando. 2013. "mapDamage2.0: Fast Approximate Bayesian Estimates of Ancient DNA Damage Parameters." *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btt193.
Kainer, David, and Robert Lanfear. 2015. "The Effects of Partitioning on Phylogenetic Inference." *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msv026.
Kalis, Arie J., Josef Merkt, and Jürgen Wunderlich. 2003. "Environmental Changes during the Holocene Climatic Optimum in Central Europe - Human Impact and Natural Causes." *Quaternary Science Reviews* 22 (1): 33–79.
Kantanen, J., C. J. Edwards, D. G. Bradley, H. Viinalass, S. Thessler, Z. Ivanova, T. Kiselyova, et al. 2009. "Maternal and Paternal Genealogy of Eurasian Taurine Cattle (Bos Taurus)." *Heredity* 103 (5): 404–15.
Karimi, Karim, Eva M. Strucken, Nasir Moghaddar, Mohammad H. Ferdosi, Ali Esmailizadeh, and Cedric Gondro. 2016. "Local and Global Patterns of Admixture and Population Structure in Iranian Native Cattle." *BMC Genetics* 17 (1): 108.
Karmin, Monika, Lauri Saag, Mário Vicente, Melissa A. Wilson Sayres, Mari Järve, Ulvi Gerst Talas, Siiri Rootsi, et al. 2015. "A Recent Bottleneck of Y Chromosome Diversity Coincides with a Global Change in Culture." *Genome Research* 25 (4): 459–66.
Kay, Gemma L., Martin J. Sergeant, Valentina Giuffra, Pasquale Bandiera, Marco Milanese, Barbara Bramanti, Raffaella Bianucci, and Mark J. Pallen. 2014. "Recovery of a Medieval Brucella Melitensis Genome Using Shotgun Metagenomics." *mBio* 5 (4): e01337–14.
Keighley, Xénia, Maiken Hemme Bro-Jørgensen, Hans Ahlgren, Paul Szpak, Marta Maria Ciucani, Fátima Sánchez Barreiro, Lesley Howse, et al. 2021. "Predicting Sample Success for Large-Scale Ancient DNA Studies on Marine Mammals." *Molecular Ecology Resources* 21 (4): 1149–66.

Kemp Brian M., David Glenn Smith. 2005. “Use of Bleach to Eliminate Contaminating DNA from the Surface of Bones and Teeth.” *Forensic Science International* 154 (1): 53–61.
Khazaeli, R. 2013. “The Subsistence Economy of the Old Nishapur from the Formation of the City up to the Mongol Period.” MA, University of Tehran.
Kijas, James W., Johannes A. Lenstra, Ben Hayes, Simon Boitard, Laercio R. Porto Neto, Magali San Cristobal, Bertrand Servin, et al. 2012. “Genome-Wide Analysis of the World’s Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection.” *PLoS Biology* 10 (2): e1001258.
Kim, Daehwan, Li Song, Florian P. Breitwieser, and Steven L. Salzberg. 2016. “Centrifuge: Rapid and Sensitive Classification of Metagenomic Sequences.” *Genome Research* 26 (12): 1721–29.
Kistler, Logan, Roselyn Ware, Oliver Smith, Matthew Collins, and Robin G. Allaby. 2017. “A New Model for Ancient DNA Decay Based on Paleogenomic Meta-Analysis.” *Nucleic Acids Research* 45 (11): 6310–20.
Knights, Dan, Justin Kuczynski, Emily S. Charlson, Jesse Zaneveld, Michael C. Mozer, Ronald G. Collman, Frederic D. Bushman, Rob Knight, and Scott T. Kelley. 2011. “Bayesian Community-Wide Culture-Independent Microbial Source Tracking.” *Nature Methods* 8 (9): 761–63.
Koch, Paul L., and Anthony D. Barnosky. 2006. “Late Quaternary Extinctions: State of the Debate,” November. https://doi.org/10.1146/annurev.ecolsys.34.011802.132415.
Korneliussen, Thorfinn Sand, Anders Albrechtsen, and Rasmus Nielsen. 2014. “ANGSD: Analysis of Next Generation Sequencing Data.” *BMC Bioinformatics* 15 (November): 356.
Kosintsev, P. A., and O. P. Bachura. 2013. “Late Pleistocene and Holocene Mammal Fauna of the Southern Urals.” *Quaternary International*. https://doi.org/10.1016/j.quaint.2012.06.022.
Krings, Matthias, Anne Stone, Ralf W. Schmitz, Heike Krainitzki, Mark Stoneking, and Svante Pääbo. 1997. “Neandertal DNA Sequences and the Origin of Modern Humans.” *Cell*. https://doi.org/10.1016/s0092-8674(00)80310-4.
Kumar, P., A. R. Freeman, R. T. Loftus, C. Gaillard, D. Q. Fuller, and D. G. Bradley. 2003. “Admixture Analysis of South Asian Cattle.” *Heredity* 91 (1): 43–50.
Kyselý, René, and Martin Hájek. 2012. “MtDNA Haplotype Identification of Aurochs Remains Originating from the Czech Republic (Central Europe).” *Environmental Archaeology*. https://doi.org/10.1179/1461410312z.00000000010.
Lachance, Joseph, and Sarah A. Tishkoff. 2013. “SNP Ascertainment Bias in Population Genetic Analyses: Why It Is Important, and How to Correct It.” *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 35 (9): 780–86.
Lambeck, Kurt, Tezer M. Esat, and Emma-Kate Potter. 2002. “Links between Climate and Sea Levels for the Past Three Million Years.” *Nature* 419 (6903): 199–206.
Landau, Serge Yan, Levana Dvash, Philippa Ryan, David Saltz, Tova Deutch, and Steven A. Rosen. 2020. “Faecal Pellets, Rock Shelters, and Seasonality: The Chemistry of Stabling in the Negev of Israel in Late Prehistory.” *Journal of Arid Environments* 181 (October): 104219.
Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, et al. 2001. “Initial Sequencing and Analysis of the Human Genome.” *Nature* 409 (6822): 860–921.
Landsteiner, K. 1900. “Zur Kenntnis Der Antifermentativen, Lytischen Und Agglutinierenden Wirkungen Des Blutserums Und Der Lymphe.” *Zentralblatt Fur Bakteriologie: International Journal of Medical Microbiology* 27: 357–62.
Lari, Martina, Ermanno Rizzi, Stefano Mona, Giorgio Corti, Giulio Catalano, Kefei Chen, Cristiano Vernesi, et al. 2011. “The Complete Mitochondrial Genome of an 11,450-Year-Old Aurochsen (Bos Primigenius) from Central Italy.” *BMC Evolutionary Biology* 11 (January): 32.

Larson, Greger, and Joachim Burger. 2013. “A Population Genetics View of Animal Domestication.” *Trends in Genetics: TIG* 29 (4): 197–205.
Larson, Greger, Elinor K. Karlsson, Angela Perri, Matthew T. Webster, Simon Y. W. Ho, Joris Peters, Peter W. Stahl, et al. 2012. “Rethinking Dog Domestication by Integrating Genetics, Archeology, and Biogeography.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (23): 8878–83.
Lawson, Daniel John, Garrett Hellenthal, Simon Myers, and Daniel Falush. 2012. “Inference of Population Structure Using Dense Haplotype Data.” *PLoS Genetics* 8 (1): e1002453.
Lazaridis, Iosif, Nick Patterson, Alissa Mittnik, Gabriel Renaud, Swapan Mallick, Karola Kirsanow, Peter H. Sudmant, et al. 2014. “Ancient Human Genomes Suggest Three Ancestral Populations for Present-Day Europeans.” *Nature* 513 (7518): 409–13.
Leaché, Adam D., Barbara L. Banbury, Joseph Felsenstein, Adrián Nieto-Montes de Oca, and Alexandros Stamatakis. 2015. “Short Tree, Long Tree, Right Tree, Wrong Tree: New Acquisition Bias Corrections for Inferring SNP Phylogenies.” *Systematic Biology* 64 (6): 1032–47.
Leonardi, Michela, Francesco Boschin, Paolo Boscato, and Andrea Manica. 2022. “Following the Niche: The Differential Impact of the Last Glacial Maximum on Four European Ungulates.” *Communications Biology* 5 (1): 1038.
Leslie, Stephen, Bruce Winney, Garrett Hellenthal, Dan Davison, Abdelhamid Boumertit, Tammy Day, Katarzyna Hutnik, et al. 2015. “The Fine-Scale Genetic Structure of the British Population.” *Nature* 519 (7543): 309–14.
Librado, Pablo, Clio Der Sarkissian, Luca Ermini, Mikkel Schubert, Hákon Jónsson, Anders Albrechtsen, Matteo Fumagalli, et al. 2015. “Tracking the Origins of Yakutian Horses and the Genetic Basis for Their Fast Adaptation to Subarctic Environments.” *Proceedings of the National Academy of Sciences of the United States of America* 112 (50): E6889–97.
Librado, Pablo, Cristina Gamba, Charleen Gaunitz, Clio Der Sarkissian, Mélanie Pruvost, Anders Albrechtsen, Antoine Fages, et al. 2017. “Ancient Genomic Changes Associated with Domestication of the Horse.” *Science* 356 (6336): 442–45.
Librado, Pablo, Naveed Khan, Antoine Fages, Mariya A. Kusliy, Tomasz Suchan, Laure Tonasso-Calvière, Stéphanie Schiavinato, et al. 2021. “The Origins and Spread of Domestic Horses from the Western Eurasian Steppes.” *Nature* 598 (7882): 634–40.
Li, Heng, and Richard Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform.” *Bioinformatics* 25 (14): 1754–60.
———. 2011. “Inference of Human Population History from Individual Whole-Genome Sequences.” *Nature* 475 (7357): 493–96.
Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics* 25 (16): 2078–79.
Lindahl, Tomas. 1993. “Instability and Decay of the Primary Structure of DNA.” *Nature* 362 (6422): 709–15.
Linseele, Veerle, Heiko Riemer, Jan Baeten, Dirk De Vos, Elena Marinova, and Claudio Ottoni. 2013. “Species Identification of Archaeological Dung Remains: A Critical Review of Potential Methods.” *Environmental Archaeology* 18 (1): 5–17.
Lisiecki, Lorraine E., and Maureen E. Raymo. 2005. “A Pliocene-Pleistocene Stack of 57 Globally Distributed Benthic $\delta^{18}O$ Records.” *Paleoceanography*. https://doi.org/10.1029/2004pa001071.
Li, Xingguang, Junjie Zai, Qiang Zhao, Qing Nie, Yi Li, Brian T. Foley, and Antoine Chaillon. 2020. “Evolutionary History, Potential Intermediate Animal Host, and Cross-species Analyses of SARS-CoV-2.” *Journal of Medical Virology*. https://doi.org/10.1002/jmv.25731.
Li, Xin, Ji Yang, Min Shen, Xing-Long Xie, Guang-Jian Liu, Ya-Xi Xu, Feng-Hua Lv, et al. 2020a. “Whole-Genome Resequencing of Wild and Domestic Sheep Identifies Genes Associated with Morphological and Agronomic Traits.” *Nature Communications* 11 (1):

2815.
———. 2020b. “Whole-Genome Resequencing of Wild and Domestic Sheep Identifies Genes Associated with Morphological and Agronomic Traits.” *Nature Communications* 11 (1): 2815.
Loftus, R. T., O. Ertugrul, A. H. Harba, M. A. El-Barody, D. E. MacHugh, S. D. Park, and D. G. Bradley. 1999. “A Microsatellite Survey of Cattle from a Centre of Origin: The Near East.” *Molecular Ecology* 8 (12): 2015–22.
Loftus, R. T., D. E. MacHugh, D. G. Bradley, P. M. Sharp, and P. Cunningham. 1994. “Evidence for Two Independent Domestications of Cattle.” *Proceedings of the National Academy of Sciences of the United States of America* 91 (7): 2757–61.
Lord, Edana, Aurelio Marangoni, Mateusz Baca, Danijela Popović, Anna V. Goropashnaya, John R. Stewart, Monika V. Knul, et al. 2022. “Population Dynamics and Demographic History of Eurasian Collared Lemmings.” *BMC Ecology and Evolution* 22 (1): 126.
Lu, Jennifer, Florian P. Breitwieser, Peter Thielen, and Steven L. Salzberg. 2017. “Bracken: Estimating Species Abundance in Metagenomics Data.” *PeerJ Computer Science* 3 (January): e104.
Lv, Feng-Hua, Wei-Feng Peng, Ji Yang, Yong-Xin Zhao, Wen-Rong Li, Ming-Jun Liu, Yue-Hui Ma, et al. 2015. “Mitogenomic Meta-Analysis Identifies Two Phases of Migration in the History of Eastern Eurasian Sheep.” *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msv139.
MacHugh, D. E., M. D. Shriver, R. T. Loftus, P. Cunningham, and D. G. Bradley. 1997. “Microsatellite DNA Variation and the Evolution, Domestication and Phylogeography of Taurine and Zebu Cattle (Bos Taurus and Bos Indicus).” *Genetics* 146 (3): 1071–86.
Mallick, Swapan, Heng Li, Mark Lipson, Iain Mathieson, Melissa Gymrek, Fernando Racimo, Mengyao Zhao, et al. 2016. “The Simons Genome Diversity Project: 300 Genomes from 142 Diverse Populations.” *Nature* 538 (7624): 201–6.
Mannen, Hideyuki, Takahiro Yonezawa, Kako Murata, Aoi Noda, Fuki Kawaguchi, Shinji Sasazaki, Anna Olivieri, Alessandro Achilli, and Antonio Torroni. 2020. “Cattle Mitogenome Variation Reveals a postglacial Expansion of Haplogroup P and an Early Incorporation into Northeast Asian Domestic Herds.” *Scientific Reports*. https://doi.org/10.1038/s41598-020-78040-8.
Mannen, H., M. Kohno, Y. Nagata, S. Tsuji, D. G. Bradley, J. S. Yeo, D. Nyamsamba, et al. 2004. “Independent Mitochondrial Origin and Historical Genetic Differentiation in North Eastern Asian Cattle.” *Molecular Phylogenetics and Evolution* 32 (2): 539–44.
Mannen, H., M. L. Morimoto, K. Oyamat, F. Mukai, and S. Tsuji. 2003. “Identification of Mitochondrial DNA Substitutions Related to Meat Quality in Japanese Black Cattle.” *Journal of Animal Science* 81 (1): 68–73.
Manning, Katie, Adrian Timpson, Stephen Shennan, and Enrico Crema. 2015. “Size Reduction in Early European Domestic Cattle Relates to Intensification of Neolithic Herding Strategies.” *PloS One* 10 (12): e0141873.
Marchi, Nina, Laura Winkelbach, Ilektra Schulz, Maxime Brami, Zuzana Hofmanová, Jens Blöcher, Carlos S. Reyna-Blanco, et al. 2022. “The Genomic Origins of the World’s First Farmers.” *Cell* 185 (11): 1842–59.e18.
Mardis, Elaine R. 2008. “Next-Generation DNA Sequencing Methods.” *Annual Review of Genomics and Human Genetics* 9: 387–402.
Margaryan, Ashot, Henrik B. Hansen, Simon Rasmussen, Martin Sikora, Vyacheslav Moiseyev, Alexandr Khoklov, Andrey Epimakhov, et al. 2018. “Ancient Pathogen DNA in Human Teeth and Petrous Bones.” *Ecology and Evolution* 8 (6): 3534–42.
Margulies, Marcel, Michael Egholm, William E. Altman, Said Attiya, Joel S. Bader, Lisa A. Bemben, Jan Berka, et al. 2005. “Genome Sequencing in Microfabricated High-Density Picolitre Reactors.” *Nature*. https://doi.org/10.1038/nature03959.
Marom, Nimrod, and Guy Bar-Oz. 2013. “The Prey Pathway: A Regional History of Cattle (Bos Taurus) and Pig (Sus Scrofa) Domestication in the Northern Jordan Valley, Israel.” *PloS One* 8 (2): e55958.

Marsolier-Kergoat, Marie-Claude, Pauline Palacio, Véronique Berthonaud, Frédéric Maksud, Thomas Stafford, Robert Bégouën, and Jean-Marc Elalouf. 2015. “Hunting the Extinct Steppe Bison (Bison Priscus) Mitochondrial Genome in the Trois-Frères Paleolithic Painted Cave.” *PloS One* 10 (6): e0128267.
Martínez-Navarro, Bienvenido, Narjess Karoui-Yaakoub, Oriol Oms, Lamjed Amri, Juan Manuel López-García, Kamel Zerai, Hugues-Alexandre Blain, et al. 2014. “The Early Middle Pleistocene Archeopaleontological Site of Wadi Sarrat (Tunisia) and the Earliest Record of Bos Primigenius.” *Quaternary Science Reviews*. https://doi.org/10.1016/j.quascirev.2014.02.016.
Martínez-Navarro, Bienvenido, Juan Antonio Pérez-Claros, Maria Rita Palombo, Lorenzo Rook, and Paul Palmqvist. 2007. “The Olduvai buffalo *Pelorovis* and the Origin of *Bos*.” *Quaternary Research*. https://doi.org/10.1016/j.yqres.2007.06.002.
Martiniano, Rui, Lara M. Cassidy, Ros Ó’Maoldúin, Russell McLaughlin, Nuno M. Silva, Licinio Manco, Daniel Fidalgo, et al. 2017. “The Population Genomics of Archaeological Transition in West Iberia: Investigation of Ancient Substructure Using Imputation and Haplotype-Based Methods.” *PLOS Genetics*. https://doi.org/10.1371/journal.pgen.1006852.
Martin, Marcel. 2011. “Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads.” *EMBnet.journal* 17 (1): 10–12.
Mashkour, M. 2013. “Animal Exploitation during the Iron Age to Achaemenid, Saasanian and Early Islamic Periods.” *SAUER EW, REKAVANDI HO, WILKINSON TJ & NOKAN-DEH J. (éds), Persia’s Imperial Power in the Late Antiquity. Oxbow Books, Oxford*, 548–80.
Mashkour, Marjan. 2014. “Faunal Report of the 2010 and 2011 Excavation Campaigns.” In *The Archaoelogy of the Salt Miners*, edited by A. Aali and T. Stöllner, 21/1-2:112–17. Bochum: Metatta.
Mashkour, M., H. Fathi, and H. Davoudi. 2020. “Interaktionen Zwischen Mensch Und Tier Im Salzbergwerk Douzlākh, Chehrābād.” In *Tod Im Salz. Eine Archäologische Ermittlung in Persien. Begleitbuch, Katalog Und Graphic Novel*, edited by Stöllner T, Aali A, Bagherpour Kashani N, 219–24.
Massilani, Diyendo, Silvia Guimaraes, Jean-Philip Brugal, E. Andrew Bennett, Malgorzata Tokarska, Rose-Marie Arbogast, Gennady Baryshnikov, et al. 2016. “Past Climate Changes, Population Dynamics and the Origin of Bison in Europe.” *BMC Biology* 14 (1): 93.
Matar, Rachel, and Maxime Merheb. 2016. “Paleogeneticist View of Leather: The Role of Mitochondrial DNA to Uncover the Mysteries of Fake Leather and Its Products.” *Bioscience Biotechnology Research Communications* 9 (1): 6–14.
Mazet, O., W. Rodríguez, S. Grusea, S. Boitard, and L. Chikhi. 2016. “On the Importance of Being Structured: Instantaneous Coalescence Rates and Human Evolution—lessons for Ancestral Population Size Inference?” *Heredity*. https://doi.org/10.1038/hdy.2015.104.
McDonough, Molly M., Lillian D. Parker, Nancy Rotzel McInerney, Michael G. Campana, and Jesús E. Maldonado. 2018. “Performance of Commonly Requested Destructive Museum Samples for Mammalian Genomic Studies.” *Journal of Mammalogy* 99 (4): 789–802.
McKay, Bailey D., and Robert M. Zink. 2010. “The Causes of Mitochondrial DNA Gene Tree Paraphyly in Birds.” *Molecular Phylogenetics and Evolution*. https://doi.org/10.1016/j.ympev.2009.08.024.
McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, et al. 2010. “The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data.” *Genome Research* 20 (9): 1297–1303.
McVean, Gilean A. T., and Niall J. Cardin. 2005. “Approximating the Coalescent with Recombination.” *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2005.1673.
Meadows, J. R. S., S. Hiendleder, and J. W. Kijas. 2011. “Haplogroup Relationships between Domestic and Wild Sheep Resolved Using a Mitogenome Panel.” *Heredity* 106 (4): 700–

706.
Meiri, Meirav, Adrian Lister, Pavel Kosintsev, Grant Zazula, and Ian Barnes. 2020. “Population Dynamics and Range Shifts of Moose (*Alces Alces*) during the Late Quaternary.” *Journal of Biogeography*. https://doi.org/10.1111/jbi.13935.
Meirmans, Patrick G., and Philip W. Hedrick. 2011. “Assessing Population Structure: F(ST) and Related Measures.” *Molecular Ecology Resources* 11 (1): 5–18.
Meisner, Jonas, and Anders Albrechtsen. 2018. “Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data.” *Genetics* 210 (2): 719–31.
Meltzer, David J. 2020. “Overkill, Glacial History, and the Extinction of North America’s Ice Age Megafauna.” *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.2015032117.
Menozzi, P., A. Piazza, and L. Cavalli-Sforza. 1978. “Synthetic Maps of Human Gene Frequencies in Europeans.” *Science* 201 (4358): 786–92.
Meo, G. P. Di, G. P. Di Meo, A. Perucatti, S. Floriot, D. Incarnato, R. Rullo, A. Caputi Jambrenghi, et al. 2005. “Chromosome Evolution and Improved Cytogenetic Maps of the Y Chromosome in Cattle, Zebu, River Buffalo, Sheep and Goat.” *Chromosome Research*. https://doi.org/10.1007/s10577-005-2688-4.
Merheb, Maxime, Stephane Vaiedelich, Thierry Maniguet, and Catherine Haenni. 2015. “DNA for Species Identification in Leather: Fraud Detection and Endangered Species Protection.” *Research Journal of Biotechnology Vol* 10: 9.
Messager, Erwan, Soumaya Belmecheri, Ulrich Von Grafenstein, Sébastien Nomade, Vincent Ollivier, Pierre Voinchet, Simon Puaud, et al. 2013. “Late Quaternary Record of the Vegetation and Catchment-Related Changes from Lake Paravani (Javakheti, South Caucasus).” *Quaternary Science Reviews* 77 (October): 125–40.
Metcalf, Jessica L., Zhenjiang Zech Xu, Sophie Weiss, Simon Lax, Will Van Treuren, Embriette R. Hyde, Se Jin Song, et al. 2016. “Microbial Community Assembly and Metabolic Function during Mammalian Corpse Decomposition.” *Science* 351 (6269): 158–62.
Meyer, Matthias, Juan-Luis Arsuaga, Cesare de Filippo, Sarah Nagel, Ayinuer Aximu-Petri, Birgit Nickel, Ignacio Martínez, et al. 2016. “Nuclear DNA Sequences from the Middle Pleistocene Sima de Los Huesos Hominins.” *Nature* 531 (7595): 504–7.
Meyer, Matthias, Qiaomei Fu, Ayinuer Aximu-Petri, Isabelle Glocke, Birgit Nickel, Juan-Luis Arsuaga, Ignacio Martínez, et al. 2013. “A Mitochondrial Genome Sequence of a Hominin from Sima de Los Huesos.” *Nature* 505 (7483): 403–6.
Meyer, Matthias, and Martin Kircher. 2010. “Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing.” *Cold Spring Harbor Protocols* 2010 (6): db.prot5448.
Meyer, Matthias, Martin Kircher, Marie-Theres Gansauge, Heng Li, Fernando Racimo, Swapan Mallick, Joshua G. Schraiber, et al. 2012. “A High-Coverage Genome Sequence from an Archaic Denisovan Individual.” *Science* 338 (6104): 222–26.
Michel, Amanda. 2014. “Skin Deep: An Outline of the Structure of Different Skins and How It Influences Behaviour in Use. A Practitioner’s Guide.” In *Why Leather?: The Material and Cultural Dimensions of Leather*, edited by Susanna Harris and Andre J. Veldmeijer, 23–40. Leiden: Sidestone Press.
Miller, Webb, Stephan C. Schuster, Andreanna J. Welch, Aakrosh Ratan, Oscar C. Bedoya-Reina, Fangqing Zhao, Hie Lim Kim, et al. 2012. “Polar and Brown Bear Genomes Reveal Ancient Admixture and Demographic Footprints of Past Climate Change.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (36): E2382–90.
Minniti, C., S. Valenzuela-Lamas, J. Evans, and U. Albarella. 2014. “Widening the Market. Strontium Isotope Analysis on Cattle Teeth from Owslebury (Hampshire, UK) Highlights Changes in Livestock Supply between the Iron Age and the Roman Period.” *Journal of Archaeological Science*. https://doi.org/10.1016/j.jas.2013.10.008.
Molak, Martyna, Eline D. Lorenzen, Beth Shapiro, and Simon Y. W. Ho. 2013. “Phylogenetic

Estimation of Timescales Using Ancient DNA: The Effects of Temporal Sampling Scheme and Uncertainty in Sample Ages." *Molecular Biology and Evolution* 30 (2): 253–62.
Mona, Stefano, Giulio Catalano, Martina Lari, Greger Larson, Paolo Boscato, Antonella Casoli, Luca Sineo, et al. 2010. "Population Dynamic of the Extinct European Aurochs: Genetic Evidence of a North-South Differentiation Pattern and No Evidence of postglacial Expansion." *BMC Evolutionary Biology* 10 (March): 83.
Moreira-Grez, Benjamin, Kang Tam, Adam T. Cross, Jean W. H. Yong, Deepak Kumaresan, Paul Nevill, Mark Farrell, and Andrew S. Whiteley. 2019. "The Bacterial Microbiome Associated With Arid Biocrusts and the Biogeochemical Influence of Biocrusts Upon the Underlying Soil." *Frontiers in Microbiology* 10 (September): 2143.
Mullis, Kary B., and Fred A. Faloona. 1987. "[21] Specific Synthesis of DNA in Vitro via a Polymerase-Catalyzed Chain Reaction." *Methods in Enzymology*. https://doi.org/10.1016/0076-6879(87)55023-6.
Murchie, Tyler J., Melanie Kuch, Ana T. Duggan, Marissa L. Ledger, Kévin Roche, Jennifer Klunk, Emil Karpinski, et al. 2021. "Optimizing Extraction and Targeted Capture of Ancient Environmental DNA for Reconstructing Past Environments Using the PalaeoChip Arctic-1.0 Bait-Set." *Quaternary Research* 99 (January): 305–28.
Mwai, Okeyo, Olivier Hanotte, Young-Jun Kwon, and Seoae Cho. 2015. "African Indigenous Cattle: Unique Genetic Resources in a Rapidly Changing World." *Asian-Australasian Journal of Animal Sciences* 28 (7): 911–21.
Na, Ri-Su, Yong-Ju Zhao, Hui-Jiang Gao, Tian-Wu An, Yong-Fu Huang, and Guang-Xin E. 2016. "Complete Mitochondrial Genome of the Yakow (Bos Primigenius Taurus × Bos Grunniens) in China." *Mitochondrial DNA. Part A, DNA Mapping, Sequencing, and Analysis* 27 (6): 3826–27.
Nasab, Hamed Vahdati, Sanaz Shirvani, and Solange Rigaud. 2019. "The Northern Iranian Central Plateau at the End of the Pleistocene and Early Holocene: The Emergence of Domestication." *Journal of World Prehistory*. https://doi.org/10.1007/s10963-019-09133-0.
Nayfach, Stephen, Beltran Rodriguez-Mueller, Nandita Garud, and Katherine S. Pollard. 2016. "An Integrated Metagenomics Pipeline for Strain Profiling Reveals Novel Patterns of Bacterial Transmission and Biogeography." *Genome Research*. https://doi.org/10.1101/gr.201863.115.
Nei, M., and W. H. Li. 1979. "Mathematical Model for Studying Genetic Variation in Terms of Restriction Endonucleases." *Proceedings of the National Academy of Sciences of the United States of America* 76 (10): 5269–73.
Nezamabadi, M., A. Aali, Th Stöllner, M. Mashkour, and M. Le Bailly. 2013. "Paleoparasitological Analysis of Samples from the Chehrabad Salt Mine (Northwestern Iran)." *International Journal of Paleopathology* 3 (3): 229–33.
Nielsen, Rasmus, Joshua S. Paul, Anders Albrechtsen, and Yun S. Song. 2011. "Genotype and SNP Calling from next-Generation Sequencing Data." *Nature Reviews. Genetics* 12 (6): 443–51.
Nishibuchi, Gohei, and Jérôme Déjardin. 2017. "The Molecular Basis of the Organization of Repetitive DNA-Containing Constitutive Heterochromatin in Mammals." *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology* 25 (1): 77–87.
Noonan, James P., Michael Hofreiter, Doug Smith, James R. Priest, Nadin Rohland, Gernot Rabeder, Johannes Krause, J. Chris Detter, Svante Pääbo, and Edward M. Rubin. 2005. "Genomic Sequencing of Pleistocene Cave Bears." *Science* 309 (5734): 597–99.
Öhrström, Lena M., Roger Seiler, Thomas Böni, Abolfazl Aali, Thomas Stöllner, and Frank J. Rühli. 2016. "Erratum to: Radiological Findings in an Ancient Iranian Salt Mummy (Chehrābād Ca. 410-350 BC)." *Skeletal Radiology* 45 (3): 433.
Okonechnikov, Konstantin, Ana Conesa, and Fernando García-Alcalde. 2016. "Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data."

Bioinformatics 32 (2): 292–94.
Olivieri, Anna, Francesca Gandini, Alessandro Achilli, Alessandro Fichera, Ermanno Rizzi, Silvia Bonfiglio, Vincenza Battaglia, et al. 2015. “Mitogenomes from Egyptian Cattle Breeds: New Clues on the Origin of Haplogroup Q and the Early Spread of Bos Taurus from the Near East.” *PloS One* 10 (10): e0141170.
Orlando, Ludovic. 2015. “The First Aurochs Genome Reveals the Breeding History of British and European Cattle.” *Genome Biology*.
———. 2020. “The Evolutionary and Historical Foundation of the Modern Horse: Lessons from Ancient Genomics.” *Annual Review of Genetics* 54 (November): 563–81.
Orlando, Ludovic, Aurélien Ginolhac, Guojie Zhang, Duane Froese, Anders Albrechtsen, Mathias Stiller, Mikkel Schubert, et al. 2013. “Recalibrating Equus Evolution Using the Genome Sequence of an Early Middle Pleistocene Horse.” *Nature* 499 (7456): 74–78.
O’Sullivan, Niall J., Matthew D. Teasdale, Valeria Mattiangeli, Frank Maixner, Ron Pinhasi, Daniel G. Bradley, and Albert Zink. 2016. “A Whole Mitochondria Analysis of the Tyrolean Iceman’s Leather Provides Insights into the Animal Sources of Copper Age Clothing.” *Scientific Reports*. https://doi.org/10.1038/srep31279.
Outram, Alan K., Alexei Kasparov, Natalie A. Stear, Victor Varfolomeev, Emma Usmanova, and Richard P. Evershed. 2012. “Patterns of Pastoralism in Later Bronze Age Kazakhstan: New Evidence from Faunal and Lipid Residue Analyses.” *Journal of Archaeological Science*. https://doi.org/10.1016/j.jas.2012.02.009.
Pääbo, S. 1985. “Molecular Cloning of Ancient Egyptian Mummy DNA.” *Nature* 314 (6012): 644–45.
———. 1989. “Ancient DNA: Extraction, Characterization, Molecular Cloning, and Enzymatic Amplification.” *Proceedings of the National Academy of Sciences of the United States of America* 86 (6): 1939–43.
Pääbo, S., J. A. Gifford, and A. C. Wilson. 1988. “Mitochondrial DNA Sequences from a 7000-Year Old Brain.” *Nucleic Acids Research* 16 (20): 9775–87.
Pääbo, Svante, Hendrik Poinar, David Serre, Viviane Jaenicke-Despres, Juliane Hebler, Nadin Rohland, Melanie Kuch, Johannes Krause, Linda Vigilant, and Michael Hofreiter. 2004. “Genetic Analyses from Ancient DNA.” *Annual Review of Genetics* 38: 645–79.
Palacio, Pauline, Véronique Berthonaud, Claude Guérin, Josie Lambourdière, Frédéric Maksud, Michel Philippe, Delphine Plaire, Thomas Stafford, Marie-Claude Marsolier-Kergoat, and Jean-Marc Elalouf. 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.
Palkopoulou, Eleftheria, Mateusz Baca, Natalia I. Abramson, Mikhail Sablin, Paweł Socha, Adam Nadachowski, Stefan Prost, et al. 2016. “Synchronous Genetic Turnovers across Western Eurasia in Late Pleistocene Collared Lemmings.” *Global Change Biology* 22 (5): 1710–21.
Palkopoulou, Eleftheria, Love Dalén, Adrian M. Lister, Sergey Vartanyan, Mikhail Sablin, Andrei Sher, Veronica Nyström Edmark, et al. 2013. “Holarctic Genetic Structure and Range Dynamics in the Woolly Mammoth.” *Proceedings. Biological Sciences / The Royal Society* 280 (1770): 20131910.
Palkopoulou, Eleftheria, Swapan Mallick, Pontus Skoglund, Jacob Enk, Nadin Rohland, Heng Li, Ayça Omrak, et al. 2015. “Complete Genomes Reveal Signatures of Demographic and Genetic Declines in the Woolly Mammoth.” *Current Biology: CB* 25 (10): 1395–1400.
Pálsdóttir, Albína Hulda, Auli Bläuer, Eve Rannamäe, Sanne Boessenkool, and Jón Hallsteinn Hallsson. 2019. “Not a Limitless Resource: Ethics and Guidelines for Destructive Sampling of Archaeofaunal Remains.” *Royal Society Open Science* 6 (10): 191059.
Pamilo, P., and M. Nei. 1988. “Relationships between Gene Trees and Species Trees.” *Molecular Biology and Evolution* 5 (5): 568–83.
Paradis, Emmanuel, Julien Claude, and Korbinian Strimmer. 2004. “APE: Analyses of

Phylogenetics and Evolution in R Language." *Bioinformatics* 20 (2): 289–90.
Parks, Matthew, and David Lambert. 2015. "Impacts of Low Coverage Depths and Post-Mortem DNA Damage on Variant Calling: A Simulation Study." *BMC Genomics* 16 (1): 19.
Park, Stephen D. E., David A. Magee, Paul A. McGettigan, Matthew D. Teasdale, Ceiridwen J. Edwards, Amanda J. Lohan, Alison Murphy, et al. 2015. "Genome Sequencing of the Extinct Eurasian Wild Aurochs, Bos Primigenius, Illuminates the Phylogeography and Evolution of Cattle." *Genome Biology* 16 (October): 234.
Patel, A. K., and R. H. Meadow. 2017. "South Asian Contributions to Animal Domestication and Pastoralism." In *The Oxford Handbook of Zooarchaeology*, edited by Umberto Albarella, Mauro Rizzetto, Hannah Russ, Kim Vickers, and Sarah Viner-Daniels, 280–303. Oxford: Oxford University Press.
Patterson, Nick, Priya Moorjani, Yontao Luo, Swapan Mallick, Nadin Rohland, Yiping Zhan, Teri Genschoreck, Teresa Webster, and David Reich. 2012. "Ancient Admixture in Human History." *Genetics* 192 (3): 1065–93.
Patterson, Nick, Alkes L. Price, and David Reich. 2006. "Population Structure and Eigenanalysis." *PLoS Genetics* 2 (12): e190.
Patton, Henry, Alun Hubbard, Karin Andreassen, Amandine Auriac, Pippa L. Whitehouse, Arjen P. Stroeven, Calvin Shackleton, Monica Winsborrow, Jakob Heyman, and Adrian M. Hall. 2017. "Deglaciation of the Eurasian Ice Sheet Complex." *Quaternary Science Reviews*. https://doi.org/10.1016/j.quascirev.2017.05.019.
Payne, Sebastian. 1973. "Kill-off Patterns in Sheep and Goats: The Mandibles from Aşvan Kale." *Anatolian Studies*. https://doi.org/10.2307/3642547.
Pebesma, Edzer. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *The R Journal* 10 (1): 439–46.
Pereira, Filipe, Simon J. M. Davis, Luísa Pereira, Brian McEvoy, Daniel G. Bradley, and António Amorim. 2006. "Genetic Signatures of a Mediterranean Influence in Iberian Peninsula Sheep Husbandry." *Molecular Biology and Evolution* 23 (7): 1420–26.
Pérez-Pardal, L., L. J. Royo, A. Beja-Pereira, I. Curik, A. Traoré, I. Fernández, J. Sölkner, et al. 2010. "Y-Specific Microsatellites Reveal an African Subfamily in Taurine (Bos Taurus) Cattle." *Animal Genetics* 41 (3): 232–41.
Perry, George H., John C. Marioni, Páll Melsted, and Yoav Gilad. 2010. "Genomic-Scale Capture and Sequencing of Endogenous DNA from Feces." *Molecular Ecology* 19 (24): 5332.
Peters, J., B. S. Arbuckle, and N. Pöllath. 2014. "Subsistence and beyond: Animals in Neolithic Anatolia." *The Neolithic*.
Pickrell, Joseph K., and Jonathan K. Pritchard. 2012. "Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data." *PLoS Genetics* 8 (11): e1002967.
Pitt, Daniel, Natalia Sevane, Ezequiel L. Nicolazzi, David E. MacHugh, Stephen D. E. Park, Licia Colli, Rodrigo Martinez, Michael W. Bruford, and Pablo Orozco-terWengel. 2019. "Domestication of Cattle: Two or Three Events?" *Evolutionary Applications* 12 (1): 123–36.
Pitulko, Vladimir V., and Aleksey K. Kasparov. 2017. "Archaeological Dogs from the Early Holocene Zhokhov Site in the Eastern Siberian Arctic." *Journal of Archaeological Science: Reports* 13 (June): 491–515.
Poinar, Hendrik N., Carsten Schwarz, Ji Qi, Beth Shapiro, Ross D. E. Macphee, Bernard Buigues, Alexei Tikhonov, et al. 2006. "Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA." *Science* 311 (5759): 392–94.
Pöllath, Nadja, Oliver Dietrich, Jens Notroff, Lee Clare, Laura Dietrich, Çiğdem Köksal-Schmidt, Klaus Schmidt, and Joris Peters. 2018. "Almost a Chest Hit: An Aurochs Humerus with Hunting Lesion from Göbekli Tepe, South-Eastern Turkey, and Its Implications." *Quaternary International*. https://doi.org/10.1016/j.quaint.2017.12.003.
Pöllath, Nadja, Renate Schafberg, and Joris Peters. 2019. "Astragalar Morphology: Approaching the Cultural Trajectories of Wild and Domestic Sheep Applying Geometric

Morphometrics." *Journal of Archaeological Science: Reports* 23 (February): 810–21.
Posth, Cosimo, Gabriel Renaud, Alissa Mittnik, Dorothée G. Drucker, Hélène Rougier, Christophe Cupillard, Frédérique Valentin, et al. 2016. "Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe." *Current Biology: CB* 26 (6): 827–33.
Prabhu, Vandana R., Moolamkudy Suresh Arjun, Karippadakam Bhavana, Ranganathan Kamalakkannan, and Muniyandi Nagarajan. 2019. "Complete Mitochondrial Genome of Indian Mithun, Bos Frontalis and Its Phylogenetic Implications." *Molecular Biology Reports* 46 (2): 2561–66.
Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. "Inference of Population Structure Using Multilocus Genotype Data." *Genetics* 155 (2): 945–59.
Pujolar, J. M., L. Dalén, M. M. Hansen, and J. Madsen. 2017. "Demographic Inference from Whole-Genome and RAD Sequencing Data Suggests Alternating Human Impacts on Goose Populations since the Last Ice Age." *Molecular Ecology*. https://doi.org/10.1111/mec.14374.
Qian, Hong, and Robert E. Ricklefs. 2001. "Diversity of Temperate Plants in East Asia." *Nature*. https://doi.org/10.1038/35093169.
Qi, W., R. J. Reiter, D. X. Tan, J. J. Garcia, L. C. Manchester, M. Karbownik, and J. R. Calvo. 2000. "Chromium(III)-Induced 8-Hydroxydeoxyguanosine in DNA and Its Reduction by Antioxidants: Comparative Effects of Melatonin, Ascorbate, and Vitamin E." *Environmental Health Perspectives* 108 (5): 399–402.
Quievryn, G., E. Peterson, J. Messer, and A. Zhitkovich. 2003. "Genotoxicity and Mutagenicity of chromium(VI)/ascorbate-Generated DNA Adducts in Human and Bacterial Cells." *Biochemistry* 42 (4). https://doi.org/10.1021/bi0271547.
Quinlan, Aaron R., and Ira M. Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics* 26 (6): 841–42.
Racimo, Fernando, Sriram Sankararaman, Rasmus Nielsen, and Emilia Huerta-Sánchez. 2015. "Evidence for Archaic Adaptive Introgression in Humans." *Nature Reviews. Genetics* 16 (6): 359–71.
Rampelli, Simone, Silvia Turroni, Carolina Mallol, Cristo Hernandez, Bertila Galván, Ainara Sistiaga, Elena Biagi, et al. 2021. "Components of a Neanderthal Gut Microbiome Recovered from Fecal Sediments from El Salt." *Communications Biology* 4 (1): 1–10.
Ramsey, Christopher Bronk. 2009. "Bayesian Analysis of Radiocarbon Dates." *Radiocarbon* 51 (1): 337–60.
Rasmussen, Morten, Yingrui Li, Stinus Lindgreen, Jakob Skou Pedersen, Anders Albrechtsen, Ida Moltke, Mait Metspalu, et al. 2010. "Ancient Human Genome Sequence of an Extinct Palaeo-Eskimo." *Nature* 463 (7282): 757–62.
Rast-Eicher, Antoinette. 2010. "Fell- Und Lederreste Aus Den Gräbern 2008." In *Spiez-Einigen, Holleeweg 3, Naturwissenschaftliche Untersuchungen Zu Den Bronzezeitlichen Bestattungen*, edited by Christine Cooper, Michaela Harbeck, Marlu Kühn, Antoinette Rast-Eicher, Mike Schweissing, Susi Ulrich-Bochsler, and Patricia Vandorpe, 175–98. Bern: Service Archéologique du Canton de Berne.
———. 2016. *Fibres: Microscopy of Archaeological Textiles and Furs*. Budapest: Archaeolingua Alapítvány.
Raxworthy, Christopher J., and Brian Tilston Smith. 2021. "Mining Museums for Historical DNA: Advances and Challenges in Museomics." *Trends in Ecology & Evolution* 36 (11): 1049–60.
Rego, Alena, Paul B. Sinclair, Wei Tao, Igor Kireev, and Andrew S. Belmont. 2008. "The Facultative Heterochromatin of the Inactive X Chromosome Has a Distinctive Condensed Ultrastructure." *Journal of Cell Science* 121 (Pt 7): 1119–27.
Reich, David, Richard E. Green, Martin Kircher, Johannes Krause, Nick Patterson, Eric Y. Durand, Bence Viola, et al. 2010. "Genetic History of an Archaic Hominin Group from Denisova Cave in Siberia." *Nature* 468 (7327): 1053–60.
Reich, David, Kumarasamy Thangaraj, Nick Patterson, Alkes L. Price, and Lalji Singh. 2009.

"Reconstructing Indian Population History." *Nature* 461 (7263): 489–94.
Reimer, Paula J., William E. N. Austin, Edouard Bard, Alex Bayliss, Paul G. Blackwell, Christopher Bronk Ramsey, Martin Butzin, et al. 2020. "The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 Cal kBP)." *Radiocarbon* 62 (4): 725–57.
Reimer, Paula J., Edouard Bard, Alex Bayliss, J. Warren Beck, Paul G. Blackwell, Christopher Bronk Ramsey, Caitlin E. Buck, et al. 2013. "IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years Cal BP." *Radiocarbon* 55 (4): 1869–87.
Renaud, Gabriel, Kristian Hanghøj, Thorfinn Sand Korneliussen, Eske Willerslev, and Ludovic Orlando. 2019. "Joint Estimates of Heterozygosity and Runs of Homozygosity for Modern and Ancient Samples." *Genetics* 212 (3): 587–614.
Richerson, Peter J., Robert Boyd, and Robert L. Bettinger. 2001. "Was Agriculture Impossible during the Pleistocene but Mandatory during the Holocene? A Climate Change Hypothesis." *American Antiquity* 66 (3): 387–411.
Rivollat, Maïté, Choongwon Jeong, Stephan Schiffels, İşil Küçükkalıpçı, Marie-Hélène Pemonge, Adam Benjamin Rohrlach, Kurt W. Alt, et al. 2020. "Ancient Genome-Wide DNA from France Highlights the Complexity of Interactions between Mesolithic Hunter-Gatherers and Neolithic Farmers." *Science Advances* 6 (22): eaaz5344.
Rizzi, Ermanno, Martina Lari, Elena Gigli, Gianluca De Bellis, and David Caramelli. 2012. "Ancient DNA Studies: New Perspectives on Old Samples." *Genetics, Selection, Evolution: GSE* 44 (July): 21.
Roberts, R. G., T. F. Flannery, L. K. Ayliffe, H. Yoshida, J. M. Olley, G. J. Prideaux, G. M. Laslett, et al. 2001. "New Ages for the Last Australian Megafauna: Continent-Wide Extinction about 46,000 Years Ago." *Science*.
Robinson, Courtney K., Jacek Wierzchos, Celeste Black, Alexander Crits-Christoph, Bing Ma, Jacques Ravel, Carmen Ascaso, et al. 2015. "Microbial Diversity and the Presence of Algae in Halite Endolithic Communities Are Correlated to Atmospheric Moisture in the Hyper-Arid Zone of the Atacama Desert." *Environmental Microbiology* 17 (2): 299–315.
Robinson, James T., Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, and Jill P. Mesirov. 2011. "Integrative Genomics Viewer." *Nature Biotechnology* 29 (1): 24–26.
Rosen, Steven A. 2013. "Evolution in the Desert: Scale and Discontinuity in the Central Negev (Israel) in the Fourth Millennium BCE." *Paléorient* 39 (1): 139–48.
Rosen, Steven A., Arkady B. Savinetsky, Yosef Plakht, Nina K. Kisseleva, Bulat F. Khassanov, Andrey M. Pereladov, and Mordecai Haiman. 2005. "Dung in the Desert: Preliminary Results of the Negev Holocene Ecology Project." *Current Anthropology* 46 (2): 317–26.
Rosli, Norsyamimi, Frankie Thomas Sitam, Jeffrine Japning Rovie-Ryan, Han Ming Gan, Yin Peng Lee, Hartini Ithnin, Millawati Gani, Mohd Firdaus Ariff Abdul Razak, Badrul Munir Md-Zain, and Mohd Tajuddin Abdullah. 2019. "The Complete Mitochondrial Genome of Malayan Gaur (Bos Gaurus Hubbacki) from Peninsular Malaysia." *Mitochondrial DNA. Part B, Resources* 4 (2): 2535–36.
Ross, Ashley A., Kirsten M. Müller, J. Scott Weese, and Josh D. Neufeld. 2018. "Comprehensive Skin Microbiome Analysis Reveals the Uniqueness of Human Skin and Evidence for Phylosymbiosis within the Class Mammalia." *Proceedings of the National Academy of Sciences of the United States of America* 115 (25): E5786–95.
Rossi, Conor, Gabriela Ruß-Popa, Valeria Mattiangeli, Fionnuala McDaid, Andrew J. Hare, Hossein Davoudi, Haeedeh Laleh, et al. 2021. "Exceptional Ancient DNA Preservation and Fibre Remains of a Sasanian Saltmine Sheep Mummy in Chehrābād, Iran." *Biology Letters* 17 (7): 20210222.
Ruß-Popa, Gabriela. 2018. "Der Gebrauch von Schaffell in Der Mitteleuropäischen Urgeschichtlichen Bekleidung." *Annalen Des Naturhistorischen Museums in Wien. Serie A Für Mineralogie Und Petrographie, Geologie Und Paläontologie, Anthropologie Und*

Prähistorie 120: 157–76.
Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1988. “Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase.” *Science* 239 (4839): 487–91.
Saitou, N., and M. Nei. 1987. “The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees.” *Molecular Biology and Evolution* 4 (4): 406–25.
Salque, Mélanie, Peter I. Bogucki, Joanna Pyzel, Iwona Sobkowiak-Tabaka, Ryszard Grygiel, Marzena Szmyt, and Richard P. Evershed. 2013. “Earliest Evidence for Cheese Making in the Sixth Millennium BC in Northern Europe.” *Nature* 493 (7433): 522–25.
Salter, Susannah J., Michael J. Cox, Elena M. Turek, Szymon T. Calus, William O. Cookson, Miriam F. Moffatt, Paul Turner, Julian Parkhill, Nicholas J. Loman, and Alan W. Walker. 2014. “Reagent and Laboratory Contamination Can Critically Impact Sequence-Based Microbiome Analyses.” *BMC Biology* 12 (1): 1–12.
Sambrook, Joseph, E. F. Fritsch, and Tom Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*.
Sandom, Christopher, Søren Faurby, Brody Sandel, and Jens-Christian Svenning. 2014. “Global Late Quaternary Megafauna Extinctions Linked to Humans, Not Climate Change.” *Proceedings of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rspb.2013.3254.
Sanger, F., S. Nicklen, and A. R. Coulson. 1977. “DNA Sequencing with Chain-Terminating Inhibitors.” *Proceedings of the National Academy of Sciences of the United States of America* 74 (12): 5463–67.
Sardina, M. T., M. Ballester, J. Marmi, R. Finocchiaro, J. B. C. H. M. van Kaam, B. Portolano, and J. M. Folch. 2006. “Phylogenetic Analysis of Sicilian Goats Reveals a New mtDNA Lineage.” *Animal Genetics* 37 (4): 376–78.
Sawyer, Susanna, Johannes Krause, Katerina Guschanski, Vincent Savolainen, and Svante Pääbo. 2012. “Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA.” *PloS One* 7 (3): e34131.
Scheu, Amelie, Adam Powell, Ruth Bollongino, Jean-Denis Vigne, Anne Tresset, Canan Çakırlar, Norbert Benecke, and Joachim Burger. 2015. “The Genetic Prehistory of Domesticated Cattle from Their Origin to the Spread across Europe.” *BMC Genetics* 16 (May): 54.
Schiffels, Stephan, and Ke Wang. 2020. “MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent.” *Methods in Molecular Biology* 2090: 147–66.
Schmieder, Robert, and Robert Edwards. 2011. “Quality Control and Preprocessing of Metagenomic Datasets.” *Bioinformatics* 27 (6): 863–64.
Schröder, Oskar, Mayke Wagner, Saskia Wutke, Yong Zhang, Yingxia Ma, Dongliang Xu, Tomasz Goslar, Reinder Neef, Pavel E. Tarasov, and Arne Ludwig. 2016. “Ancient DNA Identification of Domestic Animals Used for Leather Objects in Central Asia during the Bronze Age.” *The Holocene*. https://doi.org/10.1177/0959683616641741.
Schubert, Mikkel, Stinus Lindgreen, and Ludovic Orlando. 2016. “AdapterRemoval v2: Rapid Adapter Trimming, Identification, and Read Merging.” *BMC Research Notes* 9 (1): 1–7.
Schuenemann, Verena J., Charlotte Avanzi, Ben Krause-Kyora, Alexander Seitz, Alexander Herbig, Sarah Inskip, Marion Bonazzi, et al. 2018. “Ancient Genomes Reveal a High Diversity of Mycobacterium Leprae in Medieval Europe.” *PLoS Pathogens* 14 (5): e1006997.
Schulz, Ellen, and Thomas M. KaiSer. 2007. “Feeding Strategy of the Urus Bos Primigenius BOJANUS, 1827 from the Holocene of Denmark.” *COURIER-FORSCHUNGSINSTITUT SENCKENBERG* 259: 155.
Schwarz, Carsten, Regis Debruyne, Melanie Kuch, Elizabeth McNally, Henry Schwarcz, Andrew D. Aubrey, Jeffrey Bada, and Hendrik Poinar. 2009. “New Insights from Old Bones: DNA Preservation and Degradation in Permafrost Preserved Mammoth Remains.” *Nucleic Acids Research* 37 (10): 3215–29.
Serjeantson, Dale. 2011. “Review of Animal Remains from the Neolithic and Early Bronze

Age of Southern Britain," 29-2011, , 158.
Shackleton, N. J. 2000. "The 100,000-Year Ice-Age Cycle Identified and Found to Lag Temperature, Carbon Dioxide, and Orbital Eccentricity." *Science* 289 (5486): 1897–1902.
Shakun, Jeremy D., and Anders E. Carlson. 2010. "A Global Perspective on Last Glacial Maximum to Holocene Climate Change." *Quaternary Science Reviews* 29 (15): 1801–16.
Sharma, Rekha, Amit Kishore, Manishi Mukesh, Sonika Ahlawat, Avishek Maitra, Ashwni Kumar Pandey, and Madhu Sudan Tantia. 2015. "Genetic Diversity and Relationship of Indian Cattle Inferred from Microsatellite and Mitochondrial DNA Markers." *BMC Genetics* 16 (June): 73.
Sheridan, Alison. 2010. "The Neolithization of Britain and Ireland: The 'big Picture.'" *Landscapes in Transition*, 89–105.
Shillito, Lisa-Marie, John C. Blong, Eleanor J. Green, and Eline N. van Asperen. 2020. "The What, How and Why of Archaeological Coprolite Analysis." *Earth-Science Reviews* 207 (August): 103196.
Shinde, Vasant, Vagheesh M. Narasimhan, Nadin Rohland, Swapan Mallick, Matthew Mah, Mark Lipson, Nathan Nakatsuka, et al. 2019. "An Ancient Harappan Genome Lacks Ancestry from Steppe Pastoralists or Iranian Farmers." *Cell* 179 (3): 729–35.e10.
Sikora, Martin, Vladimir V. Pitulko, Vitor C. Sousa, Morten E. Allentoft, Lasse Vinner, Simon Rasmussen, Ashot Margaryan, et al. 2019. "The Population History of Northeastern Siberia since the Pleistocene." *Nature*, June, 1.
Sinding, Mikkel-Holger S., Marta M. Ciucani, Jazmín Ramos-Madrigal, Alberto Carmagnini, Jacob Agerbo Rasmussen, Shaohong Feng, Guangji Chen, et al. 2021. "Kouprey () Genomes Unveil Polytomic Origin of Wild Asian." *iScience* 24 (11): 103226.
Sirak, Kendra A., Elizabeth A. Sawchuk, and Mary E. Prendergast. 2022. "Ancient Human DNA and African Population History." In *Oxford Research Encyclopedia of Anthropology*. Oxford University Press.
Skoglund, Pontus, Bernd H. Northoff, Michael V. Shunkov, Anatoli P. Derevianko, Svante Pääbo, Johannes Krause, and Mattias Jakobsson. 2014. "Separating Endogenous Ancient DNA from Modern Day Contamination in a Siberian Neandertal." *Proceedings of the National Academy of Sciences of the United States of America* 111 (6): 2229–34.
Skourtanioti, Eirini, Yilmaz S. Erdal, Marcella Frangipane, Francesca Balossi Restelli, K. Aslıhan Yener, Frances Pinnock, Paolo Matthiae, et al. 2020. "Genomic History of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern Caucasus." *Cell* 181 (5): 1158–75.e28.
Smeds, Linnéa, Ilpo Kojola, and Hans Ellegren. 2019. "The Evolutionary History of Grey Wolf Y Chromosomes." *Molecular Ecology* 28 (9): 2173–91.
Smith, Oliver, Garry Momber, Richard Bates, Paul Garwood, Simon Fitch, Mark Pallen, Vincent Gaffney, and Robin G. Allaby. 2015. "Archaeology. Sedimentary DNA from a Submerged Site Reveals Wheat in the British Isles 8000 Years Ago." *Science* 347 (6225): 998–1001.
Snelling, Warren M., Readman Chiu, Jacqueline E. Schein, Matthew Hobbs, Colette A. Abbey, David L. Adelson, Jan Aerts, et al. 2007. "A Physical Map of the Bovine Genome." *Genome Biology* 8 (8): R165.
Soubrier, Julien, Graham Gower, Kefei Chen, Stephen M. Richards, Bastien Llamas, Kieren J. Mitchell, Simon Y. W. Ho, et al. 2016. "Early Cave Art and Ancient DNA Record the Origin of European Bison." *Nature Communications* 7 (October): 13158.
Stephens, Lucas, Dorian Fuller, Nicole Boivin, Torben Rick, Nicolas Gauthier, Andrea Kay, Ben Marwick, et al. 2019. "Archaeological Assessment Reveals Earth's Early Transformation through Land Use." *Science* 365 (6456): 897–902.
Stöllner, T., A. Aali, and N. Bagherpour Kashani. 2020. *Tod Im Salz. Eine Archäologische Ermittlung in Persien. Begleitbuch, Katalog Und Graphic Novel*. Edited by T. Stöllner, A. Aali, and N. Bagherpour Kashani. Bochum: Nünnerich-Asmus Verlag Media.

Stoneking, M. 1995. “Ancient DNA: How Do You Know When You Have It and What Can You Do with It?” *American Journal of Human Genetics* 57 (6): 1259–62.
Stringer, Chris. 2016. “The Origin and Evolution of Homo Sapiens.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 371 (1698). https://doi.org/10.1098/rstb.2015.0237.
Strucken, Eva M., Netsanet Z. Gebrehiwot, Marimuthu Swaminathan, Sachin Joshi, Mohammad Al Kalaldeh, and John P. Gibson. 2021. “Genetic Diversity and Effective Population Sizes of Thirteen Indian Cattle Breeds.” *Genetics, Selection, Evolution: GSE* 53 (1): 47.
Stuart, A. J. 1991. “Mammalian Extinctions in the Late Pleistocene of Northern Eurasia and North America.” *Biological Reviews of the Cambridge Philosophical Society* 66 (4): 453–562.
Sullivan, Jack, and Paul Joyce. 2005. “Model Selection in Phylogenetics.” *Annual Review of Ecology, Evolution, and Systematics*. https://doi.org/10.1146/annurev.ecolsys.36.102003.152633.
Svendsen, John Inge, Helena Alexanderson, Valery I. Astakhov, Igor Demidov, Julian A. Dowdeswell, Svend Funder, Valery Gataullin, et al. 2004. “Late Quaternary Ice Sheet History of Northern Eurasia.” *Quaternary Science Reviews*. https://doi.org/10.1016/j.quascirev.2003.12.008.
Svensson, Emma, and Anders Götherström. 2008. “Temporal Fluctuations of Y-Chromosomal Variation in Bos Taurus.” *Biology Letters* 4 (6): 752–54.
Taberlet, P., L. Fumagalli, A. G. Wust-Saucy, and J. F. Cosson. 1998. “Comparative Phylogeography and Postglacial Colonization Routes in Europe.” *Molecular Ecology* 7 (4): 453–64.
Taylor, William T. T., Mélanie Pruvost, Cosimo Posth, William Rendu, Maciej T. Krajcarz, Aida Abdykanova, Greta Brancaleoni, et al. 2021. “Evidence for Early Dispersal of Domestic Sheep into Central Asia.” *Nature Human Behaviour*, April. https://doi.org/10.1038/s41562-021-01083-y.
Teasdale, Matthew D., Sarah Fiddyment, Jiří Vnouček, Valeria Mattiangeli, Camilla Speller, Annelise Binois, Martin Carver, et al. 2017. “The York Gospels: A 1000-Year Biological Palimpsest.” *Royal Society Open Science* 4 (10): 170988.
Teasdale, M. D., N. L. van Doorn, S. Fiddyment, C. C. Webb, T. O’Connor, M. Hofreiter, M. J. Collins, and D. G. Bradley. 2015. “Paging through History: Parchment as a Reservoir of Ancient DNA for next Generation Sequencing.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, January. https://doi.org/10.1098/rstb.2013.0379.
Thomas, R. H., W. Schaffner, A. C. Wilson, and S. Pääbo. 1989. “DNA Phylogeny of the Extinct Marsupial Wolf.” *Nature* 340 (6233): 465–67.
Timmermann, Axel, and Tobias Friedrich. 2016. “Late Pleistocene Climate Drivers of Early Human Migration.” *Nature* 538 (7623): 92–95.
Tito, Raul Y., Dan Knights, Jessica Metcalf, Alexandra J. Obregon-Tito, Lauren Cleeland, Fares Najar, Bruce Roe, et al. 2012. “Insights from Characterizing Extinct Human Gut Microbiomes.” *PloS One* 7 (12): e51146.
Tito, Raúl Y., Simone Macmil, Graham Wiley, Fares Najar, Lauren Cleeland, Chunmei Qu, Ping Wang, et al. 2008. “Phylotyping and Functional Analysis of Two Ancient Human Microbiomes.” *PloS One* 3 (11): e3703.
Troy, C. S., D. E. MacHugh, J. F. Bailey, D. A. Magee, R. T. Loftus, P. Cunningham, A. T. Chamberlain, B. C. Sykes, and D. G. Bradley. 2001. “Genetic Evidence for Near-Eastern Origins of European Cattle.” *Nature* 410 (6832): 1088–91.
Turvey, Samuel T., Haowen Tong, Anthony J. Stuart, and Adrian M. Lister. 2013. “Holocene Survival of Late Pleistocene Megafauna in China: A Critical Review of the Evidence.” *Quaternary Science Reviews* 76 (September): 156–66.
Upadhyay, Maulik, Chiara Bortoluzzi, Mario Barbato, Paolo Ajmone-Marsan, Licia Colli, Catarina Ginja, Tad S. Sonstegard, et al. 2019. “Deciphering the Patterns of Genetic

Admixture and Diversity in Southern European Cattle Using Genome-Wide SNPs." *Evolutionary Applications* 12 (5): 951–63.
Upadhyay, M. R., W. Chen, J. A. Lenstra, C. R. J. Goderie, D. E. MacHugh, S. D. E. Park, D. A. Magee, et al. 2017. "Genetic Origin, Admixture and Population History of Aurochs (Bos Primigenius) and Primitive European Cattle." *Heredity* 119 (6): 469.
Urlings, Bert. 1995. "Inedible Meat by-Products. Advances in Meat Research, Volume 8." *Livestock Production Science*. https://doi.org/10.1016/0301-6226(95)90021-7.
Vahdati Nasab, Hamed, Abolfazl Aali, Mandan Kazzazi, Mark Pollard, and Thomas Stöllner. 2019. "Reappraisal of the Number of Salt Mummies Identified in Chehrābād Salt Mine, Zanjan, Iran." *Bioarchaeology of the Near East* 13: 23–47.
Vaidyanathan, Vaidyanathan Ganesan, and Balachandran Unni Nair. 2003. "Synthesis, Characterization, and DNA Binding Studies of a Chromium(III) Complex Containing a Tridentate Ligand." *European Journal of Inorganic Chemistry*. https://doi.org/10.1002/ejic.200300170.
Valk, Tom van der, Patrícia Pečnerová, David Díez-Del-Molino, Anders Bergström, Jonas Oppenheimer, Stefanie Hartmann, Georgios Xenikoudakis, et al. 2021. "Million-Year-Old DNA Sheds Light on the Genomic History of Mammoths." *Nature* 591 (7849): 265–69.
Vartanyan, S. L., V. E. Garutt, and A. V. Sher. 1993. "Holocene Dwarf Mammoths from Wrangel Island in the Siberian Arctic." *Nature* 362 (6418): 337–40.
Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, et al. 2001. "The Sequence of the Human Genome." *Science* 291 (5507): 1304–51.
Verdugo, Marta Pereira, Victoria E. Mullin, Amelie Scheu, Valeria Mattiangeli, Kevin G. Daly, Pierpaolo Maisano Delser, Andrew J. Hare, et al. 2019. "Ancient Cattle Genomics, Origins, and Rapid Turnover in the Fertile Crescent." *Science* 365 (6449): 173–76.
Verkaar, Edward L. C., Isaäc J. Nijman, Maurice Beeke, Eline Hanekamp, and Johannes A. Lenstra. 2004. "Maternal and Paternal Lineages in Cross-Breeding Bovine Species. Has Wisent a Hybrid Origin?" *Molecular Biology and Evolution* 21 (7): 1165–70.
Vigne, Jean-Denis. 2011. "The Origins of Animal Domestication and Husbandry: A Major Change in the History of Humanity and the Biosphere." *Comptes Rendus Biologies* 334 (3): 171–81.
———. 2015. "Early Domestication and Farming: What Should We Know or Do for a Better Understanding?" *Anthropozoologica*. https://doi.org/10.5252/az2015n2a5.
Vigne, Jean-Denis, Lionel Gourichon, Daniel Helmer, Louise Martin, and Joris Peters. 2017. "84 The Beginning of Animal Domestication and Husbandry in Southwest Asia." In *Quaternary in the Levant*, edited by Y. Enzel and O. Bar Yosef, 753–60. Cambridge: Cambridge Univ. Press.
Vijayalakshmi, R., M. Kanthimathi, V. Subramanian, and B. U. Nair. 2000. "DNA Cleavage by a Chromium(III) Complex." *Biochemical and Biophysical Research Communications* 271 (3): 731–34.
Vila, Emmanuelle, Philippe Abrahami, Moussab Albesso, Agraw Amane, Camille Bader, Rémi Berthon, Sofiane Bouzid, et al. 2021. "EVOSHEEP: The Makeup of Sheep Breeds in the Ancient Near East." *Antiquity* 95 (379). https://doi.org/10.15184/aqy.2020.247.
Voitkun, Victoria, Anatoly Zhitkovich, and Max Costa. 1998. "Cr(III)-Mediated Crosslinks of Glutathione or Amino Acids to the DNA Phosphate Backbone Are Mutagenic in Human Cells." *Nucleic Acids Research* 26 (8): 2024–30.
Vuissoz, Annick, Michael Worobey, Nancy Odegaard, Michael Bunce, Carlos A. Machado, Niels Lynnerup, Elizabeth E. Peacock, and M. Thomas P. Gilbert. 2007. "The Survival of PCR-Amplifiable DNA in Cow Leather." *Journal of Archaeological Science* 34 (5): 823–29.
Vuure, T. van. 2005. *Retracing the Aurochs: Morphology and Ecology of an Extinct Wild Ox*.
Walsh, Corinne M., Matthew J. Gebert, Manuel Delgado-Baquerizo, Fernando T. Maestre, and Noah Fierer. 2019. "A Global Survey of Mycobacterial Diversity in Soil." *Applied and Environmental Microbiology* 85 (17). https://doi.org/10.1128/AEM.01180-19.
Wang, Chaolong, Xiaowei Zhan, Jennifer Bragg-Gresham, Hyun Min Kang, Dwight

Stambolian, Emily Y. Chew, Kari E. Branham, et al. 2014. “Ancestry Estimation and Control of Population Stratification for Sequence-Based Association Studies.” *Nature Genetics* 46 (4): 409–15.
Wang, Kun, Johannes A. Lenstra, Liang Liu, Quanjun Hu, Tao Ma, Qiang Qiu, and Jianquan Liu. 2018. “Incomplete Lineage Sorting rather than Hybridization Explains the Inconsistent Phylogeny of the Wisent.” *Communications Biology* 1 (October): 169.
Wang, Ming-Shan, Gemma G. R. Murray, Daniel Mann, Pamela Groves, Alisa O. Vershinina, Megan A. Supple, Joshua D. Kapp, et al. 2022. “A Polar Bear Paleogenome Reveals Extensive Ancient Gene Flow from Polar Bears into Brown Bears.” *Nature Ecology & Evolution* 6 (7): 936–44.
Wang, Xiaofang, Hyun Jeong Lim, and Ahjeong Son. 2014. “Characterization of Denaturation and Renaturation of DNA for DNA Hybridization.” *Environmental Health and Toxicology* 29. https://doi.org/10.5620/eht.2014.29.e2014007.
Wang, Xiaoming, Lawrence J. Flynn, and Mikael Fortelius. 2013. *Fossil Mammals of Asia: Neogene Biostratigraphy and Chronology*. Columbia University Press.
Wang, Yujian, Hao Zhang, Lin Zhu, Yulin Xu, Na Liu, Xiaomei Sun, Liping Hu, He Huang, Kai Wei, and Ruiliang Zhu. 2018. “Dynamic Distribution of Gut Microbiota in Goats at Different Ages and Health States.” *Frontiers in Microbiology* 9 (October): 2509.
Wang, Zhaofeng, Xin Shen, Bin Liu, Jianping Su, Takahiro Yonezawa, Yun Yu, Songchang Guo, et al. 2010. “Phylogeographical Analyses of Domestic and Wild Yaks Based on Mitochondrial DNA: New Data and Reappraisal.” *Journal of Biogeography* 37 (12): 2332–44.
Warinner, C., and A. Bouwman. 2015. “Ancient DNA Investigation of the Chehrābād Mummies: Interim Report.” In *The Archaeology of the Salt Miners: Interdisciplinary Research 2010-2014*, edited by Abolfazl Aali and Thomas Stöllner, 90–96. Bochum: Deutsches Bergbau-Museum.
Warinner, Christina, Alexander Herbig, Allison Mann, James A. Fellows Yates, Clemens L. Weiß, Hernán A. Burbano, Ludovic Orlando, and Johannes Krause. 2017. “A Robust Framework for Microbial Archaeology.” *Annual Review of Genomics and Human Genetics* 18 (August): 321–56.
Warinner, Christina, Camilla Speller, Matthew J. Collins, and Cecil M. Lewis Jr. 2015. “Ancient Human Microbiomes.” *Journal of Human Evolution* 79 (February): 125–36.
Węcek, Karolina, Stefanie Hartmann, Johanna L. A. Paijmans, Ulrike Taron, Georgios Xenikoudakis, James A. Cahill, Peter D. Heintzman, et al. 2016. “Complex Admixture Preceded and Followed the Extinction of Wisent in the Wild.” *Molecular Biology and Evolution* 34 (3): 598–612.
Weiß, Clemens L., Michael Dannemann, Kay Prüfer, and Hernán A. Burbano. 2015. “Contesting the Presence of Wheat in the British Isles 8,000 Years Ago by Assessing Ancient DNA Authenticity from Low-Coverage Data,” November. https://doi.org/10.7554/eLife.10005.
Welker, Frido, Elza Duijm, Kristiaan J. van der Gaag, Bas van Geel, Peter de Knijff, Jacqueline van Leeuwen, Dick Mol, et al. 2014. “Analysis of Coprolites from the Extinct Mountain Goat Myotragus Balearicus.” *Quaternary Research* 81 (1): 106–16.
Wilkin, Shevan, Alicia Ventresca Miller, Ricardo Fernandes, Robert Spengler, William T-T Taylor, Dorcas R. Brown, David Reich, et al. 2021. “Dairying Enabled Early Bronze Age Yamnaya Steppe Expansions.” *Nature* 598 (7882): 629–33.
Wingett, Steven W., and Simon Andrews. 2018. “FastQ Screen: A Tool for Multi-Genome Mapping and Quality Control.” *F1000Research*. https://doi.org/10.12688/f1000research.15931.1.
Wojtal, Piotr, and Jarosław Wilczyński. 2015. “Hunters of the Giants: Woolly Mammoth Hunting during the Gravettian in Central Europe.” *Quaternary International: The Journal of the International Union for Quaternary Research* 379 (August): 71–81.
Woodbridge, Jessie, Ralph M. Fyfe, Neil Roberts, Sean Downey, Kevan Edinborough, and Stephen Shennan. 2014. “The Impact of the Neolithic Agricultural Transition in Britain: A

Comparison of Pollen-Based Land-Cover and Archaeological 14C Date-Inferred Population Change." *Journal of Archaeological Science*. https://doi.org/10.1016/j.jas.2012.10.025.
Wood, Derrick E., Jennifer Lu, and Ben Langmead. 2019. "Improved Metagenomic Analysis with Kraken 2." *Genome Biology* 20 (1): 257.
Wood, Jamie R., Andrea Crown, Theresa L. Cole, and Janet M. Wilmshurst. 2016. "Microscopic and Ancient DNA Profiling of Polynesian Dog (kurī) Coprolites from Northern New Zealand." *Journal of Archaeological Science: Reports* 6 (April): 496–505.
Woolhouse, Mark E. J., and Sonya Gowtage-Sequeria. 2005. "Host Range and Emerging and Reemerging Pathogens." *Emerging Infectious Diseases* 11 (12): 1842–47.
Wutz, Anton. 2011. "Gene Silencing in X-Chromosome Inactivation: Advances in Understanding Facultative Heterochromatin Formation." *Nature Reviews. Genetics* 12 (8): 542–53.
Wu, Xiaoyun, Min Chu, Xuezhi Ding, Xian Guo, Hongbo Wang, Pengjia Bao, Chunnian Liang, and Ping Yan. 2018. "Characterization of the Complete Mitochondrial Genome of Kunlun Mountain Type Wild Yak (Bos Mutus)." *Conservation Genetics Resources* 10 (1): 111–13.
Xia, Xiao-Ting, Alessandro Achilli, Johannes A. Lenstra, Bin Tong, Yun Ma, Yong-Zhen Huang, Jian-Lin Han, et al. 2021. "Mitochondrial Genomes from Modern and Ancient Turano-Mongolian Cattle Reveal an Ancient Diversity of Taurine Maternal Lineages in East Asia." *Heredity* 126 (6): 1000–1008.
Xu, Bo, and Ziheng Yang. 2016. "Challenges in Species Tree Estimation Under the Multispecies Coalescent Model." *Genetics* 204 (4): 1353–68.
Xue, Yali, Qiuju Wang, Quan Long, Bee Ling Ng, Harold Swerdlow, John Burton, Carl Skuce, et al. 2009. "Human Y Chromosome Base-Substitution Mutation Rate Measured by Direct Sequencing in a Deep-Rooting Pedigree." *Current Biology*. https://doi.org/10.1016/j.cub.2009.07.032.
Yang, Ziheng, and Bruce Rannala. 2012. "Molecular Phylogenetics: Principles and Practice." *Nature Reviews. Genetics* 13 (5): 303–14.
Zeder, Melinda A. 2006. "Central Questions in the Domestication of Plants and Animals." *Evolutionary Anthropology* 15 (3): 105–17.
———. 2012. "The Domestication of Animals." *Journal of Anthropological Research* 68 (2): 161–90.
———. 2015. "Core Questions in Domestication Research." *Proceedings of the National Academy of Sciences of the United States of America* 112 (11): 3191–98.
Zeuner, Frederick Everard, and Others. 1963. "A History of Domesticated Animals." *A History of Domesticated Animals.* https://www.cabdirect.org/cabdirect/abstract/19630102642.
Zeyland, J., Ł. Wolko, J. Bocianowski, M. Szalata, R. Słomski, A. M. Dzieduszycki, M. Ryba, H. Przystałowska, and D. Lipiński. 2013. "Complete Mitochondrial Genome of Wild Aurochs (Bos Primigenius) Reconstructed from Ancient DNA." *Polish Journal of Veterinary Sciences*. https://doi.org/10.2478/pjvs-2013-0037.
Zhang, Hucai, Johanna L. A. Paijmans, Fengqin Chang, Xiaohong Wu, Guangjie Chen, Chuzhao Lei, Xiujuan Yang, et al. 2013. "Morphological and Genetic Evidence for Early Holocene Cattle Management in Northeastern China." *Nature Communications* 4: 2755.
Zhang, Yan, Qiuying Wu, Liqin Yang, Xuanyu Chen, Chun Wang, Yan Zhang, Yutian Zeng, et al. 2019. "Characterization of the Complete Mitochondrial Genome Sequence of Golden Wild Yak and Revealed Its Phylogenetic Relationship with 9 Yak Subspecies." *Mitochondrial DNA Part B*. https://doi.org/10.1080/23802359.2019.1568215.
Zhao, Yong-Xin, Ji Yang, Feng-Hua Lv, Xiao-Ju Hu, Xing-Long Xie, Min Zhang, Wen-Rong Li, et al. 2017. "Genomic Reconstruction of the History of Native Sheep Reveals the Peopling Patterns of Nomads and the Expansion of Early Pastoralism in East Asia." *Molecular Biology and Evolution* 34 (9): 2380–95.
Zhitkovich, A. 2005. "Importance of Chromium-DNA Adducts in Mutagenicity and Toxicity of chromium(VI)." *Chemical Research in Toxicology* 18 (1).

https://doi.org/10.1021/tx049774+.
Zhitkovich, A., Y. Song, G. Quievryn, and V. Voitkun. 2001. “Non-Oxidative Mechanisms Are Responsible for the Induction of Mutagenesis by Reduction of Cr(VI) with Cysteine: Role of Ternary DNA Adducts in Cr(III)-Dependent Mutagenesis.” *Biochemistry* 40 (2): 549–60.
Zimin, Aleksey V., Arthur L. Delcher, Liliana Florea, David R. Kelley, Michael C. Schatz, Daniela Puiu, Finnian Hanrahan, et al. 2009. “A Whole-Genome Assembly of the Domestic Cow, Bos Taurus.” *Genome Biology* 10 (4): R42.

12. Appendix

Appendix Table 3.1 - Alignment statistics of screening samples. Laboratory work is defined as any handling of the sample from sample drilling to sequencing of the screening data. AH = Andrew Hare, CR = Conor Rossi, ER = Erika Rosengren, KD = Kevin Daly, MS = Mikkel Sinding, MV = Marta Verdugo, VM = Valeria Mattiangeli, VEM = Victoria Mullin.

Internal Lab ID	Screening	Trimmed Reads	Aligned mq30	Endogenous	median length	Laboratory Work
CR001	MiSeq	236691	6000	2.53	55	VM/CR
CR002	MiSeq	1059162	437206	41.28	53	VM/CR
CR003	MiSeq	276685	88205	31.88	56	VM/CR
CR004	MiSeq	283071	87617	30.95	56	VM/CR
CR005	MiSeq	396768	18622	4.69	53	VM/CR
CR006	MiSeq	305350	132213	43.30	60	VM/CR
CR007	MiSeq	363467	67004	18.43	61	VM/CR
CR008	MiSeq	201909	28	0.01	N/A	VM/CR
CR009	MiSeq	350924	13460	3.84	54	VM/CR
CR011	MiSeq	201274	34084	16.93	55	VM/CR
CR012	MiSeq	279577	23773	8.50	53	VM/CR
CR013	MiSeq	296797	82978	27.96	54	VM/CR
CR014	MiSeq	568559	322009	56.64	58	VM/CR
CR015	MiSeq	447289	173403	38.77	54	VM/CR
CR016	MiSeq	290935	32789	11.27	55	VM/CR
CR017	MiSeq	274157	77212	28.16	42	VM/CR
CR018	MiSeq	152033	82	0.05	N/A	CR
CR019	MiSeq	159951	956	0.60	43	CR
CR020	MiSeq	181070	43	0.02	N/A	CR
CR021	MiSeq	167169	79	0.05	N/A	CR
CR022	MiSeq	936769	539597	57.60	60	CR
CR023	MiSeq	624432	345907	55.40	61	CR
CR024	MiSeq	296670	154101	51.94	61	CR
CR025	MiSeq	694717	339318	48.84	59	CR
CR026	MiSeq	853733	318478	37.30	62	CR
CR027	MiSeq	190195	949	0.50	61	CR
CR028	MiSeq	1556041	828172	53.22	58	CR
CR029	MiSeq	2227578	1156513	51.92	60	CR

CR030	MiSeq	739754	27970	3.78	49	CR
CR031	MiSeq	1798316	38459	2.14	56	CR
CR032	MiSeq	201548	41400	20.54	56	CR
CR033	MiSeq	239982	36523	15.22	50	CR
CR034	MiSeq	173986	66185	38.04	58	CR
CR035	MiSeq	377827	209901	55.55	58	CR
CR036	MiSeq	400009	210536	52.63	59	CR
CR037	MiSeq	229511	85426	37.22	54	CR
CR038	MiSeq	230813	107025	46.37	57	CR
CR039	MiSeq	356961	192234	53.85	59	CR
CR040	MiSeq	659286	159274	24.16	54	CR
CR041	MiSeq	410606	220583	53.72	59	CR
CR042	MiSeq	210119	96088	45.73	56	CR
CR043	MiSeq	251902	97773	38.81	55	CR
CR044	MiSeq	171861	62473	36.35	55	CR
CR045	MiSeq	245382	134640	54.87	59	CR
CR046	MiSeq	280390	123074	43.89	59	CR
CR047	MiSeq	131345	61	0.05	N/A	CR
CR048	MiSeq	178896	61	0.03	N/A	CR
CR049	MiSeq	777391	403137	51.86	55	CR
CR050	MiSeq	177887	65900	37.05	to-do	CR
CR051	MiSeq	416	15	3.61	N/A	CR
CR052	MiSeq	324787	76817	23.65	57	CR
CR053	MiSeq	170968	2461	1.44	52	CR
CR054	MiSeq	707423	9642	1.36	49	CR
CR055	MiSeq	246688	18187	7.37	52	CR
CR056	MiSeq	98092	427	0.44	61	CR
CR057	MiSeq	217266	818	0.38	58	CR
CR058	MiSeq	270373	1330	0.49	56	CR
CR059	MiSeq	fail	fail	fail	N/A	CR
CR060	MiSeq	77029	2027	2.63	39	CR
CR061	MiSeq	94620	2027	2.14	45	CR
CR062	MiSeq	282038	38779	13.75	44	CR

CR063	MiSeq	125543	340	0.27	45	CR
CR064	MiSeq	102317	19050	18.62	50	CR
CR065	MiSeq	241086	106	0.04	N/A	CR
CR066	MiSeq	335273	235	0.07	40	CR
CR067	MiSeq	294173	276	0.09	61	CR
CR068	MiSeq	223116	93	0.04	N/A	CR
CR069	MiSeq	206870	173	0.08	47	CR
CR070	MiSeq	201189	224	0.11	58	CR
CR071	MiSeq	299171	1693	0.57	52	CR
CR072	MiSeq	138457	24	0.02	N/A	CR
CR073	MiSeq	386630	15578	4.03	42	CR
CR074	MiSeq	1680	81	4.82	N/A	CR
CR075	MiSeq	980	10	1.02	N/A	CR
CR076	MiSeq	783420	9219	1.18	60	CR
CR077	MiSeq	478119	6585	1.38	41	CR
CR078	MiSeq	97460	561	0.58	41	CR
CR079	MiSeq	196947	5629	2.86	36	CR
CR080	MiSeq	1313	11	0.84	N/A	CR
CR081	MiSeq	195607	154	0.08	60	CR
CR082	MiSeq	2794	10	0.36	N/A	CR
CR083	MiSeq	8115	35	0.43	N/A	CR
CR084	MiSeq	77127	965	1.25	52	CR
CR085	MiSeq	14915	39	0.26	N/A	CR
CR086	MiSeq	17263	154	0.89	30	CR
CR087	MiSeq	34266	168	0.49	36	CR
CR088	MiSeq	1849	11	0.59	N/A	CR
CR089	MiSeq	10292	242	2.35	39	CR
CR090	MiSeq	3632	49	1.35	N/A	CR
CR091	MiSeq	6922	27	0.39	N/A	CR
CR092	MiSeq	92417	6184	6.69	52	CR
CR093	MiSeq	6690	69	1.03	N/A	CR
CR094	MiSeq	2362	17	0.72	N/A	CR
CR095	MiSeq	1719902	7545	0.44	61	CR

CR096	MiSeq	223157	53393	23.93	61	CR
CR097	MiSeq	198771	578	0.29	61	CR
CR098	MiSeq	172461	49	0.03	N/A	CR
CR099	MiSeq	272205	1170	0.43	61	CR
CR100	MiSeq	158047	15742	9.96	56	CR
CR101	MiSeq	516070	45493	8.82	55	CR
CR102	MiSeq	751348	83363	11.10	57	CR
CR103	MiSeq	464758	5196	1.12	53	CR
CR104	MiSeq	204990	58069	28.33	49	CR
CR105	MiSeq	106247	6636	6.25	50	CR
CR106	MiSeq	14815	1323	8.93	51	CR
CR107	MiSeq	795869	1694	0.21	52	CR
CR108	MiSeq	705731	1216	0.17	50	CR
CR109	MiSeq	875192	2082	0.24	54	CR
CR110	MiSeq	364465	114826	31.51	47	CR
CR111	MiSeq	622424	918	0.15	44	CR
CR112	MiSeq	95091	518	0.54	43	CR
CR113	MiSeq	140679	422	0.30	46	CR
CR114	MiSeq	746314	93014	12.46	44	CR
CR115	MiSeq	251155	1231	0.49	49	CR
CR116	MiSeq	815428	429459	52.67	51	CR
CR117	NovaSeq	1786421	925453	51.80	44	CR
CR118	NovaSeq	1890627	932570	49.33	44	CR
CR119	NovaSeq	1164956	124402	10.68	36	CR
CR120	NovaSeq	13125595	7203124	54.88	49	CR
CR121	NovaSeq	723357	129460	17.90	37	CR
CR122	NovaSeq	438404	12871	2.94	36	CR
CR123	NovaSeq	578249	6622	1.15	39	CR
CR124	NovaSeq	600551	374	0.06	37	CR
CR125	NovaSeq	592779	256981	43.35	43	CR
CR126	NovaSeq	615214	276548	44.95	41	CR
CR127	NovaSeq	575000	151	0.03	N/A	CR
CR128	NovaSeq	450250	1066	0.24	35	CR

CR129	NovaSeq	2362551	3682	0.16	51	CR
CR130	NovaSeq	751164	13913	1.85	43	ER/CR
CR131	NovaSeq	913316	106	0.01	N/A	ER/CR
CR132	NovaSeq	761154	14281	1.88	49	ER/CR
CR133	NovaSeq	528945	5545	1.05	43	ER/CR
CR134	NovaSeq	425995	13613	3.20	45	ER/CR
CR135	NovaSeq	1007472	22134	2.20	50	ER/CR
CR136	NovaSeq	716873	56	0.01	N/A	ER/CR
CR137	NovaSeq	407287	118	0.03	N/A	ER/CR
CR138	NovaSeq	642007	50	0.01	N/A	ER/CR
CR139	NovaSeq	922280	6051	0.66	N/A	ER/CR
CR140	NovaSeq	1097824	1366	0.12	49	ER/CR
CR141	NovaSeq	450830	174	0.04	N/A	ER/CR
CR142	NovaSeq	486785	116	0.02	N/A	ER/CR
CR143	NovaSeq	387617	115	0.03	N/A	ER/CR
CR144	NovaSeq	617506	151	0.02	N/A	ER/CR
CR145	NovaSeq	2425239	159	0.01	N/A	ER/CR
CR146	NovaSeq	736812	377	0.05	N/A	ER/CR
CR147	NovaSeq	589540	4241	0.72	39	ER/CR
CR148	NovaSeq	393736	8125	2.06	36	ER/CR
CR149	NovaSeq	649749	4120	0.63	38	ER/CR
CR150	NovaSeq	2399279	45254	1.89	40	ER/CR
CR151	NovaSeq	571162	899	0.16	37	ER/CR
CR152	NovaSeq	715271	3094	0.43	35	ER/CR
CR153	NovaSeq	599385	1610	0.27	39	ER/CR
CR154	NovaSeq	642105	1297	0.20	38	ER/CR
CR155	NovaSeq	634886	2691	0.42	41	ER/CR
CR156	NovaSeq	2772321	304	0.01	N/A	ER/CR
CR157	NovaSeq	815487	12966	1.59	41	ER/CR
CR158	NovaSeq	4242559	322	0.01	N/A	ER/CR
CR159	NovaSeq	3177944	1068	0.03	51	ER/CR
CR160	NovaSeq	2074223	141	0.01	N/A	ER/CR
CR161	NovaSeq	550691	506	0.09	41	ER/CR

CR162	NovaSeq	646936	250	0.04	N/A	ER/CR
CR163	NovaSeq	1948579	3667	0.19	40	CR
CR164	NovaSeq	1847236	333	0.02	N/A	CR
CR165	NovaSeq	1400670	684100	48.84	42	CR
CR166	NovaSeq	1014108	85758	8.46	38	CR
CR167	NovaSeq	110519	33415	30.23	37	CR
CR168	NovaSeq	1852	32	1.73	N/A	CR
CR169	NovaSeq	889817	302773	34.03	39	CR
CR170	NovaSeq	158458	33439	21.10	45	CR
CR171	NovaSeq	772741	478161	61.88	45	CR
CR172	NovaSeq	240694	53803	22.35	46	CR
CR173	NovaSeq	235345	1355	0.58	37	CR
CR174	NovaSeq	282884	45779	16.18	45	CR
CR175	NovaSeq	56928	24409	42.88	43	CR
CR176	NovaSeq	253431	98684	38.94	51	CR
CR177	NovaSeq	182016	46504	25.55	53	CR
CR178	MiSeq	117125	7158	6.11	38	CR
CR179	MiSeq	116315	4744	4.08	38	CR
CR180	MiSeq	155823	71596	45.95	36	CR
CR181	MiSeq	124341	35249	28.35	36	CR
CR182	MiSeq	102412	59339	57.94	40	CR
CR183	NovaSeq	27243583	12071789	44.31	38	CR
CR184	NovaSeq	30512585	11249464	36.87	37	CR
CR185	NovaSeq	20712315	8203644	39.61	36	CR
CR186	NovaSeq	17967090	4906172	27.31	35	CR
CR187	NovaSeq	10661873	1986470	18.63	34	CR
CR188	NovaSeq	90408032	6371017	7.05	45	MS/CR
CR189	NovaSeq	4288998	28545	0.67	56	MS/CR
CR190	NovaSeq	7230146	91384	1.26	40	MS/CR
CR191	NovaSeq	5042351	149515	2.97	56	MS/CR
CR192	NovaSeq	4298151	230593	5.36	50	MS/CR
CR193	NovaSeq	11638901	697941	6.00	34	MS/CR
CR194	NovaSeq	fail	fail	fail	N/A	MS/CR

ER001	HiSeq	20167452	1885	0.01	N/A	ER
ER002	HiSeq	2897931	1102	0.04	N/A	ER
ER003	HiSeq	9007590	369353	4.10	49	ER
ER004	HiSeq	8908096	32221	0.36	50	ER
ER005	HiSeq	22887218	389614	1.70	46	ER
ER007	HiSeq	9109433	521931	5.73	36	ER
ER008	HiSeq	942454	52514	5.57	58	ER
ER009	HiSeq	663954	138	0.02	N/A	ER
ER010	HiSeq	282399	7726	2.74	57	ER
ER011	HiSeq	4817065	1024	0.02	N/A	ER
ER012	HiSeq	1605194	656	0.04	52	ER
ER013	HiSeq	737224	22	0.00	N/A	ER
ER014	HiSeq	330574	9	0.00	N/A	ER
ER015	HiSeq	fail	fail	fail	N/A	ER
ER016	HiSeq	664293	338	0.05	55	ER
ER017	HiSeq	4744461	568	0.01	N/A	ER
ER018	HiSeq	360005	29	0.01	N/A	ER
ER019	HiSeq	216385	15	0.01	N/A	ER
ER020	HiSeq	276101	42	0.02	N/A	ER
ER021	HiSeq	230650	7	0.00	N/A	ER
ER022	HiSeq	1075151	476	0.04	N/A	ER
ER023	HiSeq	4970819	74098	1.49	41	ER
ER024	HiSeq	6719825	42775	0.64	40	ER
ER025	HiSeq	1058659	1064	0.10	55	ER
ER026	HiSeq	385354	6193	1.61	53	ER
ER027	HiSeq	631440	2560	0.41	55	ER
ER028	HiSeq	379527	936	0.25	55	ER
MV018	MiSeq	132403	2657	2.01	48	MV/CR
MV027	MiSeq	127552	11792	9.24	43	MV/CR
MV045	MiSeq	124629	12757	10.24	43	MV/CR
MV055	MiSeq	109717	6223	5.67	46	MV/CR
MV193/Kalisz1	MiSeq	132882	35034	26.36	39	CR
MV194-Tula	NovaSeq	1280958	582707	45.49	39	CR

MV196-Kalisz3	NovaSeq	1255009	434632	34.63	44	CR
MV239-Yer1	NovaSeq	1872201	504794	26.96	41	CR
MV240-Kok1	NovaSeq	1502855	688043	45.78	46	CR
NHMD 231145 (K1)	MiSeq	109717	6223	5.67	48	MS/CR
NHMD 231146 (K2)	MiSeq	128421	32288	25.14	54	MS/CR
4169	MiSeq	127522	514	0.40	44	MS/CR
4476	MiSeq	113184	29885	26.40	46	MS/CR
4477	MiSeq	64144	32849	51.21	54	MS/CR
4478	MiSeq	157533	17118	10.87	46	MS/CR
AH067	MiSeq	41700	8525	20.44	51	AH/CR
AH068	MiSeq	11227	3818	34.01	55	AH/CR
AH071-1	MiSeq	6047	2335	38.61	51	AH/CR
AH071-2	MiSeq	29445	14815	50.31	60	AH/CR
AH074	MiSeq	3357	355	10.57	40	AH/CR
AH075	MiSeq	1530	8	0.52	N/A	AH/CR
AH076-1	MiSeq	36166	19	0.05	N/A	AH/CR
AH076-2	MiSeq	22040	11	0.05	N/A	AH/CR
AH099	MiSeq	4081	598	14.65	50	AH/CR
AH100	MiSeq	7675	15	0.20	N/A	AH/CR
AH101	MiSeq	11309	4	0.04	N/A	AH/CR
AH102-1	MiSeq	65801	38347	58.28	51	AH/CR
AH103	MiSeq	33859	7521	22.21	45	AH/CR
AH106	MiSeq	24283	1354	5.58	47	AH/CR
AH107	MiSeq	6002	28	0.47	N/A	AH/CR
AH108	MiSeq	13000	763	5.87	43	AH/CR
AH109	MiSeq	1743	417	23.92	58	AH/CR
AH110	MiSeq	46095	11058	23.99	51	AH/CR
AH111	MiSeq	33174	20596	62.08	57	AH/CR
AH112	MiSeq	24423	895	3.66	44	AH/CR
AH113	MiSeq	30890	63	0.20	N/A	AH/CR
AH114-1	MiSeq	30744	3026	9.84	57	AH/CR
CN 1147	MiSeq	146946	18949	12.90	51	MS/CR
CN 1264	MiSeq	86350	115	0.13	51	MS/CR

CN 1576	MiSeq	103082	46363	44.98	44	MS/CR
CN 1577	MiSeq	125135	53748	42.95	48	MS/CR
CN 1578	MiSeq	105462	2723	2.58	39	MS/CR
CN 1587	MiSeq	109017	28109	25.78	49	MS/CR
CN 1588	MiSeq	213253	34545	16.20	44	MS/CR
CN 1589	MiSeq	72969	18	0.02	N/A	MS/CR
CN 1590	MiSeq	107184	11090	10.35	50	MS/CR
CN 1591	MiSeq	37768	3299	8.73	54	MS/CR
CN 1592	MiSeq	113548	28614	25.20	51	MS/CR
CN 3353	MiSeq	104548	11715	11.21	52	MS/CR
CN 3402	MiSeq	58001	13754	23.71	54	MS/CR
CN 526	MiSeq	88416	1627	1.84	42	MS/CR
CN 527	MiSeq	90328	48983	54.23	50	MS/CR
CN 528	MiSeq	113196	20163	17.81	47	MS/CR
CN 529	MiSeq	134843	49313	36.57	40	MS/CR
CN 643	MiSeq	111791	31293	27.99	45	MS/CR
CN 867	MiSeq	87526	2139	2.44	49	MS/CR
CN 903	MiSeq	32784	2864	8.74	50	MS/CR
CN 904	MiSeq	107274	9593	8.94	53	MS/CR
CN 906	MiSeq	190989	33756	17.67	52	MS/CR
CN 907	MiSeq	250590	22830	9.11	43	MS/CR
ROS001	NovaSeq	44480823	97143	0.22	48	KD/CR
ROS002	NovaSeq	45257518	711663	1.57	48	KD/CR
WB001	NovaSeq	9638661	5747	0.06	43	CR
WB002	NovaSeq	34308788	4246	0.01	44	CR
WB004	NovaSeq	11683047	104767	0.90	43	CR

Appendix Table 4.1 - The cytosine deamination probability in single strand context (δS) of published ancient tissues.

Sample ID	Tissue	Age	Nuc δS mean	Study
MUM2	Mummified Tissue	1600	0.012	*This study*
NE6	Bone	5125	0.160	(Kistler et al. 2017)
BR2	Bone	1650	0.170	(Kistler et al. 2017)
KO1	Bone	5715	0.180	(Kistler et al. 2017)

Batagai	Bone	5154	0.202	(Kistler et al. 2017)
BR1	Bone	2085	0.210	(Kistler et al. 2017)
CO1	Bone	2800	0.210	(Kistler et al. 2017)
NE2	Bone	5175	0.210	(Kistler et al. 2017)
NE7	Bone	4425	0.220	(Kistler et al. 2017)
NE5	Bone	5100	0.230	(Kistler et al. 2017)
6DRIF-18	Tooth	1800	0.240	(Kistler et al. 2017)
NE4	Bone	5160	0.240	(Kistler et al. 2017)
KO2	Bone	5640	0.250	(Kistler et al. 2017)
NE1	Bone	5190	0.250	(Kistler et al. 2017)
M63	Tooth	2125	0.256	(Kistler et al. 2017)
6DRIF-21	Tooth	1800	0.286	(Kistler et al. 2017)
RM127	Bone	4021	0.290	(Kistler et al. 2017)
RISE504	Bone	1210	0.290	(Kistler et al. 2017)
6DRIF-23	Bone	1800	0.294	(Kistler et al. 2017)
6DRIF-3	Bone	1800	0.309	(Kistler et al. 2017)
6DRIF-22	Tooth	1800	0.319	(Kistler et al. 2017)
Khor1	Bone		0.334	*This study*
RSK2	Bone	3651	0.334	(Kistler et al. 2017)
RISE174	Bone	1496	0.340	(Kistler et al. 2017)
RISE502	Bone	3416	0.340	(Kistler et al. 2017)
RISE505	Bone	3701	0.340	(Kistler et al. 2017)
3DRIF-26	Tooth	1800	0.349	(Kistler et al. 2017)
RISE495	Bone	1150	0.350	(Kistler et al. 2017)
RISE500	Bone	1600	0.350	(Kistler et al. 2017)
RISE523	Bone	3513	0.350	(Kistler et al. 2017)
RISE555	Bone	4692	0.350	(Kistler et al. 2017)
BA64	Bone	5346	0.363	(Kistler et al. 2017)
S35	Tooth	1500	0.366	(Kistler et al. 2017)
RISE480	Bone	1750	0.370	(Kistler et al. 2017)
RISE479	Bone	1750	0.380	(Kistler et al. 2017)
KD177-2	Bone	2440	0.382	*(Daly et al. 2018)*
Khor1	Bone		0.382	*This study*

M1489	Bone	2100	0.389	(Kistler et al. 2017)
NE3	Bone	5110	0.390	(Kistler et al. 2017)
RISE373	Bone	3806	0.390	(Kistler et al. 2017)
RISE386	Bone	4186.5	0.390	(Kistler et al. 2017)
NO3423	Bone	1416	0.397	(Kistler et al. 2017)
RISE497	Bone	1150	0.400	(Kistler et al. 2017)
3DRIF-16	Bone	1800	0.400	(Kistler et al. 2017)
RISE492	Bone	2317.5	0.410	(Kistler et al. 2017)
RISE511	Bone	4809	0.410	(Kistler et al. 2017)
RISE512	Bone	3387	0.410	(Kistler et al. 2017)
KD178-2	Bone	4550	0.411	*(Daly et al. 2018)*
RSK1	Bone	3898	0.415	(Kistler et al. 2017)
KD179-2	Bone	4150	0.417	*(Daly et al. 2018)*
RISE175	Bone	3278.5	0.420	(Kistler et al. 2017)
RISE567	Bone	2300	0.420	(Kistler et al. 2017)
RISE98	Bone	4164	0.420	(Kistler et al. 2017)
KD001	Bone	4142	0.424	*(Daly et al. 2018)*
RISE1	Bone	4736.5	0.430	(Kistler et al. 2017)
RISE509	Bone	4797	0.430	(Kistler et al. 2017)
RISE524	Bone	1000	0.430	(Kistler et al. 2017)
ATP17	Bone	4889	0.440	(Kistler et al. 2017)
RISE525	Bone	1000	0.440	(Kistler et al. 2017)
CS7675	Tooth	7285	0.445	(Kistler et al. 2017)
RISE395	Bone	3873	0.450	(Kistler et al. 2017)
RISE569	Bone	2300	0.450	(Kistler et al. 2017)
CB14	Tooth	7400	0.453	(Kistler et al. 2017)
H3C6	Bone	7350	0.454	(Kistler et al. 2017)
Mebrak 344	Tooth	2125	0.457	(Kistler et al. 2017)
F19	Bone	7380	0.458	(Kistler et al. 2017)
ATP9	Bone	3634	0.460	(Kistler et al. 2017)
RISE276	Bone	2685.5	0.460	(Kistler et al. 2017)
KD157a	Bone	3960	0.466	*(Daly et al. 2018)*
RISE423	Bone	3321.5	0.470	(Kistler et al. 2017)

Samdzong 40	Tooth	1500	0.474	(Kistler et al. 2017)
G21	Bone	7390	0.475	(Kistler et al. 2017)
RISE371	Bone	4053.5	0.480	(Kistler et al. 2017)
RISE499	Bone	1150	0.480	(Kistler et al. 2017)
Mebrak 240	Tooth	2125	0.485	(Kistler et al. 2017)
RISE00	Bone	2550	0.490	(Kistler et al. 2017)
RISE240	Bone	4771	0.490	(Kistler et al. 2017)
Chokopani 1	Tooth	2775	0.494	(Kistler et al. 2017)
RISE392	Bone	4026	0.500	(Kistler et al. 2017)
RISE494	Bone	3357	0.500	(Kistler et al. 2017)
S10	Tooth	1500	0.501	(Kistler et al. 2017)
LaBrana	Bone	7816	0.510	(Kistler et al. 2017)
Matojo	Bone	4894.5	0.510	(Kistler et al. 2017)
RISE553	Bone	2885.5	0.510	(Kistler et al. 2017)
RISE568	Bone	2300	0.510	(Kistler et al. 2017)
AH005-3-31	Bone	8000	0.525	*(Daly et al. 2018)*
S41	Tooth	1500	0.539	(Kistler et al. 2017)
RISE109	Bone	3878	0.540	(Kistler et al. 2017)
RISE374	Bone	3757.5	0.550	(Kistler et al. 2017)
RISE496	Bone	3352.5	0.550	(Kistler et al. 2017)
RISE503	Bone	3634	0.550	(Kistler et al. 2017)
RISE71	Bone	4124.5	0.550	(Kistler et al. 2017)
AH006-4-34	Bone	8000	0.553	*(Daly et al. 2018)*
RISE247	Bone	3693.5	0.560	(Kistler et al. 2017)
RISE484	Bone	1750	0.560	(Kistler et al. 2017)
RISE94	Bone	4561.5	0.560	(Kistler et al. 2017)
AH005-1-5	Bone	8002	0.568	*(Daly et al. 2018)*
ATP16	Bone	5038.5	0.570	(Kistler et al. 2017)
RISE42	Bone	4096.5	0.570	(Kistler et al. 2017)
RISE493	Bone	3494	0.570	(Kistler et al. 2017)
RISE510	Bone	4674.5	0.570	(Kistler et al. 2017)
ATP20	Bone	4119.5	0.580	(Kistler et al. 2017)
RISE407	Bone	3020	0.580	(Kistler et al. 2017)

RISE483	Bone	1750	0.580	(Kistler et al. 2017)
ATP7	Bone	5089.5	0.590	(Kistler et al. 2017)
RISE446	Bone	4662	0.590	(Kistler et al. 2017)
RISE210	Bone	3377	0.600	(Kistler et al. 2017)
RISE408	Bone	3124	0.600	(Kistler et al. 2017)
RISE547	Bone	4775.5	0.600	(Kistler et al. 2017)
RISE554	Bone	2939.5	0.600	(Kistler et al. 2017)
KD111-2	Bone	9152	0.601	*(Daly et al. 2018)*
KD236	Bone	6800	0.608	*(Daly et al. 2018)*
RISE394	Bone	3866.5	0.610	(Kistler et al. 2017)
RISE508	Bone	5148	0.610	(Kistler et al. 2017)
RISE516	Bone	4133.5	0.620	(Kistler et al. 2017)
RISE546	Bone	2700	0.620	(Kistler et al. 2017)
NGDG_2014_L2_in11	Bone	4800	0.622	(Kistler et al. 2017)
KD113-2	Bone	8109	0.630	*(Daly et al. 2018)*
RISE436	Bone	4739	0.630	(Kistler et al. 2017)
RISE507	Bone	5137.5	0.630	(Kistler et al. 2017)
KD112-2	Bone	9152	0.631	*(Daly et al. 2018)*
ATP3	Bone	5389	0.640	(Kistler et al. 2017)
MotaCave	Bone	4469.5	0.640	(Kistler et al. 2017)
RISE391	Bone	4018.5	0.650	(Kistler et al. 2017)
KD238	Bone	6800	0.653	*(Daly et al. 2018)*
KD184-1	Bone	8500	0.658	*(Daly et al. 2018)*
RISE150	Bone	3804	0.660	(Kistler et al. 2017)
RISE21	Bone	3368.5	0.660	(Kistler et al. 2017)
RISE154	Bone	3860	0.670	(Kistler et al. 2017)
RISE515	Bone	4257.5	0.670	(Kistler et al. 2017)
KD089	Bone	3598	0.691	*(Daly et al. 2018)*
KD235	Bone	6800	0.692	*(Daly et al. 2018)*
RISE179	Bone	3908	0.710	(Kistler et al. 2017)
KD040	Bone	6248	0.719	*(Daly et al. 2018)*
RISE416	Bone	3559	0.720	(Kistler et al. 2017)
RISE548	Bone	2700	0.720	(Kistler et al. 2017)

RISE566	Bone	2300	0.720	(Kistler et al. 2017)
CB13	Tooth	7400	0.726	(Kistler et al. 2017)
KD042	Bone	7994	0.726	*(Daly et al. 2018)*
RISE562	Bone	2300	0.730	(Kistler et al. 2017)
RISE97	Bone	3970	0.730	(Kistler et al. 2017)
RISE254	Bone	4033.5	0.740	(Kistler et al. 2017)
RISE47	Bone	3426.5	0.740	(Kistler et al. 2017)
RISE564	Bone	2300	0.740	(Kistler et al. 2017)
KD242	Bone	8150	0.759	*(Daly et al. 2018)*
ATP2	Bone	4738.5	0.760	(Kistler et al. 2017)
RISE577	Bone	2050	0.760	(Kistler et al. 2017)
AH004-4	Bone	8000	0.776	*(Daly et al. 2018)*
RISE559	Bone	2300	0.780	(Kistler et al. 2017)
RISE563	Bone	2300	0.790	(Kistler et al. 2017)
AH005-2-6	Bone	8000	0.826	*(Daly et al. 2018)*
RISE486	Bone	3968.5	0.830	(Kistler et al. 2017)
RISE61	Bone	4686.5	0.830	(Kistler et al. 2017)
KD240-2	Bone	6800	0.833	*(Daly et al. 2018)*
KD055	Bone	8000	0.835	*(Daly et al. 2018)*
RISE145	Bone	4088	0.840	(Kistler et al. 2017)
RISE413	Bone	3817	0.840	(Kistler et al. 2017)
RISE412	Bone	3084	0.850	(Kistler et al. 2017)
RISE434	Bone	4770	0.860	(Kistler et al. 2017)
KD233	Bone	6500	0.861	*(Daly et al. 2018)*
RISE349	Bone	3924	0.870	(Kistler et al. 2017)
RISE550	Bone	4999.5	0.880	(Kistler et al. 2017)
RISE560	Bone	2300	0.880	(Kistler et al. 2017)
RISE471	Bone	2550	0.900	(Kistler et al. 2017)
RISE489	Bone	4758	0.900	(Kistler et al. 2017)
KD041	Bone	8229	0.902	*(Daly et al. 2018)*
RISE435	Bone	4695.5	0.930	(Kistler et al. 2017)
Klei10	Bone	6128	0.932	(Kistler et al. 2017)
Bar8	Bone	8137	0.937	(Kistler et al. 2017)

RISE396	Bone	3079.5	0.950	(Kistler et al. 2017)
Kennewick	Bone	8770	0.960	(Kistler et al. 2017)
RISE397	Bone	2966.5	0.960	(Kistler et al. 2017)
KD259	Bone	8230	0.991	*(Daly et al. 2018)*
KD197	Bone	6800	0.996	*(Daly et al. 2018)*
Bar31	Bone	8344	1.000	(Kistler et al. 2017)
Pal7	Bone	6407	1.000	(Kistler et al. 2017)
Rev5	Bone	8367	1.000	(Kistler et al. 2017)
RISE139	Bone	4044	1.000	(Kistler et al. 2017)
RISE207	Bone	3412.5	1.000	(Kistler et al. 2017)
RISE431	Bone	4182	1.000	(Kistler et al. 2017)
RISE487	Bone	5310	1.000	(Kistler et al. 2017)
RISE552	Bone	4511	1.000	(Kistler et al. 2017)
RISE586	Bone	2050	1.000	(Kistler et al. 2017)
RISE598	Bone	2711.5	1.000	(Kistler et al. 2017)

Appendix Table 4.2 - Median fragment lengths of published ancient ovicaprid genomes.

Sample ID	Tissue	Median length	Study
MUM2	Mummified Tissue	107	*This study*
Geor2	Bone	64	(Daly et al. 2018)
Monjukli2	Bone	59	(Daly et al. 2018)
Semnan13	Bone	59	(Daly et al. 2018)
Chalow1	Bone	57	(Daly et al. 2018)
Kazbegi1	Bone	57	(Daly et al. 2018)
Darre2	Bone	56	(Daly et al. 2018)
Potterne1	Bone	56	(Daly et al. 2018)
Azer2	Bone	56	*This study*
Semnan3	Bone	55	(Daly et al. 2018)
Monjukli8	Bone	54	(Daly et al. 2018)
Azer4	Bone	53	(Daly et al. 2018)
Bulak1	Bone	53	(Daly et al. 2018)
Hovk1	Bone	53	(Daly et al. 2018)
Monjukli1	Bone	53	(Daly et al. 2018)
Acem2	Bone	52	(Daly et al. 2018)

Fars4	Bone	52	(Daly et al. 2018)
Monjukli4	Bone	52	(Daly et al. 2018)
Semnan17	Bone	52	(Daly et al. 2018)
Semnan9	Bone	52	(Daly et al. 2018)
Tac3	Bone	52	(Daly et al. 2018)
Khor1	Bone	52	*This study*
Azer6	Bone	51	(Daly et al. 2018)
Qazvin1	Bone	51	(Daly et al. 2018)
Safi2	Bone	50	(Daly et al. 2018)
Semnan1-2	Bone	50	(Daly et al. 2018)
Yoqneam2	Bone	50	(Daly et al. 2018)
Azer3-5	Bone	49	(Daly et al. 2018)
Kov57	Bone	48	(Daly et al. 2018)
Lur12	Bone	48	(Daly et al. 2018)
Acem1	Bone	47	(Daly et al. 2018)
Bulak2	Bone	46	(Daly et al. 2018)
Bulak5	Bone	46	(Daly et al. 2018)
Direkli1-2	Bone	46	(Daly et al. 2018)
Direkli4	Bone	46	(Daly et al. 2018)
Shiqmim9	Bone	46	(Daly et al. 2018)
Ainghazal2	Bone	45	(Daly et al. 2018)
Darre1	Bone	45	(Daly et al. 2018)
Direkli5	Bone	45	(Daly et al. 2018)
Direkli6	Bone	45	(Daly et al. 2018)
Fars2-5	Bone	45	(Daly et al. 2018)
Gilat8	Bone	45	(Daly et al. 2018)
Monjukli6	Bone	45	(Daly et al. 2018)
Shiqmim1	Bone	45	(Daly et al. 2018)
Ainghazal3	Bone	44	(Daly et al. 2018)
Ainghazal4	Bone	44	(Daly et al. 2018)
Blagotin1	Bone	44	(Daly et al. 2018)
Ghosh5	Bone	44	(Daly et al. 2018)
Ainghazal1	Bone	43	(Daly et al. 2018)

Blagotin2	Bone	43	(Daly et al. 2018)
Blagotin3	Bone	43	(Daly et al. 2018)
Fars1	Bone	43	(Daly et al. 2018)
Gilat10	Bone	43	(Daly et al. 2018)
Semnan10	Bone	43	(Daly et al. 2018)
Semnan7	Bone	42	(Daly et al. 2018)
Yarmut7	Bone	42	(Daly et al. 2018)
Gilat2	Bone	41	(Daly et al. 2018)
Kohneh2	Bone	41	(Daly et al. 2018)
Blagotin16	Bone	40	(Daly et al. 2018)
Semnan8	Bone	40	(Daly et al. 2018)
Miqne5	Bone	39	(Daly et al. 2018)
Lur9	Bone	38	(Daly et al. 2018)
Yarmut1	Bone	38	(Daly et al. 2018)

Appendix 4.3 - The cytosine deamination probability in single strand context (δS) of published ancient skins.

Sample ID	Preservation type	Animal	Age (BP)	Nuc δS mean	Study
BOG5	Museum Skin	Goat	50	0.048499	(Cassidy et al. 2017)
BOG6	Museum Skin	Goat	50	0.076268	(Cassidy et al. 2017)
BOG1	Museum Skin	Goat	150	0.047499	(Cassidy et al. 2017)
SOG4	Museum Skin	Goat	150	0.081975	(Cassidy et al. 2017)
BOG4	Museum Skin	Goat	150	0.094258	(Cassidy et al. 2017)
ERR2019434	Parchment	Sheep	1000	0.041740	(Teasdale et al. 2017)
ERR2019417	Parchment	Sheep	1000	0.049633	(Teasdale et al. 2017)
ERR2019435	Parchment	Sheep	1000	0.050773	(Teasdale et al. 2017)
ERR2019416	Parchment	Sheep	1000	0.050881	(Teasdale et al. 2017)
ERR2019436	Parchment	Sheep	1000	0.061117	(Teasdale et al. 2017)
ERR2019433	Parchment	Sheep	1000	0.061145	(Teasdale et al. 2017)
ERR2019423	Parchment	Cow	1000	0.061309	(Teasdale et al. 2017)
ERR2019431	Parchment	Cow	1000	0.080282	(Teasdale et al. 2017)
ERR2019424	Parchment	Sheep	1000	0.081662	(Teasdale et al. 2017)
ERR2019418	Parchment	Cow	1000	0.082718	(Teasdale et al. 2017)
ERR2019419	Parchment	Cow	1000	0.100676	(Teasdale et al. 2017)

ERR2019420	Parchment	Cow	1000	0.108996	(Teasdale et al. 2017)
ERR2019421	Parchment	Cow	1000	0.114733	(Teasdale et al. 2017)
ERR2019426	Parchment	Cow	1000	0.114990	(Teasdale et al. 2017)
ERR2019422	Parchment	Cow	1000	0.132480	(Teasdale et al. 2017)
ERR2019427	Parchment	Cow	1000	0.145933	(Teasdale et al. 2017)
ERR2019429	Parchment	Cow	1000	0.156170	(Teasdale et al. 2017)
ERR2019428	Parchment	Cow	1000	0.173068	(Teasdale et al. 2017)
ERR2019430	Parchment	Cow	1000	0.190321	(Teasdale et al. 2017)
MUM2	Salt	Sheep	1600	0.012168	*This study*
ERR1596824	Leather	Goat	5300	0.062508	(O'Sullivan et al. 2016)
ERR1596825	Leather	Goat	5300	0.079484	(O'Sullivan et al. 2016)
ERR1596826	Leather	Sheep	5300	0.088118	(O'Sullivan et al. 2016)
ERR1596829	Leather	Sheep	5300	0.088544	(O'Sullivan et al. 2016)
ERR1596823	Leather	Cow	5300	0.160215	(O'Sullivan et al. 2016)
ERR1596828	Leather	Sheep	5300	0.160952	(O'Sullivan et al. 2016)
ERR1596832	Leather	Cow	5300	0.219441	(O'Sullivan et al. 2016)

Appendix Table 4.4 - f_3 statistics computed to quantify the amount of genetic drift shared between MUM2 and modern sheep breeds.

Breed	Test	Outgroup	$f3$	std.err	Z	SNPs
Karakas	MUM2	Asiatic Mouflon	0.184827	0.004245	43.54	27232
Moghani	MUM2	Asiatic Mouflon	0.184432	0.004175	44.176	27284
Qezel	MUM2	Asiatic Mouflon	0.184222	0.004082	45.133	27296
Afshari	MUM2	Asiatic Mouflon	0.183673	0.004175	43.989	27281
Norduz	MUM2	Asiatic Mouflon	0.183594	0.004299	42.711	27244
Garut	MUM2	Asiatic Mouflon	0.181645	0.004385	41.428	27033
LocalAwassi	MUM2	Asiatic Mouflon	0.1812	0.004	45.298	27292
Deccani	MUM2	Asiatic Mouflon	0.18002	0.00428	42.057	27231
CyprusFatTail	MUM2	Asiatic Mouflon	0.179281	0.004435	40.426	27133
IndianGarole	MUM2	Asiatic Mouflon	0.178473	0.004589	38.895	27054
BangladeshiBGE	MUM2	Asiatic Mouflon	0.17824	0.004469	39.887	27087
BangladeshiGarole	MUM2	Asiatic Mouflon	0.17573	0.004437	39.605	27063
Tibetan	MUM2	Asiatic Mouflon	0.174752	0.004073	42.906	27221
Sumatra	MUM2	Asiatic Mouflon	0.174154	0.004352	40.019	27161

EgyptianBarki	MUM2	Asiatic Mouflon	0.173716	0.003936	44.136	27268
Icelandic	MUM2	Asiatic Mouflon	0.173171	0.004245	40.797	27209
EthiopianMenz	MUM2	Asiatic Mouflon	0.173142	0.00407	42.543	27167
Changthangi	MUM2	Asiatic Mouflon	0.17311	0.004037	42.88	27261
RonderibAfrikaner	MUM2	Asiatic Mouflon	0.170808	0.004302	39.705	27067
Sakiz	MUM2	Asiatic Mouflon	0.170526	0.004454	38.285	27056
SriLankan	MUM2	Asiatic Mouflon	0.170332	0.005373	31.699	25616
Chios	MUM2	Asiatic Mouflon	0.169128	0.004218	40.095	27207
RedMaasai	MUM2	Asiatic Mouflon	0.16907	0.004102	41.217	27193
NamaquaAfrikaner	MUM2	Asiatic Mouflon	0.168173	0.004787	35.132	26663
AfecAssaf	MUM2	Asiatic Mouflon	0.167641	0.003983	42.093	27240
AfricanDorper	MUM2	Asiatic Mouflon	0.156866	0.003967	39.546	27208
BarbadosBlackBelly	MUM2	Asiatic Mouflon	0.155779	0.003964	39.3	27199
SantaInes	MUM2	Asiatic Mouflon	0.155649	0.004039	38.535	27241
BrazilianCreole	MUM2	Asiatic Mouflon	0.154707	0.003838	40.308	27274
AfricanWhiteDorper	MUM2	Asiatic Mouflon	0.153053	0.004248	36.031	26707
Comisana	MUM2	Asiatic Mouflon	0.152739	0.003841	39.766	27272
Castellana	MUM2	Asiatic Mouflon	0.152149	0.003924	38.774	27279
MoradaNova	MUM2	Asiatic Mouflon	0.151884	0.004117	36.89	27114
Churra	MUM2	Asiatic Mouflon	0.151037	0.003995	37.807	27274
Merinolandschaf	MUM2	Asiatic Mouflon	0.150845	0.003993	37.78	27234
ChineseMerino	MUM2	Asiatic Mouflon	0.150655	0.004046	37.238	27251
Arawapa	MUM2	Asiatic Mouflon	0.150431	0.004258	35.329	27197
SwissWhiteAlpineSheep	MUM2	Asiatic Mouflon	0.150333	0.004164	36.102	27219
StElizabeth	MUM2	Asiatic Mouflon	0.149948	0.003871	38.732	27192
Finnsheep	MUM2	Asiatic Mouflon	0.149657	0.003958	37.81	27232
MacarthurMerino	MUM2	Asiatic Mouflon	0.149321	0.004521	33.029	26301
Ojalada	MUM2	Asiatic Mouflon	0.149218	0.003922	38.051	27280
Rasaaragonesa	MUM2	Asiatic Mouflon	0.149031	0.003779	39.437	27293
AustralianPollMerino	MUM2	Asiatic Mouflon	0.148981	0.003882	38.374	27276
SwissMirrorSheep	MUM2	Asiatic Mouflon	0.148855	0.00406	36.661	27199
Leccese	MUM2	Asiatic Mouflon	0.148836	0.003954	37.642	27274
Altamurana	MUM2	Asiatic Mouflon	0.148819	0.003956	37.622	27260

BundnerOberlanderSheep	MUM2	Asiatic Mouflon	0.148756	0.004039	36.827	27176
AustralianMerino	MUM2	Asiatic Mouflon	0.148653	0.003748	39.664	27293
Galway	MUM2	Asiatic Mouflon	0.148388	0.004181	35.495	27132
Rambouillet	MUM2	Asiatic Mouflon	0.148148	0.00391	37.894	27263
EastFriesianWhite	MUM2	Asiatic Mouflon	0.147882	0.004157	35.576	26873
EngadineRedSheep	MUM2	Asiatic Mouflon	0.147589	0.003821	38.63	27258
BorderLeicester	MUM2	Asiatic Mouflon	0.147486	0.004437	33.24	26921
ValaisRedSheep	MUM2	Asiatic Mouflon	0.147246	0.004087	36.03	27050
GulfCoastNative	MUM2	Asiatic Mouflon	0.147145	0.003899	37.738	27288
DorsetHorn	MUM2	Asiatic Mouflon	0.146663	0.004358	33.652	27037
Lacaune	MUM2	Asiatic Mouflon	0.146614	0.003803	38.55	27296
ScottishTexel	MUM2	Asiatic Mouflon	0.146606	0.004263	34.388	27153
ValaisBlacknoseSheep	MUM2	Asiatic Mouflon	0.14636	0.004109	35.622	27039
IrishSuffolk	MUM2	Asiatic Mouflon	0.146218	0.004327	33.792	27139
ScottishBlackface	MUM2	Asiatic Mouflon	0.146137	0.003922	37.258	27219
EastFriesianBrown	MUM2	Asiatic Mouflon	0.146104	0.004192	34.857	27012
AustralianSuffolk	MUM2	Asiatic Mouflon	0.146069	0.004036	36.191	27244
AustralianCoopworth	MUM2	Asiatic Mouflon	0.14599	0.003905	37.388	27246
NewZealandRomney	MUM2	Asiatic Mouflon	0.14591	0.004078	35.782	27187
NewZealandTexel	MUM2	Asiatic Mouflon	0.145832	0.00421	34.643	27156
AustralianPollDorset	MUM2	Asiatic Mouflon	0.145581	0.00414	35.161	27146
OldNorwegianspaelsau	MUM2	Asiatic Mouflon	0.145571	0.003901	37.321	27246
SwissBlack-BrownMountainSheep	MUM2	Asiatic Mouflon	0.145536	0.00405	35.932	27201
GermanTexel	MUM2	Asiatic Mouflon	0.145535	0.004124	35.289	27205
SardinianAncestralBlack	MUM2	Asiatic Mouflon	0.145384	0.004296	33.839	27192
BlackHeadedMutton	MUM2	Asiatic Mouflon	0.145025	0.004181	34.689	27177
Wiltshire	MUM2	Asiatic Mouflon	0.144168	0.004525	31.857	26835
ImprovedAwassi	MUM2	Asiatic Mouflon	0.142946	0.004045	35.338	27063
Soay	MUM2	Asiatic Mouflon	0.139511	0.004426	31.522	26723
Boreray	MUM2	Asiatic Mouflon	0.137359	0.004534	30.293	26766

Appendix Table 4.5 - Base calls of 51 SNPs in the *PDGFD* gene associated with the derived fat-tail phenotype in MUM2.

SNP name	Position on chr15	Fat-tail allele	Thin-tail allele	MUM2 base calls	MUM2 fat-tail allele base calls	MUM2 thin-tail allele base calls
ss1139270974,ss1198779082	3854063	T	C	CCC	0	3
ss1139270977	3854426	A	G	AA	2	0
ss1139270978,ss1215832225,ss1198779102	3854617	A	G	GA	1	1
ss1139270979,ss1215832228,ss1198779106	3854930	A	T	AATTA	3	2
ss1139270980,ss1215832229,ss1198779108	3854933	C	T	CCTTC	3	2
ss1139270981,ss1215832230,ss1198779109	3854955	G	A	GGG	3	0
ss1139270982,ss1215832231,ss1198779113	3855064	T	G	GGTGGG	1	5
ss1139270984,ss1215832233,ss1198779118	3855194	A	T	TT	0	2
ss1139270986,ss1215832236,ss1198779122	3855324	T	A	AA	0	2
ss1139270987,ss1215832237,ss1198779126	3855347	A	G	AA	2	0
ss1139270990,ss1215832240,ss1198779129	3855390	T	C	CCTCTCCC	2	6
ss1139270991,ss1215832242,ss1198779131	3855407	T	C	TTTTTTTT	8	0
ss1139270992,ss1215832243,ss1198779137	3855530	A	G	GGAAGG	2	4
ss1139270994,ss1215832245,ss1198779144	3855601	T	C	GTTTT	4	0
ss1215832247,ss1198779147	3855636	C	T	CCCC	4	0
ss1215832252,ss1198779153	3855829	G	A	GGGGGGGGGGGG	12	0
ss1139270997,ss1215832253,ss1198779156	3855847	T	C	CTCTTCCTTCCT	6	6
ss1139270998,ss1215832255,ss1198779163	3855934	C	T	CCCCC	5	0
ss1139271000,ss1215832258	3856143	A	C	AAAAAAA	7	0
ss1139271002,ss1215832259,ss1198779172	3856224	A	G	AAAAAAA	7	0
ss1215832261,ss1198779175	3856258	A	C	AAAAA	5	0

ss1215832262,ss1198779177	3856263	A	G	AAAAA	5	0
ss1139271004,ss1215832263,ss1198779181	3856492	C	G	CGCGG	2	3
ss1139271005,ss1215832268,ss1198779184	3856682	G	A	G	1	0
ss1215832274,ss1198779195	3857032	G	A	GGG	3	0
ss1215832275,ss1198779196	3857037	C	T	CCC	3	0
ss1215832276,ss1198779198	3857065	C	T	CC	2	0
ss1139271006,ss1198779203	3857203	T	C	CT	1	1
ss1215832277,ss1198779205	3857217	T	C	TTT	3	0
ss1215832282,ss1198779213	3857345	A	G	AAAA	4	0
ss1139271007,ss1215832283	3857377	C	T	CCCCT	4	1
ss1139271008,ss1215832284	3857378	T	G	TTTTT	5	0
ss1215832286,ss1198779218	3857464	A	G	AAAAAA	6	0
ss1215832288,ss1198779220	3857494	T	C	TTTTTTTT	8	0
ss1139271011,ss1215832296,ss1198779236	3857817	T	C	TTTTT	5	0
ss1139271012,ss1215832297,ss1198779237	3857841	G	C	GGGGGG	6	0
ss1139271013,ss1215832298,ss1198779239	3857900	T	A	TTTTT	5	0
ss1139271014,ss1215832299,ss1198779241	3857938	C	A	CCC	3	0
ss1139271016,ss1198779258	3858241	T	G	TTTTTTTT	8	0
ss1139271017,ss1215832305,ss1198779259	3858247	G	A	GGGGGGG	7	0
ss1198779264	3858309	T	C	TTTT	4	0
ss1139271019,ss1215832309,ss1198779270	3858370	C	G	CC	2	0
ss1139271024,ss1215832322,ss1198779290	3858966	A	G	AGA	2	1
ss1139271031,ss1215832331,ss1198779312	3859314	C	T	CT	1	1
ss1139271040,ss1215832345,ss1198779342	3860194	T	C	CTCC	1	3

ss1139271042,ss1215832347,ss1198779345	3860228	T	C	CTTT	3	1
ss1215832355,ss1198 ss1139270980,ss1215832229,ss1198779108779360	3860577	G	A	GGGGGG	6	0
ss1139271048,ss1198779365	3860662	G	A	GGG	3	0
ss1215832357	3860685	C	A	CC	2	0
ss1215832359,ss1198779374	3860771	A	G	A	1	0
ss1215832363	3860894	T	C	TT	2	0

Appendix Table 5.1 - A dataset of published modern and ancient goat mitochondrial genomes used for mitochondrial analysis. * = reference sample used for ML Tree haplotype diagnosis. BEAST calibration refers to the tip date of ancient samples. Calibrations in brackets signifies that the tip date was supplied as a probability distribution due to age uncertainty.

Sample	Site	Location	Context	BEAST calibration	mtDNA	Reference
Abdul1	Tepe Abdul Hosein	Luristan, Iran	Neolithic	(8500.0,10500.0)	G	*(Daly et al. 2021)*
Abdul2	Tepe Abdul Hosein	Luristan, Iran	Neolithic	(8500.0,10500.0)	G	*(Daly et al. 2021)*
Abdul4	Tepe Abdul Hosein	Luristan, Iran	Neolithic	10139	G	*(Daly et al. 2021)*
Abdul5	Tepe Abdul Hosein	Luristan, Iran	Neolithic	(8500.0,10500.0)	G	*(Daly et al. 2021)*
Abdul6	Tepe Abdul Hosein	Luristan, Iran	Neolithic	9904	A	*(Daly et al. 2021)*
Abdul7	Tepe Abdul Hosein	Luristan, Iran	Neolithic	(8500.0,10500.0)	G	*(Daly et al. 2021)*
Abdul9	Tepe Abdul Hosein	Luristan, Iran	Neolithic	(8500.0,10500.0)	B	*(Daly et al. 2021)*
Acem1	Acemhöyük	Aksaray Plain, Turkey	Bronze Age	6408	A	*(Daly et al. 2018)*
Acem2	Acemhöyük	Aksaray Plain, Turkey	Bronze Age	(3772.0,4822.0)	A	*(Daly et al. 2018)*
Ainghazal1	Ain'Ghazal	Amman, Jordan	Neolithic	(9340.0,10233.0)	F	*(Daly et al. 2018)*
Ainghazal2	Ain'Ghazal	Amman, Jordan	Neolithic	(9340.0,10233.0)	F	*(Daly et al. 2018)*
Ainghazal3	Ain'Ghazal	Amman, Jordan	Neolithic	(9340.0,10233.0)	F	*(Daly et al. 2018)*
Ainghazal4	Ain'Ghazal	Amman, Jordan	Neolithic	(9340.0,10233.0)	F	*(Daly et al. 2018)*
AP44	Aşağı Pınar	Kirklareli, Turkey	Neolithic	(7022.0,7522.0)	A	*(Daly et al. 2018)*
AP45	Aşağı Pınar	Kirklareli, Turkey	Neolithic	(7022.0,7322.0)	A	*(Daly et al. 2018)*
AP46	Aşağı Pınar	Kirklareli, Turkey	Neolithic	7197	C	*(Daly et al. 2018)*
AP49	Aşağı Pınar	Kirklareli, Turkey	Neolithic	(7222.0,7522.0)	A	*(Daly et al. 2018)*
AP50	Aşağı Pınar	Kirklareli, Turkey	Neolithic	(7022.0,7322.0)	A	*(Daly et al. 2018)*

Azer3-5	Tepe Hasanlu	Western Azerbaijan, Iran	Bronze Age	(4122.0,4222.0)	A	*(Daly et al. 2018)*
Azer4	Tepe Hasanlu	Western Azerbaijan, Iran	Iron Age	(2352.0,2572.0)	A	*(Daly et al. 2018)*
Azer6	Soha Chay Tepe	Zanjan, Iran	Chalcolithic	(5700.0,6700.0)	A	*(Daly et al. 2018)*
Blagotin1	Blagotin-Poljna	Trstenik, Serbia	Neolithic	7197	A	*(Daly et al. 2018)*
Blagotin16	Blagotin-Poljna	Trstenik, Serbia	Neolithic	(7500.0,8500.0)	A	*(Daly et al. 2018)*
Blagotin2	Blagotin-Poljna	Trstenik, Serbia	Neolithic	8251	A	*(Daly et al. 2018)*
Blagotin3	Blagotin-Poljna	Trstenik, Serbia	Neolithic	8016	A	*(Daly et al. 2018)*
Bulak1-3	Tilla Bulak	Surkhandarja, Uzbekistan	Bronze Age	(3722.0,4022.0)	A	*(Daly et al. 2018)*
Bulak2	Tilla Bulak	Surkhandarja, Uzbekistan	Bronze Age	(3722.0,4022.0)	A	*(Daly et al. 2018)*
Bulak4	Tilla Bulak	Surkhandarja, Uzbekistan	Bronze Age	(3722.0,4022.0)	B	*(Daly et al. 2018)*
Bulak5	Tilla Bulak	Surkhandarja, Uzbekistan	Bronze Age	(3722.0,4022.0)	D	*(Daly et al. 2018)*
Cav8	Čavdar	Sofia District, Bulgaria	Neolithic	(7522.0,8022.0)	A	*(Daly et al. 2018)*
Chalow1	Chalow	Khorasan, Iran	Bronze Age	(4022.0,4322.0)	D	*(Daly et al. 2018)*
Darre1	Darre-ye Bolāghi	Fars, Iran	Chalcolithic	(6022.0,7022.0)	A	*(Daly et al. 2018)*
Darre2	Darre-ye Bolāghi	Fars, Iran	Chalcolithic	1557	A	*(Daly et al. 2018)*
Direkli1-2	Direkli Cave	Taurus Mountains, Turkey	Epipaleolithic	13281	T	*(Daly et al. 2018)*
Direkli4	Direkli Cave	Taurus Mountains, Turkey	Epipaleolithic	12037	F	*(Daly et al. 2018)*
Direkli5	Direkli Cave	Taurus Mountains, Turkey	Epipaleolithic	(11022.0,15022.0)	T	*(Daly et al. 2018)*
Direkli6	Direkli Cave	Taurus Mountains, Turkey	Epipaleolithic	(11022.0,15022.0)	T	*(Daly et al. 2018)*
Dra34	Merdžumekja	Drama, Bulgaria	Chalcolithic	6489	G	*(Daly et al. 2018)*
Fars1	Rahmat Abad	Fars, Iran	Chalcolithic	(4100.0,5100.0)	A	*(Daly et al. 2018)*
Fars2-5	Rahmat Abad	Fars, Iran	Neolithic	8932	B	*(Daly et al. 2018)*
Fars4	Mianrud	Fars, Iran	Chalcolithic	7358	A	*(Daly et al. 2018)*
Ganjdareh1	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	G'	*(Daly et al. 2021)*
Ganjdareh10	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	A	*(Daly et al. 2021)*
Ganjdareh14	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	A	*(Daly et al. 2021)*
Ganjdareh15	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	D	*(Daly et al. 2021)*
Ganjdareh16	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	G'	*(Daly et al. 2021)*
Ganjdareh18	Ganj Dareh	Iran	Neolithic	10003	A	*(Daly et al. 2021)*
Ganjdareh19	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	D	*(Daly et al. 2021)*
Ganjdareh2	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	G	*(Daly et al. 2021)*
Ganjdareh20	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	B	*(Daly et al. 2021)*
Ganjdareh21	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	G'	*(Daly et al. 2021)*

Ganjdareh22	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	B	*(Daly et al. 2021)*
Ganjdareh23	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	D	*(Daly et al. 2021)*
Ganjdareh24	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	C bez	*(Daly et al. 2021)*
Ganjdareh26	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	D bez	*(Daly et al. 2021)*
Ganjdareh27	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	G	*(Daly et al. 2021)*
Ganjdareh28	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	A	*(Daly et al. 2021)*
Ganjdareh29	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	C	*(Daly et al. 2021)*
Ganjdareh3	Ganj Dareh	Iran	Neolithic	9788	G'	*(Daly et al. 2021)*
Ganjdareh34	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	A	*(Daly et al. 2021)*
Ganjdareh35	Ganj Dareh	Iran	Neolithic	9954	F	*(Daly et al. 2021)*
Ganjdareh5	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	D	*(Daly et al. 2021)*
Ganjdareh6	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	D	*(Daly et al. 2021)*
Ganjdareh7	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	G	*(Daly et al. 2021)*
Ganjdareh8	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	B	*(Daly et al. 2021)*
Ganjdareh9	Ganj Dareh	Iran	Neolithic	(8500.0,10500.0)	B	*(Daly et al. 2021)*
Geor2	Tamara Fort	Kazbegi, Georgia	Mediaeval	(500.0,1000.0)	A	*(Daly et al. 2018)*
Ghosh5	Abu Ghosh	Judean Hills, Levant	Neolithic	(9500.0,10500.0)	F	*(Daly et al. 2018)*
Gilat10	Gilat	Northern Negev, Levant	Chalcolithic	(6222.0,6522.0)	A	*(Daly et al. 2018)*
Gilat2	Gilat	Northern Negev, Levant	Chalcolithic	(6222.0,6522.0)	A	*(Daly et al. 2018)*
Gilat8	Gilat	Northern Negev, Levant	Chalcolithic	(6222.0,6522.0)	D	*(Daly et al. 2018)*
Hovk1	Hovk-1 Cave	Tavush, Armenia	Paleolithic	(47074.0,200000.0)	F	*(Daly et al. 2018)*
Kan19	Kanlıgeçit	Kirklareli,Turkey	Bronze Age	(4222.0,4722.0)	A	*(Daly et al. 2018)*
Kan23	Kanlıgeçit	Kirklareli,Turkey	Bronze Age	4671	G	*(Daly et al. 2018)*
Kan25	Kanlıgeçit	Kirklareli,Turkey	Bronze Age	(4222.0,4722.0)	A	*(Daly et al. 2018)*
Kazbeg1	Tamara Fort	Kazbegi, Georgia	Mediaeval	1050	A	*(Daly et al. 2018)*
Kohneh2	Kohneh Tepesi	Azerbaijan, Iran	Bronze Age	(5022.0,5322.0)	A	*(Daly et al. 2018)*
Kov27	Kovačevo	Blagoevgrad, Bulgaria	Neolithic	(7622.0,8222.0)	A	*(Daly et al. 2018)*
Kov57	Kovačevo	Blagoevgrad, Bulgaria	Neolithic	(7622.0,8222.0)	A	*(Daly et al. 2018)*
Kov60	Kovačevo	Blagoevgrad, Bulgaria	Neolithic	(7622.0,8222.0)	A	*(Daly et al. 2018)*
KR059146	-	-	modern	0	A1	*(Colli et al. 2015)*
KR059147	-	-	modern	0	A1a	*(Colli et al. 2015)*
KR059152	-	-	modern	0	A2	*(Colli et al. 2015)*
KR059155	-	-	modern	0	A2a	*(Colli et al. 2015)*

KR059157	-	-	modern	0	A2a1	*(Colli et al. 2015)*
KR059165	-	-	modern	0	A3	*(Colli et al. 2015)*
KR059170	-	-	modern	0	A4	*(Colli et al. 2015)*
KR059174	-	-	modern	0	A5	*(Colli et al. 2015)*
KR059176	-	-	modern	0	A5	*(Colli et al. 2015)*
KR059178	-	-	modern	0	A6	*(Colli et al. 2015)*
KR059180	-	-	modern	0	A7	*(Colli et al. 2015)*
KR059183*	-	-	modern	0	A	*(Colli et al. 2015)*
KR059190	-	-	modern	0	A	*(Colli et al. 2015)*
KR059210	-	-	modern	0	D	*(Colli et al. 2015)*
KR059211	-	-	modern	0	D1	*(Colli et al. 2015)*
KR059213	-	-	modern	0	G	*(Colli et al. 2015)*
KR059219*	-	-	modern	0	B	*(Colli et al. 2015)*
KR059220	-	-	modern	0	B1	*(Colli et al. 2015)*
KR059221*	-	-	modern	0	C	*(Colli et al. 2015)*
KR059222	-	-	modern	0	C1	*(Colli et al. 2015)*
KR059223	-	-	modern	0	C1a	*(Colli et al. 2015)*
KR059226*	-	-	modern	0	F	*(Colli et al. 2015)*
Lur12	Tepe Abdul Hosein	Luristan, Iran	Neolithic	9977	G	*(Daly et al. 2018)*
Lur9	Kelek Asad Morad	Luristan, Iran	Neolithic	(10222.0,10522.0)	B	*(Daly et al. 2018)*
Miqne5	Tel Miqne-Ekron	Shephelah, Levant	Iron Age	(2722.0,3622.0)	A	*(Daly et al. 2018)*
Monjukli1	Monjukli Depe	Meana-Čaača, Turkmenistan	Chalcolithic	(6522.0,7122.0)	A	*(Daly et al. 2018)*
Monjukli2	Monjukli Depe	Meana-Čaača, Turkmenistan	Chalcolithic	(6522.0,7122.0)	D	*(Daly et al. 2018)*
Monjukli4	Monjukli Depe	Meana-Čaača, Turkmenistan	Chalcolithic	(6522.0,7122.0)	A	*(Daly et al. 2018)*
Monjukli6	Monjukli Depe	Meana-Čaača, Turkmenistan	Chalcolithic	(6522.0,7122.0)	D	*(Daly et al. 2018)*
Monjukli7	Monjukli Depe	Meana-Čaača, Turkmenistan	Neolithic	(7922.0,8422.0)	D	*(Daly et al. 2018)*
Monjukli8	Monjukli Depe	Meana-Čaača, Turkmenistan	Neolithic	(7922.0,8422.0)	D	*(Daly et al. 2018)*
Monjukli9	Monjukli Depe	Meana-Čaača, Turkmenistan	Neolithic	(7922.0,8422.0)	G	*(Daly et al. 2018)*
NC_005044.2	-	-	modern	0		*(Hassanin et al. 2010)*
NC_028161*	-	-	modern	0	G	*Unpublished*
Ovc11	Ovčarovo-gorata	Tărgovište, Bulgaria	Neolithic	(7522.0,7722.0)	A	*(Daly et al. 2018)*
Pie17	Pietrele	Giurgiu, Romania	Chalcolithic	(6272.0,6472.0)	A	*(Daly et al. 2018)*

Potterne1	Potterne	Wiltshire, UK	Bronze Age	(3012.0,4062.0)	A	*(Daly et al. 2018)*
Qazvin1	Tepe Chizar	Qazvin, Iran	Bronze Age	(3922.0,4422.0)	A	*(Daly et al. 2018)*
Safi2	Tel es-Safi/Gath	Ashkelon, Levant	Bronze Age	(4592.0,4922.0)	A	*(Daly et al. 2018)*
Semnan1-2	Sang-e Chakmaq	Semnan, Iran	Neolithic	9176	B	*(Daly et al. 2018)*
Semnan10	Sang-e Chakmaq	Semnan, Iran	Neolithic	(7500.0,9500.0)	G	*(Daly et al. 2018)*
Semnan13	Sang-e Chakmaq	Semnan, Iran	Neolithic	(7500.0,9500.0)	D	*(Daly et al. 2018)*
Semnan17	Sang-e Chakmaq	Semnan, Iran	Neolithic	(7500.0,9500.0)	D	*(Daly et al. 2018)*
Semnan3	Sang-e Chakmaq	Semnan, Iran	Neolithic	8131	D	*(Daly et al. 2018)*
Semnan7	Sang-e Chakmaq	Semnan, Iran	Neolithic	(7500.0,9500.0)	D	*(Daly et al. 2018)*
Semnan8	Sang-e Chakmaq	Semnan, Iran	Neolithic	(7500.0,9500.0)	D	*(Daly et al. 2018)*
Semnan9	Sang-e Chakmaq	Semnan, Iran	Neolithic	(7500.0,9500.0)	G	*(Daly et al. 2018)*
Shiqmim1	Shiqmim	Northern Negev, Levant	Chalcolithic	(5722.0,6322.0)	D	*(Daly et al. 2018)*
Shiqmim9	Shiqmim	Northern Negev, Levant	Chalcolithic	(5722.0,6322.0)	D	*(Daly et al. 2018)*
Tac1	Tachti Perda	Kakheti, Georgia	Bronze Age	(3022.0,3422.0)	A	*(Daly et al. 2018)*
Tac2	Tachti Perda	Kakheti, Georgia	Iron Age	(2722.0,3022.0)	A	*(Daly et al. 2018)*
Tac3	Tachti Perda	Kakheti, Georgia	Bronze Age	(3022.0,3422.0)	A	*(Daly et al. 2018)*
Uiv17	Uilvar	Timişoara, Romania	Neolithic	(8122.0,8422.0)	A	*(Daly et al. 2018)*
Ulu38	Ulucak Höyük	Izmir, Turkey	Neolithic	(7072.0,7272.0)	A	*(Daly et al. 2018)*
Yarmut1	Tel Yarmuth	Bet Shemesh, Levant	Bronze Age	(4522.0,4722.0)	A	*(Daly et al. 2018)*
Yarmut7	Tel Yarmuth	Bet Shemesh, Levant	Bronze Age	(4222.0,4672.0)	A	*(Daly et al. 2018)*
Yoqneam2	Tel Yoqne'am	Haifa, Levant	Bronze Age	(3562.0,3672.0)	A	*(Daly et al. 2018)*

Appendix Table 5.2 - Alignment statistics of coprolite samples.

Sample	Index Pairing	Raw Reads	Trimmed Reads	Raw Aligned	Raw endog %	mq30 Aligned	mq30 endog %	Clonality
CR073	10-62	2179623	1650265	391239	23.71	45771	2.77	0.36
CR073	10-63	1564420	1199887	287967	24.00	32918	2.74	0.37
CR073	10-64	2264657	1733743	420494	24.25	41167	2.37	0.41
CR073	10-65	8598659	6567897	1586869	24.16	59519	0.91	0.56
CR073	10-66	13479491	10295085	2548043	24.75	64448	0.63	0.61
CR073	10-67	10635192	8091167	1981631	24.49	60969	0.75	0.59
CR073	10-101	9-58418	15106781	3507339	23.22	95072	0.63	0.47
CR073	10-102	5-61685	16098588	3690138	22.92	95771	0.59	0.47
CR073	10-103	6-44036	11179820	2583277	23.11	93835	0.84	0.43
CR073	10-104	1-61159	15765181	3634278	23.05	93583	0.59	0.47

CR073	10-60	7922375	5773958	1333351	23.09	79330	1.37	0.36
CR073	10-61	28971151	21309301	4964347	23.30	141977	0.67	0.49
CR073	10-68	23528865	17117360	3955838	23.11	100765	0.59	0.47
CR073	10-69	9688737	7077144	1650968	23.33	83920	1.19	0.39
CR073	10-70	16699418	12271234	2843854	23.17	94596	0.77	0.45
CR076	9-51	10128025	5612908	451295	8.04	40502	0.72	0.65
CR076	9-52	10370388	5698670	458902	8.05	39225	0.69	0.66
CR076	9-53	21323736	11796723	907006	7.69	48468	0.41	0.71
CR076	9-54	22478825	12354441	975164	7.89	47130	0.38	0.73
CR076	9-55	15781460	8673972	684594	7.89	40229	0.46	0.70
CR079	11-58	40690206	19483024	4304337	22.09	175452	0.43	0.49
CR079	11-59	8140493	3882254	862270	22.21	74533	0.92	0.37
CR079	11-56	749173	378640	84616	22.35	10153	1.36	0.17
CR079	11-57	491315	237802	51589	21.69	6019	1.23	0.16

Appendix Table 5.3 - Mitochondrial enrichment for eight coprolite samples.

Sample	Era	Shotgun total reads	Shotgun aligned reads	Proportion	Capture total reads	Capture aligned	Proportion	Enrichment (X)	Coverage (X)	Haplotype
CR073	Late Neolithic	386,630	37	0.0001	218373	1801	0.008	86.18	16.5	A
CR076	Bronze Age	783,420	77	0.0001	319497	3246	0.01	103.37	16.25	A
CR077	Recent	478,119	15	0.00003	148903	1215	0.008	260.09	3.1	A
CR078	Recent	645062	9	0.00001	59,019	617	0.01	749.29	1.85	A
CR079	Late Neolithic	196947	19	0.0001	157,968	2017	0.013	132.35	5.13	A
CR084	Bronze Age	77,127	4	0.00005	60672	717	0.012	227.86	2.06	A
CR089	Bronze Age	10,292	1	0.0001	8596	48	0.006	57.47	0.11	A
CR092	Recent	92,417	24	0.00026	111450	10459	0.094	361.37	32.35	A

Appendix Table 5.4 - Heterozygosity of coprolites and published ancient goats bams of haplotype A.

Sample	Era	Reads	Heterozygosity of Haplotypic Sites	Heterozygosity of All Sites	Heterozygosity of Heteroplasmic Sites
acem1	Ancient Bone	118982	0.533	0.159	0.500

acem2	Ancient Bone	212113	0.362	0.117	0.241
AP44	Ancient Bone	1062	0.694	0.221	27.273
AP45	Ancient Bone	1851	0.405	0.139	14.341
AP49	Ancient Bone	3694	0.208	0.143	7.704
AP50	Ancient Bone	4237	0.346	0.101	6.870
azer3	Ancient Bone	57196	0.303	0.100	0.695
azer4	Ancient Bone	79652	0.165	0.097	0.519
azer5	Ancient Bone	77510	0.258	0.110	0.536
azer6	Ancient Bone	9819	0.349	0.074	2.220
blagotin1	Ancient Bone	154745	0.379	0.114	0.329
blagotin2	Ancient Bone	108416	0.167	0.108	0.454
blagotin3	Ancient Bone	274479	0.051	0.146	0.323
bulak1	Ancient Bone	56375	0.229	0.087	0.545
bulak2	Ancient Bone	67245	2.222	0.073	0.449
bulak3	Ancient Bone	50383	0.207	0.084	0.634
darre1	Ancient Bone	25911	0.317	0.108	1.088
darre2	Ancient Bone	74873	0.228	0.097	0.551
fars1	Ancient Bone	17956	0.362	0.112	1.601
fars4	Ancient Bone	30219	0.396	0.137	1.343
gilat10	Ancient Bone	635	2.439	0.144	43.636
gilat2	Ancient Bone	7018	0.526	0.112	3.652
gilat8	Ancient Bone	34415	0.378	0.106	0.862
Kan19	Ancient Bone	2503	0.623	0.151	10.100
Kan25	Ancient Bone	2712	0.272	0.157	10.194
kazbeg1	Ancient Bone	47986	0.298	0.114	0.833
kohneh2	Ancient Bone	9287	0.542	0.181	3.915
Kov27	Ancient Bone	824	0.847	0.098	28.090
Kov57	Ancient Bone	1641	2.500	0.319	17.189
Kov60	Ancient Bone	1859	0.717	0.168	13.067
miqne5	Ancient Bone	5330	0.158	0.096	4.852
monjukli1	Ancient Bone	11230	0.303	0.123	2.613
monjukli4	Ancient Bone	23942	0.239	0.143	1.483
Ovc11	Ancient Bone	2505	0.304	0.101	10.375

Pie17	Ancient Bone	2152	3.546	0.230	13.130
pottern1	Ancient Bone	35676	0.443	0.167	1.407
qazvin1	Ancient Bone	89205	0.118	0.085	0.365
safi2	Ancient Bone	36250	0.127	0.117	0.823
semnan1	Ancient Bone	64788	0.641	0.147	0.744
semnan2	Ancient Bone	142510	0.644	0.159	0.452
Tac2	Ancient Bone	2688	0.372	0.131	7.836
Tac3	Ancient Bone	6104	0.394	0.149	3.959
yarmuth1	Ancient Bone	13327	0.568	0.090	2.157
yarmuth7	Ancient Bone	3937	0.622	0.118	6.503
yoqneam2	Ancient Bone	92186	0.158	0.091	0.382
CR073	Neolithic Coprolite	1760	1.290	0.998	13.498
CR079	Neolithic Coprolite	1984	0.000	0.585	12.187
CR076	Bronze Age Coprolite	3228	1.099	0.222	8.797
CR084	Bronze Age Coprolite	716	0.000	0.221	35.211
CR077	Recent Coprolite	1187	2.128	1.095	19.357
CR078	Recent Coprolite	614	0.000	0.239	42.197
CR092	Recent Coprolite	10394	0.214	0.093	3.062

Appendix Table 6.1 - Repetitive sequence enrichment for leather samples and comparative samples.

Sample	Material	Study	Genome Coverage (X)	Repetitive Region Coverage (X)	Genic Region Coverage (X)	Relative Repetitive Coverage	Relative Genic Coverage	Enrichment of Repetitive Regions (X)
Leather-10	Leather (Modern)	This study	5.4200	9.43	1.58	1.74	0.29	5.97
Leather-2	Leather (Modern)	This study	4.6500	8.33	1.08	1.79	0.23	7.71
Leather-6	Leather (Modern)	This study	4.6900	7.33	2.03	1.56	0.43	3.61
ERR2019418	Parchment (Ancient)	(Teasdale et al. 2017)	0.0069	0.012	0.003	1.74	0.41	4.27
ERR2019419	Parchment (Ancient)	(Teasdale et al. 2017)	0.0042	0.008	0.001	1.85	0.31	5.86
ERR2019420	Parchment (Ancient)	(Teasdale et al. 2017)	0.0157	0.029	0.003	1.85	0.19	9.60
ERR2019421	Parchment (Ancient)	(Teasdale et al. 2017)	0.0130	0.025	0.001	1.90	0.11	17.75

ERR2019422	Parchment (Ancient)	(Teasdale et al. 2017)	0.0056	0.011	0.001	1.91	0.13	14.53
ERR2019426	Parchment (Ancient)	(Teasdale et al. 2017)	0.0039	0.007	0.002	1.73	0.40	4.31
ERR2019427	Parchment (Ancient)	(Teasdale et al. 2017)	0.0033	0.006	0.001	1.84	0.30	6.16
ERR2019428	Parchment (Ancient)	(Teasdale et al. 2017)	0.0118	0.022	0.002	1.84	0.19	9.70
ERR2019429	Parchment (Ancient)	(Teasdale et al. 2017)	0.0149	0.028	0.002	1.89	0.11	17.31
ERR2019430	Parchment (Ancient)	(Teasdale et al. 2017)	0.0058	0.011	0.001	1.91	0.12	15.79
Gyu2	Bone (Ancient)	(Verdugo et al. 2019)	3.4800	4.16	3.70	1.19	1.06	1.12
Has3	Bone (Ancient)	(Verdugo et al. 2019)	4.6700	5.54	4.89	1.19	1.05	1.13
Has4	Bone (Ancient)	(Verdugo et al. 2019)	4.3100	5.07	4.26	1.18	0.99	1.19
MV013	Bone (Ancient)	(Verdugo et al. 2019)	6.1000	7.37	4.70	1.21	0.77	1.57
MV043	Bone (Ancient)	(Verdugo et al. 2019)	4.8900	5.82	4.38	1.19	0.90	1.33
Plo3	Bone (Ancient)	(Verdugo et al. 2019)	4.0900	4.56	3.55	1.12	0.87	1.29
Plo4	Bone (Ancient)	(Verdugo et al. 2019)	4.4800	5.05	3.90	1.13	0.87	1.29
Qaz1	Bone (Ancient)	(Verdugo et al. 2019)	4.9500	5.93	5.15	1.20	1.04	1.15
SRR1188706	Blood (Modern)	(Hayes and Daetwyler 2019)	3.2100	3.64	2.53	1.14	0.79	1.44
SRR1205973	Blood (Modern)	(Hayes and Daetwyler 2019)	18.1300	19.01	15.25	1.05	0.84	1.25
SRR1205992	Blood (Modern)	(Hayes and Daetwyler 2019)	7.8300	8.23	6.45	1.05	0.82	1.28
SRR1262533	Blood (Modern)	(Hayes and Daetwyler 2019)	4.2300	4.53	2.79	1.07	0.66	1.62
SRR1262536	Blood (Modern)	(Hayes and Daetwyler 2019)	3.2600	3.57	2.33	1.09	0.72	1.53
SRR1262538	Blood (Modern)	(Hayes and Daetwyler 2019)	4.5600	4.92	3.26	1.08	0.72	1.51
SRR1262539	Blood (Modern)	(Hayes and Daetwyler 2019)	4.9700	5.28	3.19	1.06	0.64	1.66
SRR1262614	Blood (Modern)	(Hayes and Daetwyler 2019)	8.8500	9.33	6.41	1.05	0.72	1.46
SRR11569199	Skin (Modern)	(Dean et al. 2021)	0.3000	0.37	0.21	1.22	0.70	1.75
SRR11569200	Skin (Modern)	(Dean et al. 2021)	0.4500	0.56	0.33	1.24	0.72	1.71
SRR11569201	Skin (Modern)	(Dean et al. 2021)	0.4700	0.58	0.33	1.23	0.69	1.78
SRR11569202	Skin (Modern)	(Dean et al. 2021)	0.3400	0.41	0.23	1.21	0.68	1.80

SRR11569203	Skin (Modern)	(Dean et al. 2021)	0.4300	0.53	0.31	1.23	0.72	1.71

Appendix Table 6.2 - The average percentage of reads with a MAQ score in ancient parchment and bone samples.

Sample	Animal	Material	Study	Total aligned reads post PCR duplicate removal	Number of aligned reads with mapping quality = 0	Percent of reads with mapping quality = 0
ERR2019418	Bovine	Parchment (Ancient)	(Teasdale et al. 2017)	339146	224886	66.31
ERR2019419	Bovine	Parchment (Ancient)	(Teasdale et al. 2017)	207718	150273	72.34
ERR2019420	Bovine	Parchment (Ancient)	(Teasdale et al. 2017)	778498	561514	72.13
ERR2019421	Bovine	Parchment (Ancient)	(Teasdale et al. 2017)	670961	523253	77.99
ERR2019422	Bovine	Parchment (Ancient)	(Teasdale et al. 2017)	277954	212375	76.41
ERR2019426	Bovine	Parchment (Ancient)	(Teasdale et al. 2017)	162641	106910	65.73
ERR2019427	Bovine	Parchment (Ancient)	(Teasdale et al. 2017)	142638	102775	72.05
ERR2019428	Bovine	Parchment (Ancient)	(Teasdale et al. 2017)	503392	362404	71.99
ERR2019429	Bovine	Parchment (Ancient)	(Teasdale et al. 2017)	706226	548562	77.68
ERR2019430	Bovine	Parchment (Ancient)	(Teasdale et al. 2017)	256822	195910	76.28
Gyu2	Bovine	Bone (Ancient)	(Verdugo et al. 2019)	196125714	65169346	33.23
Has3	Bovine	Bone (Ancient)	(Verdugo et al. 2019)	278108156	92395392	33.22
Has4	Bovine	Bone (Ancient)	(Verdugo et al. 2019)	262063362	85994471	32.81
MV013	Bovine	Bone (Ancient)	(Verdugo et al. 2019)	316996656	105532815	33.29
MV043	Bovine	Bone (Ancient)	(Verdugo et al. 2019)	307540350	102022102	33.17
Plo3	Bovine	Bone (Ancient)	(Verdugo et al. 2019)	234239128	61730076	26.35
Plo4	Bovine	Bone (Ancient)	(Verdugo et al. 2019)	257671202	72093199	27.98
Qaz1	Bovine	Bone (Ancient)	(Verdugo et al. 2019)	298007199	99120110	33.26

Appendix Table 6.3 - The average ratio of mtDNA reads to nucDNA reads in different tissue types in both ancient and modern samples.

Sample	Material	Study	Total aligned reads to V00654 post PCR duplicate removal	Total aligned reads to bosTau8 post PCR duplicate removal	Ratio mtDNA: nucDNA
EF5R-10	Leather (Modern)	*This study*	97993	330017350	3367.76
EF5R-2-2	Leather (Modern)	*This study*	109303	307214827	2810.67

EF5R-6	Leather (Modern)	*This study*	80111	263220533	3285.70
ERR2019418	Parchment (Ancient)	(Teasdale et al. 2017)	356	339146	952.66
ERR2019419	Parchment (Ancient)	(Teasdale et al. 2017)	54	207718	3846.63
ERR2019420	Parchment (Ancient)	(Teasdale et al. 2017)	250	778498	3113.99
ERR2019421	Parchment (Ancient)	(Teasdale et al. 2017)	196	670961	3423.27
ERR2019422	Parchment (Ancient)	(Teasdale et al. 2017)	62	277954	4483.13
ERR2019426	Parchment (Ancient)	(Teasdale et al. 2017)	159	162641	1022.90
ERR2019427	Parchment (Ancient)	(Teasdale et al. 2017)	51	142638	2796.82
ERR2019428	Parchment (Ancient)	(Teasdale et al. 2017)	182	503392	2765.89
ERR2019429	Parchment (Ancient)	(Teasdale et al. 2017)	170	706226	4154.27
ERR2019430	Parchment (Ancient)	(Teasdale et al. 2017)	82	256822	3131.98
Gyu2	Bone (Ancient)	(Verdugo et al. 2019)	253591	196125714	773.39
Has3	Bone (Ancient)	(Verdugo et al. 2019)	173316	278108156	1604.63
Has4	Bone (Ancient)	(Verdugo et al. 2019)	123911	262063362	2114.93
MV013	Bone (Ancient)	(Verdugo et al. 2019)	193821	316996656	1635.51
MV043	Bone (Ancient)	(Verdugo et al. 2019)	73225	307540350	4199.94
Plo3	Bone (Ancient)	(Verdugo et al. 2019)	71096	234239128	3294.69
Plo4	Bone (Ancient)	(Verdugo et al. 2019)	85587	257671202	3010.63
Qaz1	Bone (Ancient)	(Verdugo et al. 2019)	139643	298007199	2134.06
SRR1188706	Blood (Modern)	(Hayes and Daetwyler 2019)	105	90504955	861951.95
SRR1205973	Blood (Modern)	(Hayes and Daetwyler 2019)	471820	475683853	1008.19
SRR1205992	Blood (Modern)	(Hayes and Daetwyler 2019)	269969	205315886	760.52
SRR1262533	Blood (Modern)	(Hayes and Daetwyler 2019)	1191	111671165	93762.52
SRR1262536	Blood (Modern)	(Hayes and Daetwyler 2019)	113	85945398	760578.74
SRR1262538	Blood (Modern)	(Hayes and Daetwyler 2019)	1191	120263459	100976.88
SRR1262539	Blood (Modern)	(Hayes and Daetwyler 2019)	459	130996902	285396.30
SRR1262614	Blood (Modern)	(Hayes and Daetwyler 2019)	244500	243985990	997.90
SRR11569199	Skin (Modern)	(Dean et al. 2021)	532	6277106	11799.07
SRR11569200	Skin (Modern)	(Dean et al. 2021)	4114	9571310	2326.52

SRR11569201	Skin (Modern)	(Dean et al. 2021)	4265	9900149	2321.25
SRR11569202	Skin (Modern)	(Dean et al. 2021)	423	7134042	16865.35
SRR11569203	Skin (Modern)	(Dean et al. 2021)	1526	9134175	5985.70

Appendix Table 7.1 - Sequencing results for 40 samples of *Bos* and *Bison* selected for mitochondrial sequence enrichment. Where possible a haplotype was assigned. Samples CR140, CR155, ER010, ER025 displayed clustering with other *Bos primigenius* haplotypes.

Sample	Species	Shotgun total reads	Shotgun aligned reads	Proportion of mtDNA alignment	Capture total reads	Capture reads aligned	Proportion of mtDNA alignment	Enrichment (X)	Haplotype
CR129	*Bos taurus*	1175313	4	0.000003	4443281	8661	0.0019492353	572.74	T3
CR103	*Bos taurus*	464758	104	0.000224	104213	15701	0.1506625853	673.29	T1
CR140	*Bos primigenius*	1435066	3	0.000002	4143139	38	0.0000091718	4.39	*Primigenius*
CR155	*Bos primigenius*	795654	8	0.00001	270364	36	0.0001331538	13.24	*Primigenius*
ER010	*Bos primigenius*	3988152	21	0.000005	322363	174	0.0005397642	102.51	*Primigenius*
ER025	*Bos primigenius*	13831815	9	0.000001	7910715	666	0.0000841896	129.39	*Primigenius*
CR107	*Bos primigenius*	795869	17	0.000021	1234455	12359	0.0100117056	468.71	P
CR122	*Bos primigenius*	218432	43	0.000197	2495	266	0.1066132265	541.58	P
CR133	*Bos primigenius*	681226	22	0.000032	371538	123	0.0003310563	10.25	P
CR134	*Bos primigenius*	673261	29	0.000043	219099	151	0.0006891862	16	P
CR147	*Bos primigenius*	712110	26	0.000037	407697	120	0.0002943362	8.06	P
CR148	*Bos primigenius*	519031	26	0.00005	148653	86	0.0005785285	11.55	P
CR149	*Bos primigenius*	833894	13	0.000016	316737	77	0.0002431039	15.59	P
ER003	*Bos primigenius*	16955597	4306	0.000254	103497	4	0.0000386485	0.15	P
ER004	*Bos primigenius*	27325795	214	0.000008	785376	197	0.0002508353	32.03	P
ER005	*Bos primigenius*	34673482	653	0.000019	198149	2	0.0000100934	0.54	P
ER023	*Bos primigenius*	7299394	1153	0.000158	206911	7	0.000033831	0.21	P
ER027	*Bos primigenius*	8096575	2	0	3912565	537	0.0001372501	555.63	P
ER028	*Bos primigenius*	4709631	5	0.000001	2190610	131	0.0000598007	56.33	P
CR113	*Bos primigenius*	140679	0	0	220221	222	0.0010080782	N/A	N/A
CR128	*Bos primigenius*	224087	1	0.000004	115416	103	0.0008924239	199.98	N/A

CR130	*Bison bonasus*	1012251	12	0.000012	247666	0	0	0	N/A
CR132	*Bovine*	982101	9	0.000009	108816	0	0	0	N/A
CR135	*Bovine*	1384329	35	0.000025	185801	0	0	0	N/A
CR139	*Bos primigenius*	1177651	0	0	422147	1	0.0000023688	N/A	N/A
CR150	*Bos primigenius*	2928456	15	0.000005	841944	1	0.0000011877	0.23	N/A
CR152	*Bos primigenius*	922775	29	0.000031	20739	5	0.0002410917	7.67	N/A
CR153	*Bos primigenius*	749374	0	0	525130	304	0.0005789043	N/A	N/A
CR154	*Bos primigenius*	785043	9	0.000011	151753	2	0.0000131793	1.15	N/A
ER002	*Bos primigenius*	7871369	4	0.000001	386300	1	0.0000025887	5.09	N/A
ER022	*Bovine*	13646340	23	0.000002	12	0	0	0	N/A
ER024	*Bovine*	3250000	19	0.000006	354462	0	0	0	N/A
ER026	*Bovine*	4944237	0	0	1010168	65	0.0000643457	N/A	N/A
CR115	*Bos primigenius*	17033051	305	0.000018	252325	1067	0.0042286733	236.15	K
CR108	*Bos primigenius*	705731	8	0.000011	1049520	3376	0.0032167086	283.77	E
CR109	*Bos primigenius*	875192	10	0.000011	471719	1919	0.0040680999	356.04	E
CR111	*Bos primigenius*	622424	1	0.000002	717324	700	0.0009758491	607.39	C
ROS001	*Bos primigenius*	11199858	73	0.000007	1262341	1229	0.000973588	149.37	C
ER007	*Bison bonasus*	13423508	149	0.000011	168474	0	0	0	Bb1
ER008	*Bison bonasus*	10699297	37	0.000003	889718	794	0.0008924176	258.06	Bb1

Appendix Table 7.2 - 27 ancient *Bos* and *Bison* samples downloaded NCBI. Context is reproduced from the reference. * = sample dated by context; ^ sample radiocarbon dated indirectly.

Species	Accession Number/Alias	Haplotype	Age	Location	Country	Reference
Bos primigenius	CPC98	P	6884 - 6741 cal yrBP	Carsington Pasture Cave	UK	(Park et al. 2015)
Bos primigenius	JQ437479	P	1500 cal yrBP	Pisz Forest	Poland	(Zeyland et al. 2013)
Bos primigenius	KF525852	C	10758 - 10565 cal yrBP	Kongni ditch	China	(Zhang et al. 2013)
Bison priscus	KM593920	Bp	19390 - 18940 cal yrBP	Trois-Frères	France	(Marsolier-Kergoat et al. 2015)
Bison priscus	KX898010	Bp	20200 - 14700 cal yrBP^	Igue du Gral	France	(Massilani et al. 2016)
Bison priscus	KX898014	Bp	39100 - 36300 cal yrBP^	La Berbie	Southern France	(Massilani et al. 2016)
Bison priscus	KX898018	Bp	29600 - 26000 cal yrBP^	Yakutia	Russia, East Siberia	(Massilani et al. 2016)

Bison priscus	KX898019	Bp	27300 - 26000 cal yrBP^	Yakutia	Russia, East Siberia	(Massilani et al. 2016)
Bison priscus	KX898020	Bp	~44000 - 29600 cal yrBP^	Yakutia	Russia, East Siberia	(Massilani et al. 2016)
Bison bonasus	KX553930	Bb2	1511 - 1302 cal yrBP	Styria	Austria	(Węcek et al. 2016)
Bison bonasus	KX553931	Bb2	1987 - 1886 cal yrBP	Styria	Austria	(Węcek et al. 2016)
Bison bonasus	KX553932	Bb2	1338 - 1265 cal yrBP	Upper Austria	Austria	(Węcek et al. 2016)
Bison bonasus caucasicus	KX553934	Bb2	5724 - 5657 cal yrBP	Sevan Lake region	Armenia	(Węcek et al. 2016)
Bison bonasus	KX592175	Bb2	>56300 cal yrBP	Mezmaiskaya Cave	North Caucasus	(Soubrier et al. 2016)
Bison bonasus	KX898008	Bb2	12,100 - 11,700 cal yrBP	Igue du Gral	France	(Massilani et al. 2016)
Bison bonasus	KX898009	Bb2	12400 - 11800 cal yrBP	Igue du Gral	France	(Massilani et al. 2016)
Bison bonasus	KX898011	Bb2	14300 - 13800 cal yrBP	Kesslerloch	Switzerland	(Massilani et al. 2016)
Bison bonasus	KX898012	Bb2	22500 - 22100 cal yrBP	Kudaro Cave	South Caucasus	(Massilani et al. 2016)
Bison bonasus	KX898013	Bb2	38500 - 37000 cal yrBP	Kudaro Cave	South Caucasus	(Massilani et al. 2016)
Bison bonasus	KX898017	Bb2	~51000 - 47700 cal yrBP	Mezmaiskaya Cave	North Caucasus	(Massilani et al. 2016)
Bison schoetensacki	KU886087	Bb1	36770 - 36000 cal yrBP	Siréjol cave	France	(Palacio et al. 2017)
Bison bonasus	KX592182	Bb1	>59400 cal yrBP	Mezmaiskaya Cave	North Caucasus	(Soubrier et al. 2016)
Bison bonasus	KX898005	Bb1	46700 - 34300 cal yrBP^	Aven d'Arquet	Southern France (Gard)	(Massilani et al. 2016)
Bison bonasus	KX898006	Bb1	46700 - 34300 cal yrBP^	Aven d'Arquet	Southern France (Gard)	(Massilani et al. 2016)
Bison bonasus	KX898007	Bb1	46700 - 34300 cal yrBP^	Aven d'Arquet	Southern France (Gard)	(Massilani et al. 2016)
Bison bonasus	KX898015	Bb1	~46500 - 44600 cal yrBP	Mezmaiskaya Cave	North Caucasus	(Massilani et al. 2016)
Bison bonasus	KX898016	Bb1	~47000 - 44000 cal yrBP	Mezmaiskaya Cave	North Caucasus	(Massilani et al. 2016)

Appendix Table 7.3 - 107 modern *Bovini* mitochondrial genomes downloaded from NCBI used Bayesian analysis.

Species	Common Name	NCBI Code	Haplogroup / Cluster	Country / Region	Breed	Reference
Bison bison	American Bison	NC_012346	American Bison	USA	-	(Achilli et al. 2008)
Bison bison	American Bison	GU946978	American Bison	USA	-	(Douglas et al. 2011)
Bison bison	American Bison	GU946988	American Bison	USA	-	(Douglas et al. 2011)

Bison bison	American Bison	GU946999	American Bison	USA	-	(Douglas et al. 2011)
Bison bison	American Bison	GU947006	American Bison	Canada	-	(Douglas et al. 2011)
Bison bonasus	Wisent	HQ223450	*Bb2*	Unknown	-	Direct NCBI Submission
Bison bonasus	Wisent	JN632602	*Bb2*	Poland	-	(Hassanin et al. 2012)
Bos frontalis	Gayal	MK279400	Gayal	India	-	(Prabhu et al. 2019)
Bos frontalis	Gayal	MK279401	Gayal	India	-	(Prabhu et al. 2019)
Bos gaurus	Gaur	MK770201	with Banteng	Malaysia	-	(Rosli et al. 2019)
Bos gaurus	Gaur	JN632604	with Gayal	Cambodia	-	(Hassanin et al. 2012)
Bos grunniens	Yak	MW414140	Yak	-	-	Direct NCBI Submission
Bos grunniens	Yak	MW414151	Yak	-	-	Direct NCBI Submission
Bos grunniens	Yak	MW414157	Yak	-	-	Direct NCBI Submission
Bos grunniens	Yak	MW414195	Yak	-	-	Direct NCBI Submission
Bos grunniens	Yak	MW414198	Yak	-	-	Direct NCBI Submission
Bos grunniens	Domestic Yak	GQ464250	Yak	Tibetan-Qinghai Plateau	-	(Wang et al. 2010)
Bos grunniens	Domestic Yak	GQ464260	Yak	Tibetan-Qinghai Plateau	-	(Wang et al. 2010)
Bos grunniens	Domestic Yak	GQ464266	Yak	Tibetan-Qinghai Plateau	-	(Wang et al. 2010)
Bos grunniens	Domestic Yak	GQ464267	Yak	Tibetan-Qinghai Plateau	-	(Wang et al. 2010)
Bos grunniens	Domestic Yak	GQ464276	Yak	Tibetan-Qinghai Plateau	-	(Wang et al. 2010)
Bos grunniens	Domestic Yak	GQ464284	Yak	Tibetan-Qinghai Plateau	-	(Wang et al. 2010)
Bos grunniens	Domestic Yak	GQ464302	Yak	Tibetan-Qinghai Plateau	-	(Wang et al. 2010)
Bos grunniens	Domestic Yak	GQ464311	Yak	Tibetan-Qinghai Plateau	-	(Wang et al. 2010)
Bos indicus	Zebu/Indicine	AY126697	I1	India	Nellore	Direct NCBI Submission
Bos indicus	Zebu/Indicine	EU177868	I1	Iraq	Iraqi	(Achilli et al. 2008)
Bos indicus	Zebu/Indicine	FJ971088	I1	Mongolia	Unspecified	(Achilli et al. 2009)
Bos indicus	Zebu/Indicine	GU256940	I1	China	Diqin	Direct NCBI Submission
Bos indicus	Zebu/Indicine	KP143771	I1	China	Simmental x local breed	Direct NCBI Submission
Bos indicus	Zebu/Indicine	NC_005971	I1	India	Nellore	Direct NCBI Submission
Bos indicus	Zebu/Indicine	AF492350	I2	Sri Lanka	Zwergzebu	(Hiendleder, Lewalski, and Janke 2008)
Bos indicus	Zebu/Indicine	EU177869	I2	Iraq	Iraqi	(Achilli et al. 2008)
Bos javanicus	Banteng	AB915322	with Gayal	Borneo	-	Modern

Bos javanicus	Banteng	JN632605	*with Gaur*	Cambodia	-	(Hassanin et al. 2012)
Bos javanicus	Banteng	JN632606	*no cluster*	Java	-	(Hassanin et al. 2012)
Bos mutus	wild Yak	KY829451	*Yak*	Tibetan-Qinghai Plateau	-	(Wu et al. 2018)
Bos mutus	wild Yak	MK033130	*Yak*	China	-	(Zhang et al. 2019)
Bos mutus	wild Yak	NC_025563	*Yak*	China	-	(Na et al. 2016)
Bos mutus	Wild Yak	KR106993	*Yak*	China	-	(Chunnian et al. 2016)
Bos mutus hybrid	Yakow	KM233417	*Yak*	China	-	(Na et al. 2016)
Bos taurus	Taurine	DQ124389	P	Korea	Beef cattle	Direct NCBI Submission
Bos taurus	Taurine	LC537308	P	Japan	Japanese Shorthorn	(Mannen et al. 2020)
Bos taurus	Taurine	LC537309	P	Japan	Japanese Shorthorn	(Mannen et al. 2020)
Bos taurus	Taurine	LC537310	P	Japan	Japanese Shorthorn	(Mannen et al. 2020)
Bos taurus	Taurine	LC537311	P	Japan	Japanese Shorthorn	(Mannen et al. 2020)
Bos taurus	Taurine	LC537312	P	Japan	Japanese Shorthorn	(Mannen et al. 2020)
Bos taurus	Taurine	LC537313	P	Japan	Japanese Shorthorn	(Mannen et al. 2020)
Bos taurus	Taurine	LC537314	P	Japan	Japanese Shorthorn	(Mannen et al. 2020)
Bos taurus	Taurine	LC537315	P	Japan	Japanese Shorthorn	(Mannen et al. 2020)
Bos taurus	Taurine	LC537316	P	Japan	Japanese Shorthorn	(Mannen et al. 2020)
Bos taurus	Taurine	LC537317	P	Japan	Japanese Shorthorn	(Mannen et al. 2020)
Bos taurus	Taurine	MZ901408	P	Austria	Murbodner	(Cubric-Curik et al. 2021)
Bos taurus	Taurine	EU177866	Q	Italy	Cabannina	(Achilli et al. 2008)
Bos taurus	Taurine	EU177867	Q	Italy	Cabannina	(Achilli et al. 2008)
Bos taurus	Taurine	FJ971080	Q	Italy	Romagnola	(Achilli et al. 2009)
Bos taurus	Taurine	FJ971082	Q	Italy	Italian Red Pied	(Achilli et al. 2009)
Bos taurus	Taurine	FJ971083	Q	Italy	Romagnola	(Achilli et al. 2009)
Bos taurus	Taurine	HQ184034	Q	Italy	Romagnola	(Bonfiglio et al. 2010)
Bos taurus	Taurine	HQ184035	Q	Italy	Romagnola	(Bonfiglio et al. 2010)
Bos taurus	Taurine	HQ184036	Q	Italy	Grauvieh (Grey Alpine)	(Bonfiglio et al. 2010)
Bos taurus	Taurine	HQ184037	Q	Italy	Grauvieh (Grey Alpine)	(Bonfiglio et al. 2010)
Bos taurus	Taurine	HQ184038	Q	Italy	Cabannina	(Bonfiglio et al. 2010)
Bos taurus	Taurine	HQ184039	Q	Italy	Chianina	(Bonfiglio et al. 2010)
Bos taurus	Taurine	KT184471	Q	Egypt	Domiaty	(Olivieri et al. 2015)

Bos taurus	Taurine	KT184472	Q	Egypt	Domiaty	(Olivieri et al. 2015)
Bos taurus	Taurine	MZ901639	Q1	Greece	Katerini	(Cubric-Curik et al. 2021)
Bos taurus	Taurine	MZ901646	Q1	Greece	Greece Shorthorn Pindos	(Cubric-Curik et al. 2021)
Bos taurus	Taurine	FJ971081	Q2	Italy	Chianina	(Achilli et al. 2009)
Bos taurus	Taurine	HQ184030	Q2	Italy	Chianina	(Bonfiglio et al. 2010)
Bos taurus	Taurine	FJ971084	R1	Italy	Agerolese	(Achilli et al. 2009)
Bos taurus	Taurine	FJ971085	R1	Italy	Cinisara	(Achilli et al. 2009)
Bos taurus	Taurine	FJ971086	R1	Italy	Cinisara	(Achilli et al. 2009)
Bos taurus	Taurine	HQ184040	R1	Italy	Romagnola	(Bonfiglio et al. 2010)
Bos taurus	Taurine	HQ184041	R1	Italy	Romagnola	(Bonfiglio et al. 2010)
Bos taurus	Taurine	HQ184042	R1	Italy	Romagnola	(Bonfiglio et al. 2010)
Bos taurus	Taurine	HQ184043	R1	Italy	Romagnola	(Bonfiglio et al. 2010)
Bos taurus	Taurine	HQ184044	R1	Italy	Romagnola	(Bonfiglio et al. 2010)
Bos taurus	Taurine	FJ971087	R2	Italy	Romagnola	(Achilli et al. 2009)
Bos taurus	Taurine	HQ184045	R2	Italy	Marchigiana	(Bonfiglio et al. 2010)
Bos taurus	Taurine	JN817304	T1	Ethiopia	Arsi	(Bonfiglio et al. 2012)
Bos taurus	Taurine	JN817321	T1	Egypt	Domiaty	(Bonfiglio et al. 2012)
Bos taurus	Taurine	KF163061	T1	South Africa	Nguni	(Horsburgh et al. 2013)
Bos taurus	Taurine	MZ901385	T1	Portugal	Barrosa	(Cubric-Curik et al. 2021)
Bos taurus	Taurine	MZ901462	T1	Croatia	Croatia Buša	(Cubric-Curik et al. 2021)
Bos taurus	Taurine	MZ901654	T1	Greece	Greece Shorthorn Rodopi	(Cubric-Curik et al. 2021)
Bos taurus	Taurine	MZ901703	T1	North Macedonia	Macedonian Buša	(Cubric-Curik et al. 2021)
Bos taurus	Taurine	MZ901377	T2	Turkey	Anatolian Black	(Cubric-Curik et al. 2021)
Bos taurus	Taurine	MZ901441	T2	Croatia	Croatia Buša	(Cubric-Curik et al. 2021)
Bos taurus	Taurine	MZ901640	T2	Greece	Katerini	(Cubric-Curik et al. 2021)
Bos taurus	Taurine	MZ901693	T2	Greece	Kea Rind	(Cubric-Curik et al. 2021)
Bos taurus	Taurine	MZ901758	T2	Mongolia	Turano-Mongolian	(Cubric-Curik et al. 2021)
Bos taurus	Taurine	KT184451	T3	Egypt	Menofi	(Olivieri et al. 2015)
Bos taurus	Taurine	KT184452	T3	Egypt	Domiaty	(Olivieri et al. 2015)
Bos taurus	Taurine	V00654	T3	UK	Unspecified	(Anderson et al. 1982)

Bos taurus	Taurine	AB074963	T4	Japan	Japanese Black	(Mannen et al. 2003)
Bos taurus	Taurine	DQ124372	T4	Korea	Korean cattle	Direct NCBI Submission
Bos taurus	Taurine	DQ124375	T4	Korea	Korean cattle	Direct NCBI Submission
Bos taurus	Taurine	DQ124377	T4	Korea	Korean cattle	Direct NCBI Submission
Bos taurus	Taurine	DQ124412	T4	Netherlands	Holstein	Direct NCBI Submission
Bos taurus	Taurine	EU177862	T5	Italy	Valdostana	(Achilli et al. 2008)
Bos taurus	Taurine	EU177863	T5	Italy	Piedmontese	(Achilli et al. 2008)
Bos taurus	Taurine	MZ901530	T5	Croatia	Holstein	(Cubric-Curik et al. 2021)
Bos taurus	Taurine	MZ901633	T5	Croatia	Slavonian Syrmian Podolian	(Cubric-Curik et al. 2021)
Bos taurus	Taurine	MZ901740	T5	Albania	Skodra Buša	(Cubric-Curik et al. 2021)
Bos taurus	Taurine	AY676867	T6	UK	Angus X	Direct NCBI Submission
Bos taurus	Taurine	AY676870	T6	UK	Angus X	Direct NCBI Submission
Bos taurus	Taurine	EU177840	T7	Italy	Cabannina	(Achilli et al. 2008)

Appendix Table 7.4 - Datasets of *Bovini* mitochondrial sequences curated for BEAST analysis. BEAST calibration refers to the tip date of ancient samples. Calibrations in brackets signifies that the tip date was supplied as a probability distribution due to age uncertainty.

Sample	Species	Location	Context	BEAST calibration	Dataset used
AB074963	*Bos taurus*	Japan	modern	0	01,02,03
AB915322	*Bos javanicus*	Borneo	modern	0	01
Ace1-MV135	*Bos taurus*	Turkey	4923 - 4323 cal yrBP^	(4300.0,4900.0)	01,02,03
AF492350	*Bos indicus*	Sri Lanka	modern	0	01,02,03
AY126697	*Bos indicus*	India	modern	0	01,02,03
AY676867	*Bos taurus*	UK	modern	0	01,02,03
AY676870	*Bos taurus*	UK	modern	0	01,02,03
Baikal1-MV189	*Bos primigenius*	Russia	12015 - 12529 cal yrBP	12271	01,02,03
Bar1-CR119	*Bos primigenius*	Italy	13246 - 13005 cal yrBP	13125	01,02,03,04
Bed3	*Bos primigenius*	Germany	11875 - 11399 cal yrBP	11636	01,02,03,04
Bed4	*Bos primigenius*	Germany	11949 - 11476 cal yrBP	11673	01,02,03,04
Belarus1-MV197	*Bos primigenius*	Belarus	4476 - 4173 cal yrBP	4324	01,02,03,04
Bes1-MV047	*Bos taurus x Bos indicus*	Iraqi Kurdistan	2934 - 2632 yrBP*	(2934.0,2632.0)	01,02,03
Bes2-MV048	*Bos taurus x Bos indicus*	Iraqi Kurdistan	624 - 731 cal yrBP	677	01,02,03
Blec1-CR118	*Bos primigenius*	Bulgaria	6258 - 6006 cal yrBP^	6131	01,02,03,04

Bor1-CR100	*Bos primigenius*	Kazakhstan	6740 - 6521 cal yrBP	6629	01,02,03
Bor2-CR101	*Bos primigenius*	Kazakhstan	4913 - 5099 cal yrBP	5004	01,02,03
Bor3-CR102	*Bos primigenius*	Kazakhstan	5115 - 4925 cal yrBP	5019	01,02,03,04
Ch22	*Bos primigenius*	Turkey	7750 - 7600 cal yrBP^	(7600.0,7750.0)	01,02,03
CPC98	*Bos primigenius*	UK	6884 - 6741 cal yrBP	6738	01,02,03,04
DQ124372	*Bos taurus*	Korea	modern	0	01,02,03
DQ124375	*Bos taurus*	Korea	modern	0	01,02,03
DQ124377	*Bos taurus*	Korea	modern	0	01,02,03
DQ124389	*Bos taurus*	Korea	modern	0	01,02,03
DQ124412	*Bos taurus*	Netherlands	modern	0	01,02,03
EU177840	*Bos taurus*	Italy	modern	0	01,02,03
EU177862	*Bos taurus*	Italy	modern	0	01,02,03
EU177863	*Bos taurus*	Italy	modern	0	01,02,03
EU177866	*Bos taurus*	Italy	modern	0	01,02,03
EU177867	*Bos taurus*	Italy	modern	0	01,02,03
EU177868	*Bos indicus*	Iraq	modern	0	01,02,03
EU177869	*Bos indicus*	Iraq	modern	0	01,02,03
Fal7	*Bos primigenius*	Germany	1000 yrBP*	(0.0,3000.0)	01,02,03,04
Fif1-CR166	*Bos primigenius*	Scotland	10634 - 10277 cal yrBP	10455	01,02,03,04
FJ971080	*Bos taurus*	Italy	modern	0	01,02,03
FJ971081	*Bos taurus*	Italy	modern	0	01,02,03
FJ971082	*Bos taurus*	Italy	modern	0	01,02,03
FJ971083	*Bos taurus*	Italy	modern	0	01,02,03
FJ971084	*Bos taurus*	Italy	modern	0	01,02,03
FJ971085	*Bos taurus*	Italy	modern	0	01,02,03
FJ971086	*Bos taurus*	Italy	modern	0	01,02,03
FJ971087	*Bos taurus*	Italy	modern	0	01,02,03
FJ971088	*Bos indicus*	Mongolia	modern	0	01,02,03
Fre1-CR173	*Bos primigenius*	Denmark	10825 - 10496 cal yrBP	10660	01,02,03,04
Gal1-CR120	*Bos primigenius*	Spain	9548 - 9113 cal yrBP	9330	01,02,03,04
Gal2-CR170	*Bos primigenius*	Spain	9481 - 9100 cal yrBP	9290	01,02,03,04
Gal3-CR176	*Bos primigenius*	Spain	9538 - 9109 cal yrBP	9323	01,02,03,04
Gallo1-CR165	*Bos primigenius*	Scotland	3764 - 3479 cal yrBP	3621	01,02,03,04

GQ464250	*Bos grunniens*	Tibetan-Qinghai Plateau	modern	0	01
GQ464260	*Bos grunniens*	Tibetan-Qinghai Plateau	modern	0	01
GQ464266	*Bos grunniens*	Tibetan-Qinghai Plateau	modern	0	01
GQ464267	*Bos grunniens*	Tibetan-Qinghai Plateau	modern	0	01
GQ464276	*Bos grunniens*	Tibetan-Qinghai Plateau	modern	0	01
GQ464284	*Bos grunniens*	Tibetan-Qinghai Plateau	modern	0	01
GQ464302	*Bos grunniens*	Tibetan-Qinghai Plateau	modern	0	01
GQ464311	*Bos grunniens*	Tibetan-Qinghai Plateau	modern	0	01
GU256940	*Bos indicus*	China	modern	0	01,02,03
GU946978	*Bison bison*	USA	modern	0	01
GU946988	*Bison bison*	USA	modern	0	01
GU946999	*Bison bison*	USA	modern	0	01
GU947006	*Bison bison*	Canada	modern	0	01
Gyu1-MV198	*Bos primigenius*	Armenia	14097 - 13863 cal yrBP	13979	01,02,03
Gyumri2-MV200	*Bos primigenius*	Armenia	7394 - 7239 cal yrBP	7316	01,02,03
Has3-MV169	*Bos taurus*	Iran	4470 cal BP^	(4200.0,4700.0)	01,02,03
Hjo1-CR174	*Bos primigenius*	Denmark	8245 - 8053 cal yrBP	8148	01,02,03,04
Horn_10537	*Bos primigenius*	Norway	400 yrBP*	700	01,02,03,04
Horn_10543	*Bos primigenius*	Unknown	600 yrBP*	600	01,02,03,04
Horn_11570	*Bos primigenius*	Poland	700 yrBP*	400	01,02,03
Horn_382	*Bos primigenius*	Denmark	600 yrBP*	600	01,02,03
Horn_617	*Bos primigenius*	Norway	600 yrBP*	600	01,02,03
Horn_CMLIX	*Bos primigenius*	Germany	400 yrBP*	400	01,02,03,04
HQ184030	*Bos taurus*	Italy	modern	0	01,02,03
HQ184034	*Bos taurus*	Italy	modern	0	01,02,03
HQ184035	*Bos taurus*	Italy	modern	0	01,02,03
HQ184036	*Bos taurus*	Italy	modern	0	01,02,03
HQ184037	*Bos taurus*	Italy	modern	0	01,02,03
HQ184038	*Bos taurus*	Italy	modern	0	01,02,03
HQ184039	*Bos taurus*	Italy	modern	0	01,02,03
HQ184040	*Bos taurus*	Italy	modern	0	01,02,03

HQ184041	*Bos taurus*	Italy	modern	0	01,02,03
HQ184042	*Bos taurus*	Italy	modern	0	01,02,03
HQ184043	*Bos taurus*	Italy	modern	0	01,02,03
HQ184044	*Bos taurus*	Italy	modern	0	01,02,03
HQ184045	*Bos taurus*	Italy	modern	0	01,02,03
HQ223450	*Bison bonasus*	Unknown	modern	0	01
Hxh2	*Bos primigenius*	Germany	5000 yrBP*	(3000.0,9000.0)	01,02,03,04
JN632602	*Bison bonasus*	Poland	modern	0	01
JN632604	*Bos gaurus*	Cambodia	modern	0	01
JN632605	*Bos javanicus*	Cambodia	modern	0	01
JN632606	*Bos javanicus*	Java	modern	0	01
JN817304	*Bos taurus*	Ethiopia	modern	0	01,02,03
JN817321	*Bos taurus*	Egypt	modern	0	01,02,03
JQ437479	*Bos primigenius*	Poland	~1500 cal yrBP	1500	01,02,03,04
Kalisz1-MV192	*Bos primigenius*	Poland	4487 - 4229 cal yrBP	4357	01,02,03,04
Kalisz2-MV194	*Bos primigenius*	Poland	5813 - 5668 cal yrBP	5740	01,02,03,04
Kalisz3-MV195	*Bos primigenius*	Poland	4477 - 4168 cal yrBP	4322	01,02,03
KF163061	*Bos taurus*	South Africa	modern	0	01,02,03
KM233417	*Bos mutus hybrid*	China	modern	0	01
KM593920	*Bison priscus*	France	19390 - 18940 cal yrBP	(18940.0,19390.0)	01
Kok1-MV239	*Bos primigenius*	Uzbekistan	5000-3500 yrBP*	(1500.0,7000.0)	01,02,03
KP143771	*Bos indicus*	China	modern	0	01,02,03
KR106993	*Bos mutus*	China	modern	0	01
KT184451	*Bos taurus*	Egypt	modern	0	01,02,03
KT184452	*Bos taurus*	Egypt	modern	0	01,02,03
KT184471	*Bos taurus*	Egypt	modern	0	01,02,03
KT184472	*Bos taurus*	Egypt	modern	0	01,02,03
KU886087	*Bison schoetensacki*	France	36770 - 36000 cal yrBP	(36000.0,36770.0)	01
KX553930	*Bison bonasus*	Austria	1511 - 1302 cal yrBP	(1302.0,1511.0)	01
KX553931	*Bison bonasus*	Austria	1987 - 1886 cal yrBP	(1886.0,1987.0)	01
KX553932	*Bison bonasus*	Austria	1338 - 1265 cal yrBP	(1265.0,1338.0)	01
KX553934	*Bison bonasus caucasicus*	Armenia	5724 - 5657 cal yrBP	(5657.0,5724.0)	01
KX592175	*Bison bonasus*	North Caucasus	>56300 cal yrBP	(56300.0,200000.0)	01

KX592182	*Bison bonasus*	North Caucasus	>59400 cal yrBP	(59400.0,200000.0)	01
KX898005	*Bison bonasus*	France	46700 - 34300 cal yrBP^	(34300.0,46700.0)	01
KX898006	*Bison bonasus*	France	46700 - 34300 cal yrBP^	(34300.0,46700.0)	01
KX898007	*Bison bonasus*	France	46700 - 34300 cal yrBP^	(34300.0,46700.0)	01
KX898008	*Bison bonasus*	France	12,100 - 11,700 cal yrBP	(11700.0,12100.0)	01
KX898009	*Bison bonasus*	France	12400 - 11800 cal yrBP	(11800.0,12400.0)	01
KX898010	*Bison priscus*	France	20200 - 14700 cal yrBP^	(14700.0,20200.0)	01
KX898011	*Bison bonasus*	Switzerland	14300 - 13800 cal yrBP	(13800.0,14300.0)	01
KX898012	*Bison bonasus*	South Caucasus	22500 - 22100 cal yrBP	(22100.0,22500.0)	01
KX898013	*Bison bonasus*	South Caucasus	38500 - 37000 cal yrBP	(37000.0,38500.0)	01
KX898014	*Bison priscus*	Southern France	39100 - 36300 cal yrBP^	(36300.0,39100.0)	01
KX898015	*Bison bonasus*	North Caucasus	~46500 - 44600 cal yrBP	(44600.0,46500.0)	01
KX898016	*Bison bonasus*	North Caucasus	~47000 - 44000 cal yrBP	(44000.0,47000.0)	01
KX898017	*Bison bonasus*	North Caucasus	~51000 - 47700 cal yrBP	(47700.0,51000.0)	01
KX898018	*Bison priscus*	Russia, East Siberia	29600 - 26000 cal yrBP^	(26000.0,29600.0)	01
KX898019	*Bison priscus*	Russia, East Siberia	27300 - 26000 cal yrBP^	(26000.0,27300.0)	01
KX898020	*Bison priscus*	Russia, East Siberia	~44000 - 29600 cal yrBP^	(29600.0,44000.0)	01
KY829451	*Bos mutus*	Tibetan-Qinghai Plateau	modern	0	01
LC537308	*Bos taurus*	Japan	modern	0	01,02,03
LC537309	*Bos taurus*	Japan	modern	0	01,02,03
LC537310	*Bos taurus*	Japan	modern	0	01,02,03
LC537311	*Bos taurus*	Japan	modern	0	01,02,03
LC537312	*Bos taurus*	Japan	modern	0	01,02,03
LC537313	*Bos taurus*	Japan	modern	0	01,02,03
LC537314	*Bos taurus*	Japan	modern	0	01,02,03
LC537315	*Bos taurus*	Japan	modern	0	01,02,03
LC537316	*Bos taurus*	Japan	modern	0	01,02,03
LC537317	*Bos taurus*	Japan	modern	0	01,02,03
Lil1-CR171	*Bos primigenius*	Denmark	4484 - 4228 cal yrBP	4355	01,02,03,04
Lun1-CR178	*Bos primigenius*	Sweden	9744-9525 / 9045-8656 cal yrBP^	(8656.0, 9744.0)	01,02,03,04
Lun2-CR179	*Bos primigenius*	Sweden	9744-9525 / 9045-8656 cal yrBP^	(8656.0, 9744.0)	01,02,03,04
Men2-MV013	*Bos taurus*	Turkey	7998 - 7843 cal yrBP	4355	01,02,03

MK033130	*Bos mutus*	China	modern	0	01
MK279400	*Bos frontalis*	India	modern	0	01
MK279401	*Bos frontalis*	India	modern	0	01
MK770201	*Bos gaurus*	Malaysia	modern	0	01
MW414140	*Bos grunniens*	-	modern	0	01
MW414151	*Bos grunniens*	-	modern	0	01
MW414157	*Bos grunniens*	-	modern	0	01
MW414195	*Bos grunniens*	-	modern	0	01
MW414198	*Bos grunniens*	-	modern	0	01
MZ901377	*Bos taurus*	Turkey	modern	0	01,02,03
MZ901385	*Bos taurus*	Portugal	modern	0	01,02,03
MZ901408	*Bos taurus*	Austria	modern	0	01,02,03
MZ901441	*Bos taurus*	Croatia	modern	0	01,02,03
MZ901462	*Bos taurus*	Croatia	modern	0	01,02,03
MZ901530	*Bos taurus*	Croatia	modern	0	01,02,03
MZ901633	*Bos taurus*	Croatia	modern	0	01,02,03
MZ901639	*Bos taurus*	Greece	modern	0	01,02,03
MZ901640	*Bos taurus*	Greece	modern	0	01,02,03
MZ901646	*Bos taurus*	Greece	modern	0	01,02,03
MZ901654	*Bos taurus*	Greece	modern	0	01,02,03
MZ901693	*Bos taurus*	Greece	modern	0	01,02,03
MZ901703	*Bos taurus*	North Macedonia	modern	0	01,02,03
MZ901740	*Bos taurus*	Albania	modern	0	01,02,03
MZ901758	*Bos taurus*	Mongolia	modern	0	01,02,03
NC_005971	*Bos indicus*	India	modern	0	01,02,03
NC_012346	*Bison bison*	USA	modern	0	01
NC_025563	*Bos mutus*	China	modern	0	01
Pad1-CR121	*Bos primigenius*	Italy	12267 - 11844 cal yrBP	12056	01,02,03,04
Pal1-CR125	*Bos primigenius*	Italy	16287 - 15849 cal yrBP	16068	01,02,03,04
Plo1-MV136	*Bos taurus*	Serbia	7275/7125 - 7100/7000 cal yrBP^	(7000.0,7275.0)	01,02,03
Plo2-MV137	*Bos taurus*	Serbia	7025/6950 - 6850 - 6650 cal yrBP^	(6650.0,7025.0)	01,02,03
Plo3-MV138	*Bos taurus*	Serbia	7100/7000 - 7025/6950 cal yrBP^	(6950.0,7100.0)	01,02,03
Plo4-MV139	*Bos taurus*	Serbia	7100/7000 - 7025/6950 cal yrBP^	(6950.0,7100.0)	01,02,03

Plo7-MV160	*Bos taurus*	Serbia	7100/7000 - 7025/6950 cal yrBP^	(6950.0,7100.0)	01,02,03
Plo8-MV162	*Bos taurus*	Serbia	7100/7000 - 7025/6950 cal yrBP^	(6950.0,7100.0)	01,02,03
Qaz1-MV172	*Bos taurus*	Iran	4905 - 4659 cal yrBP	4781	01,02,03
Rhi1-CR180	*Bos primigenius*	Germany	47,000 cal yrBP+	(40000.0,150000.0)	01,02,03
Rhi2-CR181	*Bos primigenius*	Germany	47009 - 45312 cal yrBP	46,160	01,02,03
Rhi3-CR182	*Bos primigenius*	Germany	Pleistocene*	(10000.0,150000.0)	01,02,03,04
Rhi4-CR183	*Bos primigenius*	Germany	Pleistocene*	(10000.0,150000.0)	01,02
Rhi5-CR184	*Bos primigenius*	Germany	Pleistocene*	(10000.0,150000.0)	01,02
Rhi6-CR185	*Bos primigenius*	Germany	Pleistocene*	(10000.0,150000.0)	01,02
Rhi7-CR186	*Bos primigenius*	Germany	Pleistocene*	(10000.0,150000.0)	01,02
Rhi8-CR187	*Bos primigenius*	Germany	Pleistocene*	(10000.0,150000.0)	01,02
ROS002	*Bos primigenius*	Kazakhstan	5116 - 4926 cal yrBP	5020	01,02,03
Shau1-CR103	*Bos taurus*	Kazakhstan	Early Bronze Age*	(3000.0,6000.0)	01,02,03
Shi1-CR104	*Bos taurus*	Kazakhstan	3983 - 3772 cal yrBP	3878	01,02,03
Siv1-CR117	*Bos primigenius*	Bulgaria	6258 - 6006 cal yrBP	6131	01,02,03,04
Ska1-CR167	*Bos primigenius*	Sweden	9610 - 9207 cal yrBP	9408	01,02,03,04
Ska3-CR169	*Bos primigenius*	Sweden	9731 - 9508 cal yrBP	9619	01,02,03,04
Tango1-MV198	*Bos primigenius*	Former Soviet Union	4142 - 3914 cal yrBP	4027	01,02,03
Tango2-MV199	*Bos primigenius*	Former Soviet Union	4599 - 4379 cal yrBP	4488	01,02,03
Th7	*Bos primigenius*	Morocco	9073 - 8730 cal yrBP	8901	01,02,03
Ton1-CR172	*Bos primigenius*	Denmark	3279 - 3074 cal yrBP	3176	01,02,03,04
Tri1	*Bos primigenius*	Germany	10291 - 9976 cal yrBP	10133	01,02,03,04
Tula1-MV194	*Bos primigenius*	Russia	40746 - 42460 cal yrBP	41602	01,02,03
UKR07-CR107	*Bos primigenius*	Ukraine	9000 - 8000 yrBP*	(3000.0,13000.0)	01,02,03,04
UKR08-CR108	*Bos primigenius*	Ukraine	9000 - 8000 yrBP*	(3000.0,13000.0)	01,02,03
UKR11-CR109	*Bos primigenius*	Ukraine	9000 - 8000 yrBP*	(3000.0,13000.0)	01,02,03
Uzz1-CR126	*Bos primigenius*	Italy	7593 - 7480 cal yrBP	7537	01,02,03,04
V00654	*Bos taurus*	UK	modern	0	01,02,03
Var1-CR110	*Bos primigenius*	Kazakhstan	7641 - 7500 cal yrBP	7570	01,02,03
Var2-CR116	*Bos primigenius*	Kazakhstan	7486 - 7238 cal yrBP	7361	01,02
VEM212	*Bos primigenius*	Britain	MIS 7 (by stratigraphy)	(50000.0,350000.0)	01,02
Yen1	*Bos primigenius*	Russia	14193 - 14019 cal yrBP	14106	01,02,03

Yer1-MV240	*Bos taurus x Bos indicus*	Uzbekistan	2100 yrBP*	(0.0,4000.0)	01,02,03
YoA	*Bos primigenius*	England	4909 - 4658 cal yrBP	4783	01,02,03

Appendix Table 8.1 - Modern bulls used in Y chromosome analysis.

NCBI Accession	Breed	Region	Cattle Type	Mean Depth Post Filtering	Haplotype
CL00006972	Achai	Pakistan/Afghanistan	Asian indicine	12.01	Y3
ERR3305588	Alentejana	Portugal	C. and N. European taurine	26.88	Y2a-Eu
SRR1425124	Angus	Scotland	C. and N. European taurine	14.01	Y1-modern
SRR1365144	Angus	Scotland	C. and N. European taurine	13.66	Y1-modern
SRR1365129	Angus	Scotland	C. and N. European taurine	13.31	Y1-modern
SRR1355237	Angus	Scotland	C. and N. European taurine	12.05	Y1-modern
SRR1425145	Angus	Scotland	C. and N. European taurine	10.92	Y1-modern
ERR4472515	Baoule	Ivory Coast	African taurine	25.77	Y2a-Af
SAMN06699039	Bashan	China (north-central)	Asian hybrid	9.82	Y2b-east
SAMN06699037	Bashan	China (north-central)	Asian hybrid	9.51	Y3
SAMN06699041	Bashan	China (north-central)	Asian hybrid	9.24	Y2b-east
SAMN06699038	Bashan	China (north-central)	Asian hybrid	8.31	Y3
SAMN06699040	Bashan	China (north-central)	Asian hybrid	8.15	Y2b-east
SAMN01915355	Blonde d'Aquitaine	France	C. and N. European taurine	6.21	Y2a-Eu
SAMN06699021	Bohai Black	China (north-central)	Asian hybrid	9.09	Y2a-Eu
SAMN06699022	Bohai Black	China (north-central)	Asian hybrid	8.78	Y2a-Eu
SAMN06699019	Bohai Black	China (north-central)	Asian hybrid	7.95	Y3
SAMN06699020	Bohai Black	China (north-central)	Asian hybrid	7.56	Y2a-Eu
SRR8426534	Boskarin	Croatia	Southern European	8.09	Y2a-Eu
ERR1766310	Brown Swiss	Switzerland	C. and N. European taurine	14.40	Y2a-Eu
SRR5507272	Chaidamu Yellow Cattle	China (north)	Asian taurine	10.67	Y2a-Eu
SRR5507271	Chaidamu Yellow Cattle	China (north)	Asian taurine	8.77	Y2a-Eu
SRR1355258	Charolais	France	C. and N. European taurine	14.10	Y2a-Eu
SRR1348577	Charolais	France	C. and N. European taurine	9.41	Y1-modern
SRR1343169	Charolais	France	C. and N. European taurine	8.84	Y2a-Eu
SRR1348574	Charolais	France	C. and N. European taurine	8.80	Y1-modern

SRR1348584	Charolais	France	C. and N. European taurine	8.73	Y2a-Eu
SRR1355254	Charolais	France	C. and N. European taurine	8.55	Y2a-Eu
SAMN05216091	Chianina	Italy	Southern European	11.29	Y2a-Eu
CL100006973	Cholistani	Pakistan	Asian indicine	14.13	Y3
SAMN05803883	Dexter	Ireland	C. and N. European taurine	14.32	Y1-modern
CL100012095	Dhanni	Pakistan	Asian indicine	17.01	Y3
CL100007106	Gabrali	Pakistan/Afghanistan	Asian indicine	15.55	Y3
SRR1365112	Gelbvieh	Germany	C. and N. European taurine	12.21	Y2a-Eu
SRR1355236	Gelbvieh	Germany	C. and N. European taurine	12.10	Y2a-Eu
SRR1348575	Gelbvieh	Germany	C. and N. European taurine	10.07	Y2a-Eu
SRR1346388	Gelbvieh	Germany	C. and N. European taurine	9.43	Y2a-Eu
SRR1343166	Gelbvieh	Germany	C. and N. European taurine	7.37	Y2a-Eu
SRR2016752	Gir	India	South Asian indicine	11.26	Y3
SRR2016754	Gir	India	South Asian indicine	8.64	Y3
SRR2016753	Gir	India	South Asian indicine	2.47	Y3
SRR934432	Hanwoo	Korea	Asian taurine	10.80	Y2b-east
SRR934418	Hanwoo	Korea	Asian taurine	10.21	Y2b-east
SRR934415	Hanwoo	Korea	Asian taurine	10.19	Y2b-east
SRR934419	Hanwoo	Korea	Asian taurine	9.93	Y2b-east
SRR934417	Hanwoo	Korea	Asian taurine	9.82	Y2b-east
SRR1365131	Hereford	England	C. and N. European taurine	13.92	Y1-modern
SRR1365126	Hereford	England	C. and N. European taurine	13.48	Y1-modern
SRR1365124	Hereford	England	C. and N. European taurine	13.27	Y1-modern
SRR1365128	Hereford	England	C. and N. European taurine	13.10	Y1-modern
SRR1365137	Hereford	England	C. and N. European taurine	10.98	Y1-modern
SRR1348583	Holstein	Netherlands	C. and N. European taurine	18.57	Y1-modern
SRR1425144	Holstein	Netherlands	C. and N. European taurine	14.22	Y1-modern
SRR1425133	Holstein	Netherlands	C. and N. European taurine	13.68	Y1-modern
SRR1425129	Holstein	Netherlands	C. and N. European taurine	13.64	Y1-modern
SRR1346386	Holstein	Netherlands	C. and N. European taurine	10.31	Y1-modern
SRR3497161	Jersey	Jersey	C. and N. European taurine	12.46	Y2b-east

SRR3497466	Jersey	Jersey	C. and N. European taurine	12.19	Y2b-east
SRR3497611	Jersey	Jersey	C. and N. European taurine	11.49	Y2b-east
SRR3497462	Jersey	Jersey	C. and N. European taurine	10.51	Y2b-east
SRR5507282	Jian	China (Southeast)	Asian indicine	9.84	Y3
SRR5507281	Jian	China (Southeast)	Asian indicine	8.87	Y3
SRR5507285	Jiangxi	China (Southeast)	Asian indicine	9.11	Y3
SRR5507286	Jiangxi	China (Southeast)	Asian indicine	8.31	Y3
SRR5507284	Jiangxi	China (Southeast)	Asian indicine	8.22	Y3
SAMN06699017	Jiaxian Red	China (north-central)	Asian hybrid	11.99	Y3
SAMN06699015	Jiaxian Red	China (north-central)	Asian hybrid	10.44	Y2b-east
SAMN06699014	Jiaxian Red	China (north-central)	Asian hybrid	9.63	Y2b-east
SAMN06699018	Jiaxian Red	China (north-central)	Asian hybrid	9.56	Y2b-east
SAMN06699016	Jiaxian Red	China (north-central)	Asian hybrid	9.45	Y2b-east
SAMN06698986	Kazakh	China (northwest)	Asian taurine	11.20	Y2a-Eu
SAMN06698990	Kazakh	China (northwest)	Asian taurine	9.74	Y2a-Eu
SAMN06698989	Kazakh	China (northwest)	Asian taurine	9.21	Y2a-Eu
SAMN06698991	Kazakh	China (northwest)	Asian taurine	8.90	Y2a-Eu
SAMN06698992	Kazakh	China (northwest)	Asian taurine	8.48	Y2a-Eu
SAMD00013963	Kuchinoshima	Japan	Asian taurine	12.96	Y2a-Eu
ERR3293557	Lagune	West/Central Africa	African taurine	20.74	Y2a-Af
SRR8426539	Limia	Spain	Southern European	7.89	Y2a-Eu
SRR1365134	Limousin	France	C. and N. European taurine	10.36	Y2a-Eu
SRR1365136	Limousin	France	C. and N. European taurine	10.13	Y2a-Eu
SRR866389	Limousin	France	C. and N. European taurine	3.53	Y2a-Eu
SRR866419	Limousin	France	C. and N. European taurine	3.21	Y2a-Eu
SAMN06699024	Lingnan/Linnan	China (north-central)	Asian hybrid	9.40	Y3
SAMN06699031	Lingnan/Linnan	China (north-central)	Asian hybrid	8.32	Y3
SAMN06699027	Lingnan/Linnan	China (north-central)	Asian hybrid	8.30	Y3
SAMN06699026	Lingnan/Linnan	China (north-central)	Asian hybrid	8.00	Y3
SAMN06699025	Lingnan/Linnan	China (north-central)	Asian hybrid	6.15	Y3
CL100026723	Lohani	Pakistan	Asian indicine	19.43	Y3
SAMN06699004	Luxi	China (north-central)	Asian hybrid	9.44	Y2a-Eu

SAMN06699008	Luxi	China (north-central)	Asian hybrid	8.76	Y2a-Eu
SAMN06699005	Luxi	China (north-central)	Asian hybrid	8.66	Y2a-Eu
SAMN06699006	Luxi	China (north-central)	Asian hybrid	7.73	Y2a-Eu
SAMN06699007	Luxi	China (north-central)	Asian hybrid	7.22	Y2a-Eu
SRR1365125	Maine Anjou	France	C. and N. European taurine	12.98	Y1-modern
SRR1355245	Maine Anjou	France	C. and N. European taurine	12.33	Y1-modern
SRR1365108	Maine Anjou	France	C. and N. European taurine	11.95	Y1-modern
SRR1355238	Maine Anjou	France	C. and N. European taurine	11.01	Y1-modern
SRR1355259	Maine-Anjou	France	C. and N. European taurine	14.16	Y1-modern
SRR8426544	Maremmana	Italy	Southern European	8.54	Y2a-Eu
SRR8426535	Maremmana	Italy	Southern European	6.33	Y2a-Eu
SRR8426540	Maronesa	Portugal	C. and N. European taurine	4.86	Y2a-Eu
DRR001771,78,81	Mishima	Japan	Asian taurine	8.35	Y2b-east
SAMN06698981	Mongolian	China (northwest)	Asian taurine	10.79	Y2a-Eu
SAMN06698984	Mongolian	China (northwest)	Asian taurine	10.46	Y2a-Eu
SAMN06698983	Mongolian	China (northwest)	Asian taurine	10.16	Y2a-Eu
SAMN06698982	Mongolian	China (northwest)	Asian taurine	10.03	Y2a-Eu
SAMN06698985	Mongolian	China (northwest)	Asian taurine	9.10	Y2a-Eu
ERR883699	Montbeliarde	France	C. and N. European taurine	9.20	Y2a-Eu
ERR883701	Montbeliarde	France	C. and N. European taurine	8.52	Y2a-Eu
ERR883698	Montbeliarde	France	C. and N. European taurine	8.50	Y2a-Eu
ERR883694	Montbeliarde	France	C. and N. European taurine	8.07	Y2a-Eu
ERR883700	Montbeliarde	France	C. and N. European taurine	6.18	Y2a-Eu
SRR5630653	Muturu	West Africa	African taurine	6.40	Y2a-Af
ERR883704	Normande	France	C. and N. European taurine	9.87	Y1-modern
ERR883714	Normande	France	C. and N. European taurine	9.04	Y1-modern
ERR883705	Normande	France	C. and N. European taurine	7.71	Y1-modern
ERR883707	Normande	France	C. and N. European taurine	7.18	Y1-modern
ERR883706	Normande	France	C. and N. European taurine	6.90	Y1-modern
SRR8426541	Pajuna	Spain	Iberia	8.43	Y1-modern*
SRR8426542	Pajuna	Spain	Iberia	8.29	Y2a-Eu

SRR1525606	Piedmontese	Italy	Southern European	5.99	Y2a-Eu
SRR1525596	Piedmontese	Italy	Southern European	5.87	Y2a-Eu
SRR1525603	Piedmontese	Italy	Southern European	5.60	Y2a-Eu
SRR1525604	Piedmontese	Italy	Southern European	5.41	Y2a-Eu
SRR1525599	Piedmontese	Italy	Southern European	4.91	Y2b-east
SRR1525614	Red Angus	Scotland	C. and N. European taurine	5.75	Y1-modern
CL100026724	Red Sindhi	Pakistan	Asian indicine	18.74	Y3
SRR1346378	Salers	France	C. and N. European taurine	10.04	Y2a-Eu
SRR1525608	Salers	France	C. and N. European taurine	5.54	Y2a-Eu
ERR3305587	Scottish Highland	Scotland	C. and N. European taurine	32.72	Y2a-Eu
SRR1525703	Simmental	Switzerland	C. and N. European taurine	12.88	Y2a-Eu
SRR1525705	Simmental	Switzerland	C. and N. European taurine	11.66	Y2a-Eu
SRR1525702	Simmental	Switzerland	C. and N. European taurine	11.53	Y2a-Eu
SRR1525700	Simmental	Switzerland	C. and N. European taurine	10.77	Y2a-Eu
SRR1525701	Simmental	Switzerland	C. and N. European taurine	10.47	Y2a-Eu
ERR3323643	South Anatolian Red	Anatolia	Asian hybrid	6.93	Y2b-Af
SAMN06699003	Tibetan	Tibet	Asian taurine	11.28	Y2b-east
SAMN06698999	Tibetan	Tibet	Asian taurine	9.52	Y2b-east
SAMN06699001	Tibetan	Tibet	Asian taurine	9.01	Y2b-east
SAMN06698995	Tibetan	Tibet	Asian taurine	8.40	Y2b-east
SAMN06698996	Tibetan	Tibet	Asian taurine	8.08	Y2b-east
ERR3293561	Wagyu	Japan	Asian taurine	15.84	Y2b-east
SAMN06699048	Wannan	China (Southeast)	Asian indicine	19.06	Y3
SAMN06699047	Wannan	China (Southeast)	Asian indicine	18.86	Y3
SAMN06699051	Wannan	China (Southeast)	Asian indicine	10.06	Y3
SAMN06699050	Wannan	China (Southeast)	Asian indicine	8.59	Y3
SAMN06699049	Wannan	China (Southeast)	Asian indicine	8.27	Y3
ERR2734952	Yakutian	Siberia	Asian taurine	9.29	Y2b-east
SRR6234774	Yanbian	China (north)	Asian taurine	8.46	Y2a-Eu

Appendix Table 8.2 - Samples above 8X selected for BEAST analysis.

Sample	Species/Breed	Location	Context	BEAST calibration	Dataset used
CL00006972	Achai	Pakistan/Afghanistan	Asian indicine	0	1

ERR3305588	Alentejana	Portugal	C. and N. European taurine	0	1,2,3
SRR1425124	Angus	Scotland	C. and N. European taurine	0	1,2,3
SRR1365144	Angus	Scotland	C. and N. European taurine	0	1,2,3
SRR1365129	Angus	Scotland	C. and N. European taurine	0	1,2,3
SRR1355237	Angus	Scotland	C. and N. European taurine	0	1,2,3
SRR1425145	Angus	Scotland	C. and N. European taurine	0	1,2,3
ERR4472515	Baoule	Ivory Coast	African taurine	0	1,2,3
SAMN06699039	Bashan	China (north-central)	Asian hybrid	0	1,2,3
SAMN06699041	Bashan	China (north-central)	Asian hybrid	0	1,2,3
SAMN06699040	Bashan	China (north-central)	Asian hybrid	0	1,2,3
SAMN06699037	Bashan	China (north-central)	Asian hybrid	0	1
SAMN06699038	Bashan	China (north-central)	Asian hybrid	0	1
SAMN06699021	Bohai Black	China (north-central)	Asian hybrid	0	1,2,3
SAMN06699022	Bohai Black	China (north-central)	Asian hybrid	0	1,2,3
Siv1	Bos primigenius	Bulgaria	6258 - 6006 cal yrBP	6131	1,2,3
Yen1	Bos primigenius	Russia	14193 - 14019 cal yrBP	14106	1,2,3
Gal1	Bos primigenius	Spain	9538 - 9109 cal yrBP	9330	1,2,3
Gyu1	Bos primigenius	Armenia	14097 - 13863 cal yrBP	13979	1,2
Baikal1	Bos primigenius	Russia	12015 - 12529 cal yrBP	12271	1,2
Var2	Bos primigenius	Kazakhstan	7486 - 7238 cal yrBP	7361	1,2
Rhi1	Bos primigenius	Germany	47,000+ cal yrBP	(40000.0,160000.0)	1,2,3
Sub1	Bos taurus	Turkey	Neolithic	8102	1,2,3
New1	Bos taurus	Ireland	Neolithic	4748	1,2,3
Bes2	Bos taurus x Bos indicus	Iraqi Kurdistan	Mediaeval	677	1
SRR8426534	Boskarin	Croatia	Southern European	0	1,2,3
ERR1766310	Brown Swiss	Switzerland	C. and N. European taurine	0	1,2,3
SRR5507272	Chaidamu Yellow Cattle	China (north)	Asian taurine	0	1,2,3
SRR5507271	Chaidamu Yellow Cattle	China (north)	Asian taurine	0	1,2,3
SRR1348577	Charolais	France	C. and N. European taurine	0	1,2,3
SRR1348574	Charolais	France	C. and N. European taurine	0	1,2,3
SRR1355258	Charolais	France	C. and N. European taurine	0	1,2,3

SRR1343169	Charolais	France	C. and N. European taurine	0	1,2,3
SRR1348584	Charolais	France	C. and N. European taurine	0	1,2,3
SRR1355254	Charolais	France	C. and N. European taurine	0	1,2,3
SAMN05216091	Chianina	Italy	Southern European	0	1,2,3
CL100006973	Cholistani	Pakistan	Asian indicine	0	1
SAMN05803883	Dexter	Ireland	C. and N. European taurine	0	1,2,3
CL100012095	Dhanni	Pakistan	Asian indicine	0	1
CL100007106	Gabrali	Pakistan/Afghanistan	Asian indicine	0	1
SRR1365112	Gelbvieh	Germany	C. and N. European taurine	0	1,2,3
SRR1355236	Gelbvieh	Germany	C. and N. European taurine	0	1,2,3
SRR1348575	Gelbvieh	Germany	C. and N. European taurine	0	1,2,3
SRR1346388	Gelbvieh	Germany	C. and N. European taurine	0	1,2,3
SRR2016752	Gir	India	South Asian indicine	0	1
SRR2016754	Gir	India	South Asian indicine	0	1
SRR934432	Hanwoo	Korea	Asian taurine	0	1,2,3
SRR934418	Hanwoo	Korea	Asian taurine	0	1,2,3
SRR934415	Hanwoo	Korea	Asian taurine	0	1,2,3
SRR934419	Hanwoo	Korea	Asian taurine	0	1,2,3
SRR934417	Hanwoo	Korea	Asian taurine	0	1,2,3
SRR1365131	Hereford	England	C. and N. European taurine	0	1,2,3
SRR1365126	Hereford	England	C. and N. European taurine	0	1,2,3
SRR1365124	Hereford	England	C. and N. European taurine	0	1,2,3
SRR1365128	Hereford	England	C. and N. European taurine	0	1,2,3
SRR1365137	Hereford	England	C. and N. European taurine	0	1,2,3
SRR1348583	Holstein	Netherlands	C. and N. European taurine	0	1,2,3
SRR1425144	Holstein	Netherlands	C. and N. European taurine	0	1,2,3
SRR1425133	Holstein	Netherlands	C. and N. European taurine	0	1,2,3
SRR1425129	Holstein	Netherlands	C. and N. European taurine	0	1,2,3
SRR1346386	Holstein	Netherlands	C. and N. European taurine	0	1,2,3
SRR3497161	Jersey	Jersey	C. and N. European taurine	0	1,2,3
SRR3497466	Jersey	Jersey	C. and N. European taurine	0	1,2,3

SRR3497611	Jersey	Jersey	C. and N. European taurine	0	1,2,3
SRR3497462	Jersey	Jersey	C. and N. European taurine	0	1,2,3
SRR5507282	Jian	China (Southeast)	Asian indicine	0	1
SRR5507281	Jian	China (Southeast)	Asian indicine	0	1
SRR5507285	Jiangxi	China (Southeast)	Asian indicine	0	1
SRR5507286	Jiangxi	China (Southeast)	Asian indicine	0	1
SRR5507284	Jiangxi	China (Southeast)	Asian indicine	0	1
SAMN06699015	Jiaxian Red	China (north-central)	Asian hybrid	0	1,2,3
SAMN06699014	Jiaxian Red	China (north-central)	Asian hybrid	0	1,2,3
SAMN06699018	Jiaxian Red	China (north-central)	Asian hybrid	0	1,2,3
SAMN06699016	Jiaxian Red	China (north-central)	Asian hybrid	0	1,2,3
SAMN06699017	Jiaxian Red	China (north-central)	Asian hybrid	0	1
SAMN06698986	Kazakh	China (northwest)	Asian taurine	0	1,2,3
SAMN06698990	Kazakh	China (northwest)	Asian taurine	0	1,2,3
SAMN06698989	Kazakh	China (northwest)	Asian taurine	0	1,2,3
SAMN06698991	Kazakh	China (northwest)	Asian taurine	0	1,2,3
SAMN06698992	Kazakh	China (northwest)	Asian taurine	0	1,2,3
SAMD00013963	Kuchinoshima	Japan	Asian taurine	0	1,2,3
ERR3293557	Lagune	West/Central Africa	African taurine	0	1,2,3
SRR1365134	Limousin	France	C. and N. European taurine	0	1,2,3
SRR1365136	Limousin	France	C. and N. European taurine	0	1,2,3
SAMN06699024	Lingnan/Linnan	China (north-central)	Asian hybrid	0	1
SAMN06699031	Lingnan/Linnan	China (north-central)	Asian hybrid	0	1
SAMN06699027	Lingnan/Linnan	China (north-central)	Asian hybrid	0	1
SAMN06699026	Lingnan/Linnan	China (north-central)	Asian hybrid	0	1
CL100026723	Lohani	Pakistan	Asian indicine	0	1
SAMN06699004	Luxi	China (north-central)	Asian hybrid	0	1,2,3
SAMN06699008	Luxi	China (north-central)	Asian hybrid	0	1,2,3
SAMN06699005	Luxi	China (north-central)	Asian hybrid	0	1,2,3
SRR1365125	Maine Anjou	France	C. and N. European taurine	0	1,2,3
SRR1355245	Maine Anjou	France	C. and N. European taurine	0	1,2,3
SRR1365108	Maine Anjou	France	C. and N. European taurine	0	1,2,3

SRR1355238	Maine Anjou	France	C. and N. European taurine	0	1,2,3
SRR1355259	Maine-Anjou	France	C. and N. European taurine	0	1,2,3
SRR8426544	Maremmana	Italy	Southern European	0	1,2,3
DRR001771,78,81	Mishima	Japan	Asian taurine	0	1,2,3
SAMN06698981	Mongolian	China (northwest)	Asian taurine	0	1,2,3
SAMN06698984	Mongolian	China (northwest)	Asian taurine	0	1,2,3
SAMN06698983	Mongolian	China (northwest)	Asian taurine	0	1,2,3
SAMN06698982	Mongolian	China (northwest)	Asian taurine	0	1,2,3
SAMN06698985	Mongolian	China (northwest)	Asian taurine	0	1,2,3
ERR883699	Montbeliarde	France	C. and N. European taurine	0	1,2,3
ERR883701	Montbeliarde	France	C. and N. European taurine	0	1,2,3
ERR883698	Montbeliarde	France	C. and N. European taurine	0	1,2,3
ERR883694	Montbeliarde	France	C. and N. European taurine	0	1,2,3
ERR883704	Normande	France	C. and N. European taurine	0	1,2,3
ERR883714	Normande	France	C. and N. European taurine	0	1,2,3
SRR8426541	Pajuna	Spain	Iberia	0	1,2,3
SRR8426542	Pajuna	Spain	Iberia	0	1,2,3
CL100026724	Red Sindhi	Pakistan	Asian indicine	0	1
SRR1346378	Salers	France	C. and N. European taurine	0	1,2,3
ERR3305587	Scottish Highland	Scotland	C. and N. European taurine	0	1,2,3
SRR1525703	Simmental	Switzerland	C. and N. European taurine	0	1,2,3
SRR1525705	Simmental	Switzerland	C. and N. European taurine	0	1,2,3
SRR1525702	Simmental	Switzerland	C. and N. European taurine	0	1,2,3
SRR1525700	Simmental	Switzerland	C. and N. European taurine	0	1,2,3
SRR1525701	Simmental	Switzerland	C. and N. European taurine	0	1,2,3
SAMN06699003	Tibetan	Tibet	Asian taurine	0	1,2,3
SAMN06698999	Tibetan	Tibet	Asian taurine	0	1,2,3
SAMN06699001	Tibetan	Tibet	Asian taurine	0	1,2,3
SAMN06698995	Tibetan	Tibet	Asian taurine	0	1,2,3
SAMN06698996	Tibetan	Tibet	Asian taurine	0	1,2,3
ERR3293561	Wagyu	Japan	Asian taurine	0	1,2,3

SAMN06699048	Wannan	China (Southeast)	Asian indicine	0	1
SAMN06699047	Wannan	China (Southeast)	Asian indicine	0	1
SAMN06699051	Wannan	China (Southeast)	Asian indicine	0	1
SAMN06699050	Wannan	China (Southeast)	Asian indicine	0	1
SAMN06699049	Wannan	China (Southeast)	Asian indicine	0	1
ERR2734952	Yakutian	Siberia	Asian taurine	0	1,2,3
SRR6234774	Yanbian	China (north)	Asian taurine	0	1,2,3

Appendix Table 8.3 - Reduced dataset of ancient and modern bulls used for SNP calling.

NCBI Accession	Breed	Region	Cattle Type	Mean Depth Post Filtering	Haplotype
SRR1355259	Maine-Anjou	France	C. and N. European taurine	14.16	Y1-modern
SRR1348583	Holstein	Netherlands	C. and N. European taurine	18.57	Y1-modern
SRR1425124	Angus	Scotland	C. and N. European taurine	14.01	Y1-modern
ERR3293561	Wagyu	Japan	Asian taurine	15.84	Y2b-east
SAMN06699003	Tibetan	Tibet	Asian taurine	11.28	Y2b-east
SRR934418	Hanwoo	Korea	Asian taurine	10.21	Y2b-east
ERR3305587	Scottish Highland	Scotland	C. and N. European taurine	32.72	Y2a-Eu
SRR1355258	Charolais	France	C. and N. European taurine	14.10	Y2a-Eu
ERR3305588	Alentejana	Portugal	C. and N. European taurine	26.88	Y2a-Eu
ERR3293557	Lagune	West/Central Africa	African taurine	20.74	Y2a-Af
ERR4472515	Baoule	Ivory Coast	African taurine	25.77	Y2a-Af
SAMN06699048	Wannan	China (Southeast)	Asian indicine	19.06	Y3
SAMN06699017	Jiaxian Red	China (north-central)	Asian hybrid	11.99	Y3
SRR5507282	Jian	China (Southeast)	Asian indicine	9.84	Y3
CL100026724	Red Sindhi	Pakistan	Asian indicine	18.74	Y3
CL100026723	Lohani	Pakistan	Asian indicine	19.43	Y3
SRR2016752	Gir	India	South Asian indicine	11.26	Y3
Gal1	*Bos primigenius*	Spain	Galicia	17.08	Y2c
New1	*Bos taurus*	Ireland	Newgrange	16.38	Y2b-earlydom
Yen1	*Bos primigenius*	Russia	Yenisei River	14.90	Y4b
Sub1	*Bos taurus*	Turkey	Suberde	13.85	Y2b-earlydom
Bes2	*Bos taurus x Bos indicus*	Iraqi Kurdistan	Bestansur	13.80	Y3

Baikal1	*Bos primigenius*	Baikal	Russia	13.58	Y4b
Siv1	*Bos primigenius*	Bulgaria	Vratsa	9.03	Y1-basal
Var2	*Bos primigenius*	Kazakhstan	Varfolomeevka	9.00	Y4a
Gyu1	*Bos primigenius*	Armenia	Gyumri	18.18	Y4a
Dzh1	*Bos taurus*	Bulgaria	Dzulhunitsa	9.68	Y2b-earlydom
Rhi1	*Bos primigenius*	Germany	Upper Rhine Valley	9.54	Y1-Y2-out

Appendix Table 9.1 - List of modern breeds used in Chapter 9.

NCBI Accession	Breed	Region	Cattle Type	Mean Depth Post Filtering
ERR3305588	Alentejana	Portugal	Iberian taurine	26.88
SRR1425145	Angus	Scotland	East Asian taurine	10.92
ERR4472510	Baoule	Ivory Coast	West African taurine	28.09
SAMN06699039	Bashan	China (north-central)	East Asian hybrid	9.82
108375150	Belgian Blue	Belgium	C. and N. European taurine	8.07
SAMN05803880	Belted Galloway	Scotland	C. and N. European taurine	13.05
SAMN01915355	Blonde d'Aquitaine	France	C. and N. European taurine	6.21
SAMN06699021	Bohai Black	China (north-central)	East Asian hybrid	9.09
SRR3546729	Boran	East Africa	East African hybrid	7.51
Blk123*	Busha	Former Yugolsavia	Southern European taurine	2.93
SRR5507272	Chaidamu Yellow Cattle	China (north)	East Asian taurine	10.67
SRR1348584	Charolais	France	C. and N. European taurine	8.73
SAMN05216091	Chianina	Italy	Southern European taurine	11.29
ERR3890053	Danish Red	Denmark	C. and N. European taurine	7.44
SRR1346378	Devon	England	East Asian taurine	10.17
SAMN05803888	Dexter	Ireland	C. and N. European taurine	12.5
ERR1742687	Eringer	Switzerland	C. and N. European taurine	17.03
ERR2734948	Finncattle	Finland	C. and N. European taurine	9.33
SRR1355236	Gelbvieh	Germany	C. and N. European taurine	12.1
SRR2016752	Gir	India	South Asian indicine	11.26
ERR4494255	Gourounsi	Burkina Faso	West African taurine	27.16
SRR5507286	Jiangxi	China (Southeast)	East Asian indicine	8.31
SRR934418	Hanwoo	Korea	East Asian taurine	10.21
SAMN08862747	Hariana	Pakistan/India	South Asian indicine	29.85

SRR1365126	Hereford	England	C. and N. European taurine	13.48
SRR1425144	Holstein	Netherlands	C. and N. European taurine	14.22
SRR3497465	Jersey	Jersey	C. and N. European taurine	11.5
SRR5507282	Jian	China (Southeast)	East Asian indicine	9.84
SAMN06699014	Jiaxian Red	China (north-central)	East Asian hybrid	9.63
SAMN06698986	Kazakh	China (northwest)	Central Asian taurine	11.2
SRR3694652	Kenana	East Africa	East African hybrid	8.3
SAMD00013963	Kuchinoshima	Japan	East Asian taurine	12.96
ERR3293557	Lagune	West/Central Africa	West African taurine	20.74
SRR8426539	Limia	Spain	Iberian taurine	7.89
SAMN01915358	Limousin	France	C. and N. European taurine	3.53
SAMN06699031	Lingnan/Linnan	China (north-central)	East Asian hybrid	8.3
SAMN06699005	Luxi	China (north-central)	East Asian hybrid	8.66
SRR1365108	Maine Anjou	France	C. and N. European taurine	11.95
830315*	Maltese	Malta	Southern European taurine	6.33
SRR8426543	Maremmana	Italy	Southern European taurine	8.54
SRR8426540	Maronesa	Portuguese	Iberian taurine	4.86
DRR001771,78,81	Mishima	Japan	East Asian taurine	8.8
SAMN06698983	Mongolian	China (northwest)	East Asian taurine	10.16
ERR883698	Montbeliarde	France	C. and N. European taurine	8.5
SRR5630649	Muturu	West Africa	West African taurine	8.05
SRR3694478	N'dama	West Africa	West African taurine	8.44
ERR883714	Normande	France	C. and N. European taurine	9.04
SRR3490689	Ogaden	Ethiopia	East African hybrid	7.57
ERR5711624	Original Braunvieh	Switzerland	C. and N. European taurine	13.24
SRR8426541	Pajuna	Spain	Iberian taurine	8.43
SRR1525606	Piedmontese	Italy	Southern European taurine	5.99
SRR8426536	Podolica	Italy	Southern European taurine	10
ERR454989	Rashoki	Iran	Southwest Asian hybrid	11.13
SAMN08862748	Sahiwal	India	South Asian indicine	17.9
SRR8426538	Sayaguesa	Spain	Iberian taurine	7.59

ERR3305587	Scottish Highland	Scotland	C. and N. European taurine	32.72
SAMN05216073	Shorthorn	England	C. and N. European taurine	9.17
ERR3293559	Sikia	Greece	Southern European taurine	17.37
SRR1525702	Simmental	Switzerland	C. and N. European taurine	11.53
ERR3305591	Somba	West/Central Africa	West African taurine	28.37
ERR3323643	South Anatolian Red	Anatolia	Southwest Asian hybrid	6.93
ERR1256961	Swedish Red	Sweden	C. and N. European taurine	7.31
SAMN08862749	Tharparkar	Pakistan/India	South Asian indicine	14.05
SAMN06698999	Tibetan	Tibet	East Asian taurine	9.52
ERR3376618	Tiroler Grauvieh	Austria	C. and N. European taurine	15.34
ERR3293561	Wagyu	Japan	East Asian taurine	15.84
SAMN06699047	Wannan	China (Southeast)	East Asian indicine	18.86
ERR4414059	Water Buffalo	India	Outgroup	35.92
ERR2734952	Yakutian	Siberia	East Asian taurine	9.29
SAMN07431001	Yanbian	China (north)	East Asian taurine	8.46

Appendix Table 9.2 - Relative allele-sharing of four populations of *Bos primigenius* variation: Southwest Asian aurochs, Pleistocene East Eurasian aurochs, European aurochs and *Bos namadicus*.

P1	P2	P3	P4	*D* statistic	Standard Error	Z Score	Number of Sites
East	SW Asia	Indicine	WaterBuffalo	0.024	0.0050	4.81	2114611
SW Asia	Europe	Indicine	WaterBuffalo	0.052	0.0042	12.35	2114994
East	Europe	Indicine	WaterBuffalo	0.064	0.0039	-16.50	2114901
Europe	SW Asia	East	WaterBuffalo	-0.011	0.0042	-2.61	2115049
Indicine	Europe	East	WaterBuffalo	-0.481	0.0045	-107.80	2114901
Indicine	SW Asia	East	WaterBuffalo	-0.493	0.0045	-109.11	2114611
East	Europe	SW Asia	WaterBuffalo	-0.218	0.0041	-53.47	2115049
Indicine	East	SW Asia	WaterBuffalo	-0.511	0.0057	-88.86	2114611
Indicine	Europe	SW Asia	WaterBuffalo	-0.615	0.0050	-123.49	2114994
East	Indicine	Europe	WaterBuffalo	-0.646	0.0039	166.98	2114994
SW Asia	East	Europe	WaterBuffalo	0.228	0.0043	53.43	2115049
Indicine	SW Asia	Europe	WaterBuffalo	-0.529	0.0048	-110.73	2114901

Appendix Table 9.3 - Relative allele-sharing of indicine domesticates with European aurochs.

P1	P2	P3	P4	*D* statistic	Standard Error	Z Score	Number of Sites
Bar1	Bed3	Indicine	WaterBuffalo	0.012	0.009	1.38	314415
Bed4	Bed3	Indicine	WaterBuffalo	0.001	0.005	0.20	1679859
Belarus1	Bed3	Indicine	WaterBuffalo	0.005	0.005	1.03	1608323
Blec1	Bed3	Indicine	WaterBuffalo	0.014	0.006	2.47	1250400
Cpc	Bed3	Indicine	WaterBuffalo	-0.003	0.005	-0.59	1913197
Fal7	Bed3	Indicine	WaterBuffalo	-0.064	0.006	-10.66	1149772
Fif1	Bed3	Indicine	WaterBuffalo	-0.030	0.015	-2.04	133415
Gal1	Bed3	Indicine	WaterBuffalo	-0.003	0.005	-0.71	2109220
Gal2	Bed3	Indicine	WaterBuffalo	-0.017	0.006	-2.91	1049974
Gal3	Bed3	Indicine	WaterBuffalo	-0.012	0.005	-2.33	1966686
Gallo1	Bed3	Indicine	WaterBuffalo	-0.078	0.014	-5.48	121168
Hjo1	Bed3	Indicine	WaterBuffalo	-0.015	0.005	-2.86	1365453
Horn-10537	Bed3	Indicine	WaterBuffalo	-0.048	0.017	-2.87	94422
Horn-10543	Bed3	Indicine	WaterBuffalo	-0.043	0.034	-1.25	18129
Horn-CMLIX	Bed3	Indicine	WaterBuffalo	-0.093	0.029	-3.21	30663
Hxh2	Bed3	Indicine	WaterBuffalo	0.003	0.005	0.68	2109244
Kalisz1	Bed3	Indicine	WaterBuffalo	0.010	0.006	1.67	861678
Kalisz2	Bed3	Indicine	WaterBuffalo	0.003	0.005	0.56	1862337
Kalisz3	Bed3	Indicine	WaterBuffalo	0.000	0.005	0.04	1599389
Lil1	Bed3	Indicine	WaterBuffalo	-0.004	0.005	-0.73	2084916
Lun1	Bed3	Indicine	WaterBuffalo	-0.011	0.008	-1.36	387269
Lun2	Bed3	Indicine	WaterBuffalo	-0.008	0.014	-0.56	131872
Pad1	Bed3	Indicine	WaterBuffalo	-0.004	0.010	-0.38	248649
Pal1	Bed3	Indicine	WaterBuffalo	0.018	0.005	3.69	1806767
Rhi1	Bed3	Indicine	WaterBuffalo	-0.005	0.005	-1.14	1929141
Rhi2	Bed3	Indicine	WaterBuffalo	-0.017	0.005	-3.17	1513251
Rhi3	Bed3	Indicine	WaterBuffalo	0.004	0.005	0.68	1924461
Rhi4	Bed3	Indicine	WaterBuffalo	-0.029	0.009	-3.25	343287
Rhi5	Bed3	Indicine	WaterBuffalo	-0.012	0.009	-1.42	311228
Rhi6	Bed3	Indicine	WaterBuffalo	-0.028	0.010	-2.86	227272
Rhi7	Bed3	Indicine	WaterBuffalo	-0.006	0.013	-0.51	133945
Rhi8	Bed3	Indicine	WaterBuffalo	-0.026	0.020	-1.31	53729

Siv1	Bed3	Indicine	WaterBuffalo	0.018	0.005	3.63	2106233
Ska1	Bed3	Indicine	WaterBuffalo	0.000	0.005	-0.08	1261525
Ska3	Bed3	Indicine	WaterBuffalo	0.001	0.005	0.13	1820616
Tango1	Bed3	Indicine	WaterBuffalo	0.051	0.005	10.42	1589452
Ton1	Bed3	Indicine	WaterBuffalo	-0.005	0.005	-0.94	1803444
Tri1	Bed3	Indicine	WaterBuffalo	0.001	0.005	0.18	1816504
Uzz1	Bed3	Indicine	WaterBuffalo	0.009	0.005	1.66	1591869
YoA	Bed3	Indicine	WaterBuffalo	-0.002	0.005	-0.47	2100709

Appendix Table 9.4 - Relative allele-sharing of Neolithic Southwest Asian domesticates with European aurochs.

P1	P2	P3	P4	*D* statistic	Standard Error	Z Score	Number of Sites
Bar1	Bed3	SWAsia	WaterBuffalo	0.025	0.008	3.01	314474
Bed4	Bed3	SWAsia	WaterBuffalo	0.000	0.005	0.06	1680175
Belarus1	Bed3	SWAsia	WaterBuffalo	0.032	0.005	6.26	1608607
Blec1	Bed3	SWAsia	WaterBuffalo	0.070	0.006	12.45	1250624
Cpc	Bed3	SWAsia	WaterBuffalo	0.003	0.005	0.68	1913548
Fal7	Bed3	SWAsia	WaterBuffalo	-0.006	0.006	-0.94	1149956
Fif1	Bed3	SWAsia	WaterBuffalo	-0.019	0.012	-1.52	133441
Gal1	Bed3	SWAsia	WaterBuffalo	-0.009	0.005	-1.78	2109577
Gal2	Bed3	SWAsia	WaterBuffalo	-0.003	0.005	-0.57	1050174
Gal3	Bed3	SWAsia	WaterBuffalo	-0.006	0.005	-1.13	1967006
Gallo1	Bed3	SWAsia	WaterBuffalo	0.007	0.012	0.60	121193
Hjo1	Bed3	SWAsia	WaterBuffalo	0.001	0.005	0.11	1365710
Horn-10537	Bed3	SWAsia	WaterBuffalo	-0.031	0.014	-2.28	94438
Horn-10543	Bed3	SWAsia	WaterBuffalo	-0.006	0.029	-0.19	18133
Horn-CMLIX	Bed3	SWAsia	WaterBuffalo	-0.030	0.023	-1.30	30673
Hxh2	Bed3	SWAsia	WaterBuffalo	0.004	0.005	0.84	2109600
Kalisz1	Bed3	SWAsia	WaterBuffalo	0.044	0.005	8.02	861855
Kalisz2	Bed3	SWAsia	WaterBuffalo	0.024	0.005	4.50	1862673
Kalisz3	Bed3	SWAsia	WaterBuffalo	0.036	0.005	7.20	1599692
Lil1	Bed3	SWAsia	WaterBuffalo	0.048	0.005	8.83	2085268
Lun1	Bed3	SWAsia	WaterBuffalo	-0.005	0.008	-0.63	387332
Lun2	Bed3	SWAsia	WaterBuffalo	-0.003	0.012	-0.27	131902

Pad1	Bed3	SWAsia	WaterBuffalo	0.023	0.009	2.56	248705
Pal1	Bed3	SWAsia	WaterBuffalo	0.033	0.005	6.62	1807099
Rhi1	Bed3	SWAsia	WaterBuffalo	-0.121	0.005	-24.48	1929496
Rhi2	Bed3	SWAsia	WaterBuffalo	-0.128	0.005	-24.76	1513533
Rhi3	Bed3	SWAsia	WaterBuffalo	0.002	0.005	0.33	1924808
Rhi4	Bed3	SWAsia	WaterBuffalo	-0.137	0.007	-18.64	343343
Rhi5	Bed3	SWAsia	WaterBuffalo	-0.130	0.008	-16.17	311279
Rhi6	Bed3	SWAsia	WaterBuffalo	-0.139	0.009	-15.61	227317
Rhi7	Bed3	SWAsia	WaterBuffalo	-0.125	0.011	-11.86	133966
Rhi8	Bed3	SWAsia	WaterBuffalo	-0.147	0.016	-9.04	53741
Siv1	Bed3	SWAsia	WaterBuffalo	0.071	0.005	13.64	2106590
Ska1	Bed3	SWAsia	WaterBuffalo	0.013	0.005	2.32	1261765
Ska3	Bed3	SWAsia	WaterBuffalo	0.009	0.005	1.74	1820938
Tango1	Bed3	SWAsia	WaterBuffalo	-0.002	0.005	-0.35	1589730
Tango2	Bed3	SWAsia	WaterBuffalo	0.052	0.005	10.54	1702277
Ton1	Bed3	SWAsia	WaterBuffalo	0.031	0.005	6.21	1803750
Tri1	Bed3	SWAsia	WaterBuffalo	0.007	0.005	1.43	1816828
Uzz1	Bed3	SWAsia	WaterBuffalo	0.031	0.005	6.01	1592163
YoA	Bed3	SWAsia	WaterBuffalo	0.013	0.005	2.63	2101063

Appendix Table 9.5 - Relative allele-sharing of Eastern aurochs with European aurochs.

P1	P2	P3	P4	*D* statistic	Standard Error	Z Score	Number of Sites
Bar1	Bed3	East	WaterBuffalo	0.0165	0.0095	1.75	314440
Bed4	Bed3	East	WaterBuffalo	0.0012	0.0056	0.22	1679971
Belarus1	Bed3	East	WaterBuffalo	0.0130	0.0057	2.30	1608421
Blec1	Bed3	East	WaterBuffalo	0.0306	0.0062	4.93	1250472
Cpc	Bed3	East	WaterBuffalo	0.0047	0.0053	0.88	1913304
Fal7	Bed3	East	WaterBuffalo	-0.0671	0.0066	-10.19	1149816
Fif1	Bed3	East	WaterBuffalo	-0.0279	0.0139	-2.02	133430
Gal1	Bed3	East	WaterBuffalo	-0.0002	0.0060	-0.03	2109308
Gal2	Bed3	East	WaterBuffalo	-0.0095	0.0066	-1.43	1050058
Gal3	Bed3	East	WaterBuffalo	-0.0017	0.0056	-0.30	1966759
Gallo1	Bed3	East	WaterBuffalo	-0.0561	0.0136	-4.13	121184
Hjo1	Bed3	East	WaterBuffalo	-0.0027	0.0060	-0.45	1365541

Horn-10537	Bed3	East	WaterBuffalo	-0.0717	0.0152	-4.71	94424
Horn-10543	Bed3	East	WaterBuffalo	-0.0416	0.0337	-1.24	18132
Horn-CMLIX	Bed3	East	WaterBuffalo	-0.0662	0.0285	-2.32	30666
Hxh2	Bed3	East	WaterBuffalo	0.0083	0.0055	1.50	2109331
Kalisz1	Bed3	East	WaterBuffalo	0.0130	0.0063	2.08	861751
Kalisz2	Bed3	East	WaterBuffalo	0.0082	0.0056	1.47	1862443
Kalisz3	Bed3	East	WaterBuffalo	0.0193	0.0056	3.44	1599493
Lil1	Bed3	East	WaterBuffalo	-0.0026	0.0057	-0.45	2085010
Lun1	Bed3	East	WaterBuffalo	-0.0090	0.0085	-1.06	387289
Lun2	Bed3	East	WaterBuffalo	-0.0075	0.0128	-0.59	131887
Pad1	Bed3	East	WaterBuffalo	0.0105	0.0098	1.07	248674
Pal1	Bed3	East	WaterBuffalo	0.0353	0.0054	6.48	1806872
Rhi1	Bed3	East	WaterBuffalo	-0.0008	0.0055	-0.15	1929266
Rhi2	Bed3	East	WaterBuffalo	-0.0079	0.0058	-1.36	1513344
Rhi3	Bed3	East	WaterBuffalo	0.0104	0.0055	1.89	1924557
Rhi4	Bed3	East	WaterBuffalo	-0.0324	0.0086	-3.79	343308
Rhi5	Bed3	East	WaterBuffalo	-0.0099	0.0093	-1.06	311245
Rhi6	Bed3	East	WaterBuffalo	-0.0199	0.0100	-2.00	227282
Rhi7	Bed3	East	WaterBuffalo	-0.0165	0.0119	-1.39	133955
Rhi8	Bed3	East	WaterBuffalo	-0.0330	0.0191	-1.73	53735
Siv1	Bed3	East	WaterBuffalo	0.0271	0.0054	4.99	2106318
Ska1	Bed3	East	WaterBuffalo	0.0022	0.0061	0.36	1261600
Ska3	Bed3	East	WaterBuffalo	0.0055	0.0057	0.96	1820704
Tango2	Bed3	East	WaterBuffalo	0.0782	0.0064	12.24	1702077
Ton1	Bed3	East	WaterBuffalo	0.0075	0.0057	1.32	1803517
Tri1	Bed3	East	WaterBuffalo	0.0088	0.0054	1.61	1816592
Uzz1	Bed3	East	WaterBuffalo	0.0174	0.0060	2.88	1591967
YoA	Bed3	East	WaterBuffalo	0.0023	0.0056	0.41	2100796

Appendix Table 9.6 - Relative allele-sharing of Th7 with European aurochs.

P1	P2	P3	P4	*D* statistic	Standard Error	Z Score	Number of Sites
Bar1	Bed3	Th7	WaterBuffalo	0.0664	0.0316	2.10	23474
Bed4	Bed3	Th7	WaterBuffalo	-0.0023	0.0160	-0.14	110716
Belarus1	Bed3	Th7	WaterBuffalo	-0.0096	0.0168	-0.57	103780

Blec1	Bed3	Th7	WaterBuffalo	0.0149	0.0171	0.87	83635
Cpc	Bed3	Th7	WaterBuffalo	-0.0202	0.0148	-1.36	124525
Fal7	Bed3	Th7	WaterBuffalo	-0.0616	0.0180	-3.42	75597
Fif1	Bed3	Th7	WaterBuffalo	0.0173	0.0609	0.28	9014
Gal1	Bed3	Th7	WaterBuffalo	-0.0216	0.0149	-1.45	134391
Gal2	Bed3	Th7	WaterBuffalo	-0.0169	0.0196	-0.86	69847
Gal3	Bed3	Th7	WaterBuffalo	-0.0235	0.0156	-1.50	127102
Gallo1	Bed3	Th7	WaterBuffalo	0.0000	0.0592	0.00	7836
Hjo1	Bed3	Th7	WaterBuffalo	-0.0073	0.0177	-0.41	91514
Horn-10537	Bed3	Th7	WaterBuffalo	-0.1376	0.0688	-2.00	6152
Horn-10543	Bed3	Th7	WaterBuffalo	0.0244	0.1541	0.16	1241
Horn-CMLIX	Bed3	Th7	WaterBuffalo	-0.0238	0.1175	-0.20	2044
Hxh2	Bed3	Th7	WaterBuffalo	-0.0219	0.0151	-1.45	134385
Kalisz1	Bed3	Th7	WaterBuffalo	0.0351	0.0213	1.65	57852
Kalisz2	Bed3	Th7	WaterBuffalo	-0.0070	0.0165	-0.42	120371
Kalisz3	Bed3	Th7	WaterBuffalo	0.0018	0.0167	0.11	105085
Lil1	Bed3	Th7	WaterBuffalo	-0.0078	0.0140	-0.55	133426
Lun1	Bed3	Th7	WaterBuffalo	0.0083	0.0306	0.27	27326
Lun2	Bed3	Th7	WaterBuffalo	0.0206	0.0508	0.41	9433
Pad1	Bed3	Th7	WaterBuffalo	-0.0036	0.0378	-0.10	18700
Pal1	Bed3	Th7	WaterBuffalo	0.0244	0.0149	1.64	119561
Rhi1	Bed3	Th7	WaterBuffalo	-0.0727	0.0144	-5.05	126321
Rhi2	Bed3	Th7	WaterBuffalo	-0.1201	0.0157	-7.65	102714
Rhi3	Bed3	Th7	WaterBuffalo	-0.0122	0.0146	-0.84	125056
Rhi4	Bed3	Th7	WaterBuffalo	-0.0999	0.0315	-3.17	24172
Rhi5	Bed3	Th7	WaterBuffalo	-0.0876	0.0329	-2.66	22649
Rhi6	Bed3	Th7	WaterBuffalo	-0.1694	0.0372	-4.55	16102
Rhi7	Bed3	Th7	WaterBuffalo	-0.0058	0.0455	-0.13	9939
Rhi8	Bed3	Th7	WaterBuffalo	0.0508	0.0759	0.67	4068
Siv1	Bed3	Th7	WaterBuffalo	0.0317	0.0136	2.34	134289
Ska1	Bed3	Th7	WaterBuffalo	0.0098	0.0186	0.53	86594
Ska3	Bed3	Th7	WaterBuffalo	-0.0058	0.0159	-0.37	119656
Tango2	Bed3	Th7	WaterBuffalo	0.0195	0.0150	1.30	110101

Ton1	Bed3	Th7	WaterBuffalo	0.0034	0.0151	0.22	117548
Tri1	Bed3	Th7	WaterBuffalo	0.0156	0.0153	1.02	120993
Uzz1	Bed3	Th7	WaterBuffalo	0.0193	0.0165	1.17	105978
YoA	Bed3	Th7	WaterBuffalo	-0.0159	0.0148	-1.07	133997

Appendix Table 9.7 - Excess allele-sharing in post-LGM European aurochs with Southwest Asian genomes with Pleistocene Rhine Valley genomes as a reference.

P1	P2	P3	P4	*D* statistic	Standard Error	Z Score	Number of Sites
Blec1	Rhine	SWAsia	WaterBuffalo	0.1873	0.0053	35.43	1223281
Siv1	Rhine	SWAsia	WaterBuffalo	0.1816	0.0048	37.76	2038122
Kalisz1	Rhine	SWAsia	WaterBuffalo	0.1673	0.0051	33.09	847200
Tango2	Rhine	SWAsia	WaterBuffalo	0.1658	0.0046	35.86	1659080
Lil1	Rhine	SWAsia	WaterBuffalo	0.1653	0.0049	33.87	2018211
Pad1	Rhine	SWAsia	WaterBuffalo	0.1559	0.0085	18.44	245643
Kalisz3	Rhine	SWAsia	WaterBuffalo	0.1548	0.0049	31.31	1558997
Pal1	Rhine	SWAsia	WaterBuffalo	0.1535	0.0046	33.20	1759632
Belarus1	Rhine	SWAsia	WaterBuffalo	0.1523	0.0049	30.95	1567734
Ton1	Rhine	SWAsia	WaterBuffalo	0.1513	0.0047	32.11	1751843
Uzz1	Rhine	SWAsia	WaterBuffalo	0.1506	0.0049	30.85	1555816
Kalisz2	Rhine	SWAsia	WaterBuffalo	0.1432	0.0049	29.26	1810469
Bar1	Rhine	SWAsia	WaterBuffalo	0.1428	0.0071	20.05	310204
Ska1	Rhine	SWAsia	WaterBuffalo	0.1402	0.0051	27.36	1238039
YoA	Rhine	SWAsia	WaterBuffalo	0.1372	0.0046	29.53	2032877
Gallo1	Rhine	SWAsia	WaterBuffalo	0.1372	0.0111	12.36	119303
Lun2	Rhine	SWAsia	WaterBuffalo	0.1370	0.0117	11.76	129734
Ska3	Rhine	SWAsia	WaterBuffalo	0.1346	0.0050	27.20	1775327
Lun1	Rhine	SWAsia	WaterBuffalo	0.1341	0.0066	20.35	381166
Tri1	Rhine	SWAsia	WaterBuffalo	0.1325	0.0045	29.28	1766671
Bed4	Rhine	SWAsia	WaterBuffalo	0.1311	0.0051	25.77	1640810
Rhi3	Rhine	SWAsia	WaterBuffalo	0.1309	0.0045	29.25	1873155
Gal2	Rhine	SWAsia	WaterBuffalo	0.1298	0.0052	25.04	1025519
Cpc	Rhine	SWAsia	WaterBuffalo	0.1294	0.0047	27.41	1858539
Bed3	Rhine	SWAsia	WaterBuffalo	0.1268	0.0047	26.81	2037449
Hxh2	Rhine	SWAsia	WaterBuffalo	0.1264	0.0046	27.21	2040588

Fal7	Rhine	SWAsia	WaterBuffalo	0.1242	0.0058	21.37	1112721
Hjo1	Rhine	SWAsia	WaterBuffalo	0.1233	0.0050	24.59	1329828
Gal3	Rhine	SWAsia	WaterBuffalo	0.1200	0.0050	23.85	1904516
Gal1	Rhine	SWAsia	WaterBuffalo	0.1183	0.0049	24.03	2040553
Fif1	Rhine	SWAsia	WaterBuffalo	0.1133	0.0111	10.23	131899
Horn-10543	Rhine	SWAsia	WaterBuffalo	0.1060	0.0277	3.83	17805
Horn-CMLIX	Rhine	SWAsia	WaterBuffalo	0.0940	0.0223	4.22	30073
Horn-10537	Rhine	SWAsia	WaterBuffalo	0.0902	0.0123	7.32	92204

Appendix Table 9.8 - Relative allele-sharing of European aurochs with ancient domesticates.

P1	P2	P3	P4	*D* statistic	Standard Error	Z Score	Number of Sites
Ace1	SWAsia	Iberia	WaterBuffalo	0.01202	0.00790	1.52	211078
Bel1	SWAsia	Iberia	WaterBuffalo	0.03743	0.00948	3.95	168830
Bel2	SWAsia	Iberia	WaterBuffalo	0.03448	0.00825	4.18	256666
Bla1	SWAsia	Iberia	WaterBuffalo	-0.01935	0.01687	-1.15	36564
Bub1	SWAsia	Iberia	WaterBuffalo	-0.01628	0.00634	-2.57	615869
Ch22	SWAsia	Iberia	WaterBuffalo	0.02241	0.00617	3.63	546531
Che1	SWAsia	Iberia	WaterBuffalo	-0.00550	0.00617	-0.89	502782
Far1	SWAsia	Iberia	WaterBuffalo	-0.01340	0.01369	-0.98	57596
Gyu2	SWAsia	Iberia	WaterBuffalo	0.03454	0.00470	7.35	1533140
Has1	SWAsia	Iberia	WaterBuffalo	-0.01684	0.01309	-1.29	66793
Has3	SWAsia	Iberia	WaterBuffalo	0.00463	0.00477	0.97	1415606
Has4	SWAsia	Iberia	WaterBuffalo	-0.06693	0.01025	-6.53	966966
Has5	SWAsia	Iberia	WaterBuffalo	-0.03304	0.01255	-2.63	69962
Horn-382	SWAsia	Iberia	WaterBuffalo	-0.01852	0.00798	-2.32	224846
Horn-617	SWAsia	Iberia	WaterBuffalo	-0.02883	0.00579	-4.98	568996
Horn11570	SWAsia	Iberia	WaterBuffalo	0.01099	0.00489	2.25	1091565
Mar1	SWAsia	Iberia	WaterBuffalo	-0.01966	0.01184	-1.66	75041
Plo1	SWAsia	Iberia	WaterBuffalo	0.03748	0.00813	4.61	246632
Plo2	SWAsia	Iberia	WaterBuffalo	0.02531	0.00601	4.21	747549
Plo3	SWAsia	Iberia	WaterBuffalo	0.01930	0.00524	3.68	1918986
Plo4	SWAsia	Iberia	WaterBuffalo	0.02031	0.00479	4.24	1956107
Plo5	SWAsia	Iberia	WaterBuffalo	0.01469	0.00850	1.73	167479

Plo6	SWAsia	Iberia	WaterBuffalo	0.00886	0.00765	1.16	237332
Plo7	SWAsia	Iberia	WaterBuffalo	0.00781	0.00797	0.98	212090
Plo8	SWAsia	Iberia	WaterBuffalo	0.01746	0.00674	2.59	366621
Qaz1	SWAsia	Iberia	WaterBuffalo	0.00180	0.00476	0.38	1480630
Sac3	SWAsia	Iberia	WaterBuffalo	-0.02160	0.01295	-1.67	75768
Sar38	SWAsia	Iberia	WaterBuffalo	-0.02577	0.00557	-4.63	778547
Shi1	SWAsia	Iberia	WaterBuffalo	-0.00280	0.00645	-0.43	673785
Shimao	SWAsia	Iberia	WaterBuffalo	-0.04356	0.00509	-8.55	1317734
Stu1	SWAsia	Iberia	WaterBuffalo	0.02329	0.00737	3.16	275762
Tsa3	SWAsia	Iberia	WaterBuffalo	-0.08439	0.01798	-4.69	34715

Appendix Table 9.9 - Relative allele-sharing of Eastern aurochs with ancient domesticates.

P1	P2	P3	P4	*D* statistic	Standard Error	Z Score	Number of Sites
Ace1	SWAsia	East	WaterBuffalo	0.0025	0.0094	0.269	211050
Bel1	SWAsia	East	WaterBuffalo	-0.0166	0.0099	-1.687	168812
Bel2	SWAsia	East	WaterBuffalo	-0.0086	0.0089	-0.965	256627
Bla1	SWAsia	East	WaterBuffalo	-0.0237	0.0195	-1.216	36558
Bub1	SWAsia	East	WaterBuffalo	-0.0735	0.0064	-11.458	615780
Ch22	SWAsia	East	WaterBuffalo	-0.0014	0.0064	-0.225	546464
Che1	SWAsia	East	WaterBuffalo	-0.0353	0.0065	-5.429	502707
Far1	SWAsia	East	WaterBuffalo	-0.0106	0.0164	-0.642	57588
Gyu2	SWAsia	East	WaterBuffalo	0.0513	0.0052	9.940	1532911
Has1	SWAsia	East	WaterBuffalo	-0.0123	0.0155	-0.794	66780
Has3	SWAsia	East	WaterBuffalo	0.0149	0.0051	2.911	1415375
Has4	SWAsia	East	WaterBuffalo	-0.0537	0.0102	-5.259	966812
Has5	SWAsia	East	WaterBuffalo	-0.0376	0.0147	-2.558	69949
Horn-382	SWAsia	East	WaterBuffalo	-0.0896	0.0089	-10.036	224810
Horn-617	SWAsia	East	WaterBuffalo	-0.0999	0.0068	-14.597	568889
Horn11570	SWAsia	East	WaterBuffalo	-0.0594	0.0056	-10.669	1091363
Mar1	SWAsia	East	WaterBuffalo	0.0109	0.0142	0.769	75035
Plo1	SWAsia	East	WaterBuffalo	-0.0195	0.0095	-2.059	246595
Plo2	SWAsia	East	WaterBuffalo	-0.0082	0.0059	-1.376	747438
Plo3	SWAsia	East	WaterBuffalo	-0.0137	0.0053	-2.569	1918663

Plo4	SWAsia	East	WaterBuffalo	-0.0241	0.0051	-4.763	1955777
Plo5	SWAsia	East	WaterBuffalo	-0.0239	0.0099	-2.416	167451
Plo6	SWAsia	East	WaterBuffalo	-0.0319	0.0087	-3.654	237305
Plo7	SWAsia	East	WaterBuffalo	-0.0193	0.0087	-2.211	212060
Plo8	SWAsia	East	WaterBuffalo	-0.0174	0.0072	-2.434	366560
Qaz1	SWAsia	East	WaterBuffalo	0.0153	0.0051	2.985	1480406
Sac3	SWAsia	East	WaterBuffalo	-0.0130	0.0137	-0.950	75755
Sar38	SWAsia	East	WaterBuffalo	-0.0299	0.0065	-4.566	778437
Shi1	SWAsia	East	WaterBuffalo	0.0637	0.0076	8.425	673688
Shimao	SWAsia	East	WaterBuffalo	0.0225	0.0072	3.133	1317527
Stu1	SWAsia	East	WaterBuffalo	-0.0253	0.0083	-3.032	275713
Tsa3	SWAsia	East	WaterBuffalo	-0.0414	0.0206	-2.011	34713

Appendix Table 9.10 - Relative allele-sharing of indicine with ancient domesticates.

P1	P2	P3	P4	*D* statistic	Standard Error	Z Score	Number of Sites
Ace1	SWAsia	Indicine	WaterBuffalo	-0.0083	0.0092	-0.89	211030
Bel1	SWAsia	Indicine	WaterBuffalo	-0.0306	0.0101	-3.03	168804
Bel2	SWAsia	Indicine	WaterBuffalo	-0.0320	0.0086	-3.71	256612
Bla1	SWAsia	Indicine	WaterBuffalo	-0.0678	0.0197	-3.44	36559
Bub1	SWAsia	Indicine	WaterBuffalo	-0.0780	0.0065	-12.03	615756
Ch22	SWAsia	Indicine	WaterBuffalo	-0.0135	0.0067	-2.01	546433
Che1	SWAsia	Indicine	WaterBuffalo	-0.0330	0.0063	-5.26	502676
Far1	SWAsia	Indicine	WaterBuffalo	0.0011	0.0164	0.06	57585
Gyu2	SWAsia	Indicine	WaterBuffalo	0.0035	0.0046	0.76	1532805
Has1	SWAsia	Indicine	WaterBuffalo	0.0910	0.0184	4.95	66775
Has3	SWAsia	Indicine	WaterBuffalo	-0.0016	0.0046	-0.35	1415287
Has4	SWAsia	Indicine	WaterBuffalo	0.1021	0.0137	7.43	966754
Has5	SWAsia	Indicine	WaterBuffalo	0.0564	0.0153	3.69	69947
Horn-382	SWAsia	Indicine	WaterBuffalo	-0.0870	0.0092	-9.44	224793
Horn-617	SWAsia	Indicine	WaterBuffalo	-0.1005	0.0068	-14.75	568864
Horn11570	SWAsia	Indicine	WaterBuffalo	-0.0726	0.0053	-13.78	1091354
Mar1	SWAsia	Indicine	WaterBuffalo	-0.0108	0.0141	-0.77	75027
Plo1	SWAsia	Indicine	WaterBuffalo	-0.0355	0.0090	-3.96	246571
Plo2	SWAsia	Indicine	WaterBuffalo	-0.0273	0.0059	-4.62	747393

Plo3	SWAsia	Indicine	WaterBuffalo	-0.0214	0.0046	-4.69	1918593
Plo4	SWAsia	Indicine	WaterBuffalo	-0.0286	0.0045	-6.33	1955702
Plo5	SWAsia	Indicine	WaterBuffalo	-0.0281	0.0100	-2.82	167445
Plo6	SWAsia	Indicine	WaterBuffalo	-0.0338	0.0086	-3.91	237274
Plo7	SWAsia	Indicine	WaterBuffalo	-0.0299	0.0088	-3.40	212046
Plo8	SWAsia	Indicine	WaterBuffalo	-0.0351	0.0071	-4.92	366541
Qaz1	SWAsia	Indicine	WaterBuffalo	0.0111	0.0047	2.37	1480307
Sac3	SWAsia	Indicine	WaterBuffalo	-0.0157	0.0148	-1.07	75757
Sar38	SWAsia	Indicine	WaterBuffalo	-0.0425	0.0063	-6.79	778393
Shi1	SWAsia	Indicine	WaterBuffalo	0.0181	0.0072	2.51	673653
Shimao	SWAsia	Indicine	WaterBuffalo	-0.0499	0.0055	-8.99	1317484
Stu1	SWAsia	Indicine	WaterBuffalo	-0.0427	0.0087	-4.92	275693
Tsa3	SWAsia	Indicine	WaterBuffalo	0.1021	0.0220	4.64	34710

Appendix Table 9.11 - Relative allele-sharing of Th7 with ancient domesticates.

P1	P2	P3	P4	*D* statistic	Standard Error	Z Score	Number of Sites
Ace1	SWAsia	Th7	WaterBuffalo	0.0057	0.0347	0.16	14129
Bel1	SWAsia	Th7	WaterBuffalo	-0.0818	0.0346	-2.36	12583
Bel2	SWAsia	Th7	WaterBuffalo	-0.0200	0.0285	-0.70	18409
Bla1	SWAsia	Th7	WaterBuffalo	0.0000	0.0888	0.00	2476
Bub1	SWAsia	Th7	WaterBuffalo	-0.0319	0.0203	-1.57	42654
Ch22	SWAsia	Th7	WaterBuffalo	-0.0005	0.0204	-0.03	38096
Che1	SWAsia	Th7	WaterBuffalo	-0.0173	0.0225	-0.77	35191
Far1	SWAsia	Th7	WaterBuffalo	0.0358	0.0669	0.54	3963
Gyu2	SWAsia	Th7	WaterBuffalo	0.0007	0.0129	0.06	98819
Has1	SWAsia	Th7	WaterBuffalo	0.0460	0.0576	0.80	4660
Has3	SWAsia	Th7	WaterBuffalo	-0.0094	0.0144	-0.65	93530
Has4	SWAsia	Th7	WaterBuffalo	-0.0893	0.0177	-5.03	65549
Has5	SWAsia	Th7	WaterBuffalo	-0.0706	0.0564	-1.25	4840
Horn-382	SWAsia	Th7	WaterBuffalo	-0.0446	0.0333	-1.34	15593
Horn-617	SWAsia	Th7	WaterBuffalo	-0.0343	0.0232	-1.48	36343
Horn11570	SWAsia	Th7	WaterBuffalo	-0.0580	0.0161	-3.60	69478
Mar1	SWAsia	Th7	WaterBuffalo	0.0240	0.0572	0.42	5182
Plo1	SWAsia	Th7	WaterBuffalo	-0.0401	0.0318	-1.26	17401

Plo2	SWAsia	Th7	WaterBuffalo	0.0044	0.0181	0.24	52637
Plo3	SWAsia	Th7	WaterBuffalo	-0.0219	0.0124	-1.77	125027
Plo4	SWAsia	Th7	WaterBuffalo	-0.0258	0.0121	-2.13	127096
Plo5	SWAsia	Th7	WaterBuffalo	-0.0151	0.0359	-0.42	12225
Plo6	SWAsia	Th7	WaterBuffalo	-0.0027	0.0295	-0.09	17193
Plo7	SWAsia	Th7	WaterBuffalo	-0.0409	0.0320	-1.28	14936
Plo8	SWAsia	Th7	WaterBuffalo	-0.0433	0.0247	-1.75	26292
Qaz1	SWAsia	Th7	WaterBuffalo	0.0023	0.0132	0.18	98250
Sac3	SWAsia	Th7	WaterBuffalo	0.0033	0.0563	0.06	5117
Sar38	SWAsia	Th7	WaterBuffalo	-0.0218	0.0190	-1.15	51887
Shi1	SWAsia	Th7	WaterBuffalo	0.0020	0.0196	0.10	46365
Shimao	SWAsia	Th7	WaterBuffalo	-0.0868	0.0140	-6.18	87906
Stu1	SWAsia	Th7	WaterBuffalo	-0.0491	0.0280	-1.75	19415
Tsa3	SWAsia	Th7	WaterBuffalo	0.0037	0.0805	0.05	2670

Appendix Table 9.12 - Excess allele-sharing of modern breeds with indicine cattle.

P1	P2	P3	P4	*D* statistic	Standard Error	Z Score	Number of Sites
Alentejana	SWAsia	Indicine	WaterBuffalo	-0.024	0.005	-4.79	2112961
Angus	SWAsia	Indicine	WaterBuffalo	-0.060	0.005	-12.01	2112150
Baoule	SWAsia	Indicine	WaterBuffalo	0.031	0.007	4.54	2113310
Bashan	SWAsia	Indicine	WaterBuffalo	0.319	0.010	32.76	2079674
BelgianBlue	SWAsia	Indicine	WaterBuffalo	-0.033	0.005	-6.60	2107053
BeltedGalloway	SWAsia	Indicine	WaterBuffalo	-0.062	0.005	-12.60	2110905
BlondeDAquitaine	SWAsia	Indicine	WaterBuffalo	-0.032	0.005	-6.55	2097491
Bohai	SWAsia	Indicine	WaterBuffalo	0.176	0.011	16.09	2110939
Boran	SWAsia	Indicine	WaterBuffalo	0.465	0.008	61.74	2096600
BushaB	SWAsia	Indicine	WaterBuffalo	0.073	0.007	10.35	1917867
Chaidamu	SWAsia	Indicine	WaterBuffalo	0.161	0.010	16.83	2111166
Charolais	SWAsia	Indicine	WaterBuffalo	0.005	0.010	0.52	2110727
Chianina	SWAsia	Indicine	WaterBuffalo	0.110	0.009	12.41	2110881
DanishRed	SWAsia	Indicine	WaterBuffalo	-0.040	0.005	-7.51	2108796
Devon	SWAsia	Indicine	WaterBuffalo	-0.056	0.005	-10.83	2111821
Dexter	SWAsia	Indicine	WaterBuffalo	-0.043	0.007	-6.35	2110684
Eringer	SWAsia	Indicine	WaterBuffalo	-0.027	0.006	-4.38	2113086

Finncattle	SWAsia	Indicine	WaterBuffalo	-0.048	0.004	-11.44	2099641
Gelbvieh	SWAsia	Indicine	WaterBuffalo	-0.038	0.006	-6.47	2097854
Gourounsi	SWAsia	Indicine	WaterBuffalo	0.178	0.012	15.22	2113169
Hanwoo	SWAsia	Indicine	WaterBuffalo	0.019	0.006	2.94	2106682
Hereford	SWAsia	Indicine	WaterBuffalo	-0.121	0.006	-19.11	2113331
Holstein	SWAsia	Indicine	WaterBuffalo	-0.046	0.005	-8.88	2112817
Jersey	SWAsia	Indicine	WaterBuffalo	-0.048	0.006	-8.44	2107840
Kazakh	SWAsia	Indicine	WaterBuffalo	0.086	0.009	9.16	2105531
Kenana	SWAsia	Indicine	WaterBuffalo	0.466	0.008	59.72	2100622
Kuchinoshima	SWAsia	Indicine	WaterBuffalo	0.024	0.007	3.59	2080214
Lagune	SWAsia	Indicine	WaterBuffalo	0.110	0.011	10.10	2111982
Limia	SWAsia	Indicine	WaterBuffalo	-0.036	0.005	-7.68	2108121
Limousin	SWAsia	Indicine	WaterBuffalo	-0.045	0.005	-9.49	2020284
Lingnan	SWAsia	Indicine	WaterBuffalo	0.344	0.010	34.56	2100684
Luxi	SWAsia	Indicine	WaterBuffalo	0.272	0.011	25.75	2108152
MaineAnjou	SWAsia	Indicine	WaterBuffalo	-0.049	0.005	-9.77	2111893
Maltese	SWAsia	Indicine	WaterBuffalo	0.123	0.011	11.13	2089747
Maremmana	SWAsia	Indicine	WaterBuffalo	0.116	0.010	12.17	2108165
Maronesa	SWAsia	Indicine	WaterBuffalo	-0.038	0.005	-7.54	1724176
Mishima	SWAsia	Indicine	WaterBuffalo	0.018	0.006	2.99	2084634
Mongolian	SWAsia	Indicine	WaterBuffalo	0.244	0.010	24.12	2098906
Montbeliarde	SWAsia	Indicine	WaterBuffalo	-0.024	0.006	-4.37	2088747
Muturu	SWAsia	Indicine	WaterBuffalo	0.023	0.006	3.78	2086579
Ndama	SWAsia	Indicine	WaterBuffalo	0.221	0.012	17.86	2105759
Normande	SWAsia	Indicine	WaterBuffalo	-0.036	0.005	-6.76	2089678
Ogaden	SWAsia	Indicine	WaterBuffalo	0.477	0.008	62.83	2098308
OriginalBraunvieh	SWAsia	Indicine	WaterBuffalo	-0.020	0.006	-3.48	2112427
Pajuna	SWAsia	Indicine	WaterBuffalo	-0.019	0.005	-3.71	2108167
Piedmontese	SWAsia	Indicine	WaterBuffalo	-0.016	0.006	-2.42	2100183
Podolica	SWAsia	Indicine	WaterBuffalo	0.125	0.009	14.70	2106719
Rashoki	SWAsia	Indicine	WaterBuffalo	0.339	0.009	36.47	2109734
Sayaguesa	SWAsia	Indicine	WaterBuffalo	-0.027	0.006	-4.96	2108397
ScottishHighland	SWAsia	Indicine	WaterBuffalo	-0.048	0.005	-9.78	2113426

Shorthorn	SWAsia	Indicine	WaterBuffalo	-0.051	0.005	-9.31	2107560
Sikia	SWAsia	Indicine	WaterBuffalo	0.156	0.010	15.08	2112597
Simmental	SWAsia	Indicine	WaterBuffalo	-0.030	0.005	-5.55	2112491
Somba	SWAsia	Indicine	WaterBuffalo	0.071	0.009	7.61	2113224
SouthAnatolianRed	SWAsia	Indicine	WaterBuffalo	0.325	0.009	37.67	2091846
SwedishRed	SWAsia	Indicine	WaterBuffalo	-0.046	0.005	-9.17	2095571
TirolerGrauvieh	SWAsia	Indicine	WaterBuffalo	-0.016	0.005	-2.88	2112696
Wagyu	SWAsia	Indicine	WaterBuffalo	-0.006	0.006	-1.08	2112726
Xizang	SWAsia	Indicine	WaterBuffalo	0.048	0.008	5.81	2105317
Yakut	SWAsia	Indicine	WaterBuffalo	0.094	0.009	10.29	2105944
Yanbian	SWAsia	Indicine	WaterBuffalo	0.036	0.007	5.35	2109827

Appendix Table 9.13 - Excess allele-sharing of modern breeds with Eastern aurochs.

P1	P2	P3	P4	*D* statistic	Standard Error	Z Score	Number of Sites
Alentejana	SWAsia	East	WaterBuffalo	-0.0431	0.0055	-7.88	2113007
Angus	SWAsia	East	WaterBuffalo	-0.0701	0.0057	-12.35	2112199
Baoule	SWAsia	East	WaterBuffalo	0.0059	0.0056	1.06	2113357
Bashan	SWAsia	East	WaterBuffalo	-0.3731	0.0074	-50.32	2079665
BelgianBlue	SWAsia	East	WaterBuffalo	-0.0473	0.0054	-8.69	2107108
BeltedGalloway	SWAsia	East	WaterBuffalo	-0.0613	0.0052	-11.81	2110959
BlondeDAquitaine	SWAsia	East	WaterBuffalo	-0.0545	0.0050	-11.00	2097542
Bohai	SWAsia	East	WaterBuffalo	-0.2226	0.0087	-25.52	2110983
Boran	SWAsia	East	WaterBuffalo	-0.4219	0.0061	-68.97	2096509
BushaB	SWAsia	East	WaterBuffalo	-0.0923	0.0057	-16.10	1917913
Chaidamu	SWAsia	East	WaterBuffalo	-0.0993	0.0069	-14.41	2111207
Charolais	SWAsia	East	WaterBuffalo	-0.0736	0.0065	-11.41	2110774
Chianina	SWAsia	East	WaterBuffalo	-0.0908	0.0062	-14.75	2110926
DanishRed	SWAsia	East	WaterBuffalo	-0.0555	0.0049	-11.35	2108847
Devon	SWAsia	East	WaterBuffalo	-0.0640	0.0055	-11.59	2111877
Dexter	SWAsia	East	WaterBuffalo	-0.0661	0.0052	-12.62	2110735
Eringer	SWAsia	East	WaterBuffalo	-0.0502	0.0054	-9.34	2113134
Finncattle	SWAsia	East	WaterBuffalo	-0.0420	0.0051	-8.25	2099689
Gelbvieh	SWAsia	East	WaterBuffalo	-0.0533	0.0054	-9.78	2097908
Gourounsi	SWAsia	East	WaterBuffalo	-0.0837	0.0079	-10.58	2113209

Hanwoo	SWAsia	East	WaterBuffalo	0.0237	0.0064	3.73	2106740
Hereford	SWAsia	East	WaterBuffalo	-0.1286	0.0064	-20.24	2113386
Holstein	SWAsia	East	WaterBuffalo	-0.0478	0.0058	-8.26	2112868
Jersey	SWAsia	East	WaterBuffalo	-0.0556	0.0051	-10.79	2107898
Kazakh	SWAsia	East	WaterBuffalo	-0.0959	0.0068	-14.15	2105580
Kenana	SWAsia	East	WaterBuffalo	-0.4095	0.0059	-69.36	2100553
Kuchinoshima	SWAsia	East	WaterBuffalo	0.0550	0.0068	8.09	2080289
Lagune	SWAsia	East	WaterBuffalo	-0.0453	0.0078	-5.84	2112000
Limia	SWAsia	East	WaterBuffalo	-0.0376	0.0049	-7.63	2108163
Limousin	SWAsia	East	WaterBuffalo	-0.0578	0.0050	-11.58	2020332
Lingnan	SWAsia	East	WaterBuffalo	-0.4153	0.0070	-59.43	2100641
Luxi	SWAsia	East	WaterBuffalo	-0.3595	0.0075	-48.10	2108139
MaineAnjou	SWAsia	East	WaterBuffalo	-0.0557	0.0053	-10.49	2111947
Maltese	SWAsia	East	WaterBuffalo	-0.1037	0.0071	-14.60	2089777
Maremmana	SWAsia	East	WaterBuffalo	-0.0984	0.0064	-15.34	2108205
Maronesa	SWAsia	East	WaterBuffalo	-0.0523	0.0052	-10.15	1724216
Mishima	SWAsia	East	WaterBuffalo	0.0411	0.0065	6.34	2084695
Mongolian	SWAsia	East	WaterBuffalo	-0.1769	0.0081	-21.86	2098939
Montbeliarde	SWAsia	East	WaterBuffalo	-0.0319	0.0058	-5.46	2088808
Muturu	SWAsia	East	WaterBuffalo	0.0140	0.0056	2.50	2086629
Ndama	SWAsia	East	WaterBuffalo	-0.1289	0.0096	-13.46	2105786
Normande	SWAsia	East	WaterBuffalo	-0.0385	0.0054	-7.08	2089736
Ogaden	SWAsia	East	WaterBuffalo	-0.4265	0.0058	-73.35	2098203
OriginalBraunvieh	SWAsia	East	WaterBuffalo	-0.0474	0.0053	-8.88	2112469
Pajuna	SWAsia	East	WaterBuffalo	-0.0342	0.0056	-6.08	2108220
Piedmontese	SWAsia	East	WaterBuffalo	-0.0505	0.0052	-9.76	2100232
Podolica	SWAsia	East	WaterBuffalo	-0.1081	0.0064	-16.78	2106752
Rashoki	SWAsia	East	WaterBuffalo	-0.2739	0.0070	-39.15	2109740
Sayaguesa	SWAsia	East	WaterBuffalo	-0.0396	0.0057	-6.92	2108439
ScottishHighland	SWAsia	East	WaterBuffalo	-0.0573	0.0056	-10.17	2113476
Shorthorn	SWAsia	East	WaterBuffalo	-0.0568	0.0055	-10.31	2107607
Sikia	SWAsia	East	WaterBuffalo	-0.1114	0.0070	-15.96	2112635
Simmental	SWAsia	East	WaterBuffalo	-0.0484	0.0055	-8.74	2112544

Somba	SWAsia	East	WaterBuffalo	-0.0145	0.0063	-2.30	2113275
SouthAnatolianRed	SWAsia	East	WaterBuffalo	-0.2365	0.0069	-34.12	2091845
SwedishRed	SWAsia	East	WaterBuffalo	-0.0501	0.0052	-9.61	2095633
TirolerGrauvieh	SWAsia	East	WaterBuffalo	-0.0508	0.0053	-9.65	2112746
Wagyu	SWAsia	East	WaterBuffalo	0.0173	0.0059	2.93	2112773
Xizang	SWAsia	East	WaterBuffalo	0.0481	0.0087	5.53	2105379
Yakut	SWAsia	East	WaterBuffalo	-0.0420	0.0076	-5.50	2106000
Yanbian	SWAsia	East	WaterBuffalo	0.0057	0.0059	0.96	2109897

Appendix 9.14 - Excess allele-sharing of modern breeds with European aurochs

P1	P2	P3	P4	*D* statistic	Standard Error	Z Score	Number of Sites
Alentejana	SWAsia	European aurochs	WaterBuffalo	0.0247	0.0049	5.05	2113394
Angus	SWAsia	European aurochs	WaterBuffalo	0.0209	0.0053	3.96	2112586
Baoule	SWAsia	European aurochs	WaterBuffalo	-0.0622	0.0052	-11.92	2113738
Bashan	SWAsia	European aurochs	WaterBuffalo	-0.4883	0.0071	-68.61	2080030
BelgianBlue	SWAsia	European aurochs	WaterBuffalo	0.0104	0.0057	1.84	2107485
BeltedGalloway	SWAsia	European aurochs	WaterBuffalo	0.0262	0.0049	5.35	2111341
BlondeDAquitaine	SWAsia	European aurochs	WaterBuffalo	0.0214	0.0050	4.30	2097924
Bohai	SWAsia	European aurochs	WaterBuffalo	-0.2284	0.0101	-22.52	2111363
Boran	SWAsia	European aurochs	WaterBuffalo	-0.5399	0.0065	-83.69	2096868
BushaB	SWAsia	European aurochs	WaterBuffalo	-0.0634	0.0057	-11.14	1918250
Chaidamu	SWAsia	European aurochs	WaterBuffalo	-0.1448	0.0077	-18.76	2111590
Charolais	SWAsia	European aurochs	WaterBuffalo	-0.0133	0.0074	-1.80	2111159
Chianina	SWAsia	European aurochs	WaterBuffalo	-0.0838	0.0066	-12.75	2111311
DanishRed	SWAsia	European aurochs	WaterBuffalo	0.0255	0.0048	5.26	2109228
Devon	SWAsia	European aurochs	WaterBuffalo	0.0273	0.0052	5.22	2112257
Dexter	SWAsia	European aurochs	WaterBuffalo	0.0171	0.0056	3.03	2111117
Eringer	SWAsia	European aurochs	WaterBuffalo	0.0120	0.0051	2.36	2113518
Finncattle	SWAsia	European aurochs	WaterBuffalo	0.0316	0.0047	6.69	2100072
Gelbvieh	SWAsia	European aurochs	WaterBuffalo	0.0096	0.0046	2.07	2098291
Gourounsi	SWAsia	European aurochs	WaterBuffalo	-0.1666	0.0085	-19.49	2113585
Hanwoo	SWAsia	European aurochs	WaterBuffalo	-0.0406	0.0060	-6.78	2107115
Hereford	SWAsia	European aurochs	WaterBuffalo	-0.0408	0.0060	-6.75	2113774
Holstein	SWAsia	European aurochs	WaterBuffalo	0.0294	0.0050	5.93	2113253

Jersey	SWAsia	European aurochs	WaterBuffalo	0.0228	0.0050	4.57	2108273
Kazakh	SWAsia	European aurochs	WaterBuffalo	-0.0884	0.0078	-11.34	2105954
Kenana	SWAsia	European aurochs	WaterBuffalo	-0.5381	0.0062	-87.11	2100914
Kuchinoshima	SWAsia	European aurochs	WaterBuffalo	-0.0121	0.0052	-2.32	2080652
Lagune	SWAsia	European aurochs	WaterBuffalo	-0.1223	0.0079	-15.56	2112383
Limia	SWAsia	European aurochs	WaterBuffalo	0.0226	0.0049	4.58	2108549
Limousin	SWAsia	European aurochs	WaterBuffalo	0.0145	0.0048	3.01	2020686
Lingnan	SWAsia	European aurochs	WaterBuffalo	-0.5373	0.0066	-80.81	2101007
Luxi	SWAsia	European aurochs	WaterBuffalo	-0.4308	0.0086	-50.28	2108525
MaineAnjou	SWAsia	European aurochs	WaterBuffalo	0.0211	0.0050	4.18	2112331
Maltese	SWAsia	European aurochs	WaterBuffalo	-0.0914	0.0077	-11.90	2090144
Maremmana	SWAsia	European aurochs	WaterBuffalo	-0.0931	0.0070	-13.25	2108589
Maronesa	SWAsia	European aurochs	WaterBuffalo	0.0022	0.0048	0.47	1724522
Mishima	SWAsia	European aurochs	WaterBuffalo	-0.0257	0.0057	-4.50	2085051
Mongolian	SWAsia	European aurochs	WaterBuffalo	-0.2290	0.0086	-26.51	2099312
Montbeliarde	SWAsia	European aurochs	WaterBuffalo	0.0128	0.0056	2.28	2089181
Muturu	SWAsia	European aurochs	WaterBuffalo	-0.0508	0.0051	-10.02	2086992
Ndama	SWAsia	European aurochs	WaterBuffalo	-0.2124	0.0097	-21.93	2106161
Normande	SWAsia	European aurochs	WaterBuffalo	0.0217	0.0049	4.46	2090113
Ogaden	SWAsia	European aurochs	WaterBuffalo	-0.5584	0.0059	-95.14	2098575
OriginalBraunvieh	SWAsia	European aurochs	WaterBuffalo	0.0049	0.0052	0.94	2112858
Pajuna	SWAsia	European aurochs	WaterBuffalo	0.0184	0.0053	3.48	2108595
Piedmontese	SWAsia	European aurochs	WaterBuffalo	0.0069	0.0053	1.29	2100610
Podolica	SWAsia	European aurochs	WaterBuffalo	-0.1017	0.0070	-14.47	2107138
Rashoki	SWAsia	European aurochs	WaterBuffalo	-0.3351	0.0084	-39.93	2110119
Sayaguesa	SWAsia	European aurochs	WaterBuffalo	0.0250	0.0048	5.20	2108821
ScottishHighland	SWAsia	European aurochs	WaterBuffalo	0.0236	0.0050	4.75	2113860
Shorthorn	SWAsia	European aurochs	WaterBuffalo	0.0150	0.0051	2.93	2107992
Sikia	SWAsia	European aurochs	WaterBuffalo	-0.1140	0.0078	-14.55	2113019
Simmental	SWAsia	European aurochs	WaterBuffalo	0.0050	0.0049	1.02	2112927
Somba	SWAsia	European aurochs	WaterBuffalo	-0.0824	0.0064	-12.94	2113653
SouthAnatolianRed	SWAsia	European aurochs	WaterBuffalo	-0.3130	0.0078	-40.07	2092208
SwedishRed	SWAsia	European aurochs	WaterBuffalo	0.0265	0.0050	5.32	2096005

TirolerGrauvieh	SWAsia	European aurochs	WaterBuffalo	0.0152	0.0049	3.12	2113128
Wagyu	SWAsia	European aurochs	WaterBuffalo	-0.0165	0.0049	-3.36	2113155
Xizang	SWAsia	European aurochs	WaterBuffalo	-0.1257	0.0076	-16.54	2105735
Yakut	SWAsia	European aurochs	WaterBuffalo	-0.1000	0.0077	-12.95	2106368
Yanbian	SWAsia	European aurochs	WaterBuffalo	-0.0526	0.0056	-9.45	2110266

Table 9.15 - Excess allele-sharing of modern breeds with Th7.

P1	P2	P3	P4	*D* statistic	Standard Error	Z Score	Number of Sites
Alentejana	SWAsia	Th7	WaterBuffalo	-0.0525	0.0114	-4.62	134515
Angus	SWAsia	Th7	WaterBuffalo	-0.0798	0.0114	-7.02	134452
Baoule	SWAsia	Th7	WaterBuffalo	0.0623	0.0115	5.42	134530
Bashan	SWAsia	Th7	WaterBuffalo	-0.4913	0.0106	-46.46	132281
BelgianBlue	SWAsia	Th7	WaterBuffalo	-0.0504	0.0114	-4.44	134160
BeltedGalloway	SWAsia	Th7	WaterBuffalo	-0.0744	0.0124	-6.01	134410
BlondeDAquitaine	SWAsia	Th7	WaterBuffalo	-0.0487	0.0120	-4.05	133713
Bohai	SWAsia	Th7	WaterBuffalo	-0.2960	0.0131	-22.60	134400
Boran	SWAsia	Th7	WaterBuffalo	-0.5112	0.0102	-50.20	133509
BushaB	SWAsia	Th7	WaterBuffalo	-0.1336	0.0120	-11.17	122403
Chaidamu	SWAsia	Th7	WaterBuffalo	-0.1808	0.0127	-14.28	134403
Charolais	SWAsia	Th7	WaterBuffalo	-0.0927	0.0116	-8.01	134396
Chianina	SWAsia	Th7	WaterBuffalo	-0.1388	0.0123	-11.31	134423
DanishRed	SWAsia	Th7	WaterBuffalo	-0.0727	0.0115	-6.32	134239
Devon	SWAsia	Th7	WaterBuffalo	-0.0833	0.0117	-7.09	134449
Dexter	SWAsia	Th7	WaterBuffalo	-0.0762	0.0121	-6.31	134404
Eringer	SWAsia	Th7	WaterBuffalo	-0.0575	0.0116	-4.95	134514
Finncattle	SWAsia	Th7	WaterBuffalo	-0.0514	0.0109	-4.70	133844
Gelbvieh	SWAsia	Th7	WaterBuffalo	-0.0713	0.0112	-6.34	133900
Gourounsi	SWAsia	Th7	WaterBuffalo	-0.0488	0.0139	-3.50	134524
Hanwoo	SWAsia	Th7	WaterBuffalo	-0.0750	0.0118	-6.36	134235
Hereford	SWAsia	Th7	WaterBuffalo	-0.1526	0.0114	-13.37	134544
Holstein	SWAsia	Th7	WaterBuffalo	-0.0592	0.0113	-5.23	134507
Jersey	SWAsia	Th7	WaterBuffalo	-0.0842	0.0114	-7.37	134300
Kazakh	SWAsia	Th7	WaterBuffalo	-0.1452	0.0117	-12.42	134059
Kenana	SWAsia	Th7	WaterBuffalo	-0.5156	0.0098	-52.42	133803

Kuchinoshima	SWAsia	Th7	WaterBuffalo	-0.0396	0.0121	-3.26	132349
Lagune	SWAsia	Th7	WaterBuffalo	0.0006	0.0142	0.04	134467
Limia	SWAsia	Th7	WaterBuffalo	-0.0448	0.0114	-3.94	134229
Limousin	SWAsia	Th7	WaterBuffalo	-0.0475	0.0124	-3.83	128873
Lingnan	SWAsia	Th7	WaterBuffalo	-0.5366	0.0096	-55.75	133707
Luxi	SWAsia	Th7	WaterBuffalo	-0.4429	0.0115	-38.46	134218
MaineAnjou	SWAsia	Th7	WaterBuffalo	-0.0672	0.0115	-5.83	134477
Maltese	SWAsia	Th7	WaterBuffalo	-0.1294	0.0125	-10.32	133088
Maremmana	SWAsia	Th7	WaterBuffalo	-0.1265	0.0124	-10.23	134241
Maronesa	SWAsia	Th7	WaterBuffalo	-0.0595	0.0130	-4.57	109094
Mishima	SWAsia	Th7	WaterBuffalo	-0.0552	0.0124	-4.44	132671
Mongolian	SWAsia	Th7	WaterBuffalo	-0.2629	0.0123	-21.30	133581
Montbeliarde	SWAsia	Th7	WaterBuffalo	-0.0668	0.0116	-5.75	133149
Muturu	SWAsia	Th7	WaterBuffalo	0.0688	0.0123	5.59	132862
Ndama	SWAsia	Th7	WaterBuffalo	-0.1053	0.0156	-6.76	134130
Normande	SWAsia	Th7	WaterBuffalo	-0.0627	0.0118	-5.31	133240
Ogaden	SWAsia	Th7	WaterBuffalo	-0.5300	0.0099	-53.52	133669
OriginalBraunvieh	SWAsia	Th7	WaterBuffalo	-0.0715	0.0114	-6.29	134447
Pajuna	SWAsia	Th7	WaterBuffalo	-0.0375	0.0112	-3.36	134252
Piedmontese	SWAsia	Th7	WaterBuffalo	-0.0589	0.0118	-4.98	133853
Podolica	SWAsia	Th7	WaterBuffalo	-0.1490	0.0122	-12.21	134135
Rashoki	SWAsia	Th7	WaterBuffalo	-0.3460	0.0119	-29.15	134298
Sayaguesa	SWAsia	Th7	WaterBuffalo	-0.0541	0.0115	-4.70	134249
ScottishHighland	SWAsia	Th7	WaterBuffalo	-0.0667	0.0119	-5.60	134545
Shorthorn	SWAsia	Th7	WaterBuffalo	-0.0843	0.0115	-7.35	134256
Sikia	SWAsia	Th7	WaterBuffalo	-0.1619	0.0123	-13.18	134492
Simmental	SWAsia	Th7	WaterBuffalo	-0.0689	0.0113	-6.12	134491
Somba	SWAsia	Th7	WaterBuffalo	0.0427	0.0130	3.28	134528
SouthAnatolianRed	SWAsia	Th7	WaterBuffalo	-0.2669	0.0126	-21.25	133342
SwedishRed	SWAsia	Th7	WaterBuffalo	-0.0762	0.0116	-6.57	133631
TirolerGrauvieh	SWAsia	Th7	WaterBuffalo	-0.0732	0.0111	-6.61	134493
Wagyu	SWAsia	Th7	WaterBuffalo	-0.0549	0.0111	-4.94	134494
Xizang	SWAsia	Th7	WaterBuffalo	-0.1437	0.0124	-11.59	134012

Yakut	SWAsia	Th7	WaterBuffalo	-0.1247	0.0125	-9.93	134139
Yanbian	SWAsia	Th7	WaterBuffalo	-0.0813	0.0113	-7.19	134341

Table 9.16 - ADMIXTURE cross-validation error.

Ancestral Components (K)	CV error
1	0.71838
2	0.70574
3	0.71174
4	0.72214
5	0.75367
6	0.79032
7	0.79026
8	0.87232
9	0.87635
10	0.93035

Table 9.17 - results of K=4 ADMIXTURE.

Sample	Europe	East	Indicine	Taurine	Group
Ace1	0.000012	0.029017	0.00001	0.970961	Early SW Asian domesticates
Alent	0.078076	0.00001	0.00001	0.921904	Modern Iberian taurine
Angus	0.08395	0.00001	0.00001	0.91603	Modern Central and European taurine
Baikal1	0.00001	0.99997	0.00001	0.00001	Pleistocene East Eurasian aurochs
Baoule	0.00001	0.047078	0.00001	0.952902	Modern West African taurine
Bar1	0.99997	0.00001	0.00001	0.00001	Italian Aurochs
Bashan	0.00001	0.033127	0.597585	0.369277	Modern East Asian hybrid
Bed3	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Pre-domestic contact
Bed4	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Pre-domestic contact
Bel1	0.088517	0.00001	0.00001	0.911463	Ancient European domesticate
Bel2	0.087331	0.00001	0.00001	0.912649	Ancient European domesticate
Belarus1	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Post-domestic contact
Belgian Blue	0.00001	0.00001	0.00001	0.99997	Modern Central and European taurine
Belted Galloway	0.098186	0.00001	0.00001	0.901794	Modern Central and European taurine
Bes1	0.00001	0.009694	0.420428	0.569868	Post-Bronze Age SW Asian domesticates
Bes2	0.00001	0.00001	0.825912	0.174068	Post-Bronze Age SW Asian domesticates

Blec1	0.75813	0.00001	0.00001	0.24185	Balkan aurochs
Blonde D'Aquitaine	0.020999	0.00001	0.00001	0.978981	Modern Central and European taurine
Bohai	0.003676	0.00001	0.225583	0.770731	Modern East Asian taurine
Bor1	0.00001	0.99997	0.00001	0.00001	Holocene Central Asian aurochs
Bor2	0.112767	0.706262	0.00001	0.180961	Holocene Central Asian aurochs
Bor3	0.02889	0.74847	0.00001	0.22263	Holocene Central Asian aurochs
Boran	0.00001	0.00001	0.789843	0.210137	Modern East African hybrid
Bub1	0.00001	0.00001	0.00001	0.99997	Ancient European domesticate
Busha	0.00001	0.00001	0.069269	0.930711	Modern South European taurine
Ch22	0.079343	0.006163	0.00001	0.914484	Holocene SW Asian aurochs
Chaidamu	0.00001	0.062692	0.137286	0.800012	Modern East Asian taurine
Charolais	0.00001	0.00001	0.00001	0.99997	Modern Central and European taurine
Che1	0.00001	0.00001	0.00001	0.99997	Ancient European domesticate
Chianina	0.00001	0.00001	0.077564	0.922416	Modern South European taurine
Cpc	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Pre-domestic contact
Danish Red	0.063733	0.00001	0.00001	0.936247	Modern Central and European taurine
Devon	0.105062	0.00001	0.00001	0.894918	Modern Central and European taurine
Dexter	0.097084	0.00001	0.00001	0.902896	Modern Central and European taurine
Eringer	0.00001	0.00001	0.00001	0.99997	Modern Central and European taurine
Fal7	0.916795	0.00001	0.00001	0.083185	C. and N. European aurochs - Post-domestic contact
Fif1	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Pre-domestic contact
Finncattle	0.077373	0.00001	0.00001	0.922607	Modern Central and European taurine
Gal1	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Pre-domestic contact
Gal2	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Pre-domestic contact
Gal3	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Pre-domestic contact
Gallo1	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Post-domestic contact
Gelbvieh	0.00001	0.00001	0.00001	0.99997	Modern Central and European taurine
Gir	0.00001	0.00001	0.99997	0.00001	Modern South Asian indicine
Gourounsi	0.00001	0.028063	0.136991	0.834937	Modern West African taurine
Gyu1	0.00001	0.99997	0.00001	0.00001	Pleistocene East Eurasian aurochs
Gyu2	0.225231	0.121763	0.00001	0.652996	Holocene SW Asian aurochs
Hanwoo	0.00001	0.145421	0.00001	0.854559	Modern East Asian taurine
Hariana	0.00001	0.00001	0.99997	0.00001	Modern South Asian indicine

Has3	0.00001	0.030194	0.00001	0.969786	Early SW Asian domesticates
Has4	0.00001	0.021693	0.052949	0.925347	Post-Bronze Age SW Asian domesticates
Hereford	0.098305	0.00001	0.00001	0.901675	Modern Central and European taurine
Highland	0.123368	0.00001	0.00001	0.876612	Modern Central and European taurine
Hjo1	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Pre-domestic contact
Holstein	0.08918	0.00001	0.00001	0.9108	Modern Central and European taurine
Horn-10533	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Post-domestic contact
Horn-10534	0.100934	0.00001	0.00001	0.899046	C. and N. European aurochs - Post-domestic contact
Horn-10535	0.096199	0.00001	0.00001	0.903781	C. and N. European aurochs - Post-domestic contact
Horn-10536	0.838928	0.00001	0.00001	0.161052	C. and N. European aurochs - Post-domestic contact
Horn-10537	0.983443	0.00001	0.00001	0.016537	C. and N. European aurochs - Post-domestic contact
Hxh2	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Post-domestic contact
Jersey	0.068935	0.00001	0.00001	0.931045	Modern Central and European taurine
Jian	0.000012	0.00001	0.999968	0.00001	Modern East Asian indicine
Jiangxi	0.00001	0.00001	0.99997	0.00001	Modern East Asian indicine
Jiaxian	0.00001	0.050193	0.460718	0.48908	Modern East Asian hybrid
Kalisz1	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Post-domestic contact
Kalisz2	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Post-domestic contact
Kalisz3	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Post-domestic contact
Kaz1	0.00001	0.015132	0.138608	0.84625	Post-Bronze Age SW Asian domesticates
Kaz2	0.00001	0.000144	0.249618	0.750228	Post-Bronze Age SW Asian domesticates
Kaz3	0.021086	0.00001	0.155685	0.823219	Post-Bronze Age SW Asian domesticates
Kaz4	0.00001	0.011969	0.205903	0.782118	Post-Bronze Age SW Asian domesticates
Kazakh	0.00001	0.015247	0.080105	0.904638	Modern Central Asian taurine
Kenana	0.00001	0.00001	0.808055	0.191925	Modern East African hybrid
Kok1	0.00001	0.00001	0.716001	0.283979	Ancient Central Asian domesticates
Kuchinoshima	0.00001	0.117918	0.00001	0.882062	Modern East Asian taurine
Lagune	0.00001	0.034917	0.076317	0.888756	Modern West African taurine
Lil1	0.921102	0.00001	0.00001	0.078878	C. and N. European aurochs - Post-domestic contact
Limia	0.072324	0.00001	0.00001	0.927656	Modern Iberian taurine

Limousin	0.00001	0.00001	0.00001	0.99997	Modern Central and European taurine
Lingnan	0.00001	0.014089	0.707099	0.278802	Modern East Asian hybrid
Lun1	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Pre-domestic contact
Luxi	0.00001	0.00001	0.476246	0.523734	Modern East Asian hybrid
Maine Anjou	0.034777	0.00001	0.00001	0.965203	Modern Central and European taurine
Maltese	0.00001	0.00001	0.080878	0.919102	Modern South European taurine
Maremmana	0.00001	0.00001	0.084883	0.915097	Modern South European taurine
Maronesa	0.045067	0.00001	0.00001	0.954913	Modern Iberian taurine
Men1	0.00001	0.00001	0.00001	0.99997	Early SW Asian domesticates
Men2	0.00001	0.00001	0.00001	0.99997	Early SW Asian domesticates
Mishima	0.00001	0.116727	0.00001	0.883253	Modern East Asian taurine
Mongolian	0.00001	0.042802	0.228517	0.728671	Modern East Asian taurine
Montbeliarde	0.00001	0.00001	0.00001	0.99997	Modern Central and European taurine
Muturu	0.00001	0.03908	0.00001	0.9609	Modern West African taurine
N'dama	0.00001	0.028358	0.197793	0.773839	Modern West African taurine
Normande	0.045429	0.00001	0.00001	0.954551	Modern Central and European taurine
Ogaden	0.00001	0.00001	0.848004	0.151976	Modern East African hybrid
Original Braunvieh	0.00001	0.00001	0.00001	0.99997	Modern Central and European taurine
Pad1	0.99997	0.00001	0.00001	0.00001	Italian Aurochs
Pajuna	0.061129	0.00001	0.00001	0.938851	Modern Iberian taurine
Pal1	0.978011	0.00001	0.00001	0.021969	Italian Aurochs
Piedmontese	0.00001	0.00001	0.00001	0.99997	Modern Central and European taurine
Plo1	0.096011	0.00001	0.00001	0.903969	Ancient European domesticate
Plo2	0.00001	0.00001	0.00001	0.99997	Ancient European domesticate
Plo3	0.00001	0.00001	0.00001	0.99997	Ancient European domesticate
Plo4	0.00001	0.00001	0.00001	0.99997	Ancient European domesticate
Plo5	0.00001	0.00001	0.00001	0.99997	Ancient European domesticate
Plo6	0.00001	0.00001	0.00001	0.99997	Ancient European domesticate
Plo7	0.010192	0.00001	0.00001	0.989788	Ancient European domesticate
Plo8	0.015088	0.00001	0.00001	0.984892	Ancient European domesticate
Podolica	0.00001	0.00001	0.094097	0.905883	Modern South European taurine
Qaz1	0.00001	0.024605	0.00001	0.975375	Early SW Asian domesticates
Rashoki	0.00001	0.00001	0.396894	0.603086	Modern SW Asian hybrid

Rhi1	0.861853	0.138127	0.00001	0.00001	Pleistocene Rhine Valley aurochs
Rhi2	0.880313	0.119667	0.00001	0.00001	Pleistocene Rhine Valley aurochs
Rhi3	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Pre-domestic contact
Rhi4	0.927996	0.071984	0.00001	0.00001	Pleistocene Rhine Valley aurochs
Rhi5	0.943147	0.056833	0.00001	0.00001	Pleistocene Rhine Valley aurochs
Rhi6	0.932335	0.067645	0.00001	0.00001	Pleistocene Rhine Valley aurochs
ROS002	0.00001	0.99997	0.00001	0.00001	Holocene Central Asian aurochs
Sahiwal	0.00001	0.00001	0.99997	0.00001	Modern South Asian indicine
Sayaguesa	0.062069	0.00001	0.00001	0.937911	Modern Iberian taurine
Shi1	0.00001	0.1609	0.00001	0.83908	Ancient Central Asian domesticates
Shimao	0.00001	0.132947	0.00001	0.867033	Ancient East Asian domesticates
Shorthorn	0.0387	0.00001	0.00001	0.96128	Modern Central and European taurine
Sikia	0.00001	0.00001	0.127145	0.872835	Modern South European taurine
Simmental	0.00001	0.00001	0.00001	0.99997	Modern Central and European taurine
Siv1	0.74561	0.00001	0.00001	0.25437	Balkan aurochs
Ska1	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Pre-domestic contact
Ska3	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Pre-domestic contact
Somba	0.00001	0.038849	0.039552	0.921589	Modern West African taurine
South Anatolian Red	0.00001	0.00001	0.341441	0.658539	Modern SW Asian hybrid
Stu1	0.048131	0.00001	0.00001	0.951848	Ancient European domesticate
Sub1	0.00001	0.00001	0.00001	0.99997	Early SW Asian domesticates
Swedish Red	0.038504	0.00001	0.00001	0.961476	Modern Central and European taurine
Tango1	0.330548	0.445731	0.00001	0.22371	Unknown
Tango2	0.69106	0.091418	0.00001	0.217512	Unknown
Tharparker	0.00001	0.00001	0.99997	0.00001	Modern South Asian indicine
Tibetan	0.00001	0.303754	0.00001	0.696226	Modern East Asian taurine
Tiroler Grauvieh	0.00001	0.00001	0.00001	0.99997	Modern Central and European taurine
Ton1	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Post-domestic contact
Tri1	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Pre-domestic contact
Tsa1	0.00001	0.00001	0.406253	0.593727	Post-Bronze Age SW Asian domesticates
Tula1	0.00001	0.99997	0.00001	0.00001	Pleistocene East Eurasian aurochs
Uzz1	0.99997	0.00001	0.00001	0.00001	Italian Aurochs
Var1	0.265262	0.49538	0.00001	0.239347	Holocene Central Asian aurochs

Var2	0.302646	0.434032	0.00001	0.263312	Holocene Central Asian aurochs
Wagyu	0.00001	0.069347	0.00001	0.930633	Modern East Asian taurine
Wannan	0.00001	0.00001	0.99997	0.00001	Modern East Asian indicine
Yakut	0.00001	0.092364	0.077053	0.830574	Modern East Asian taurine
Yanbian	0.00001	0.151449	0.00001	0.848531	Modern East Asian taurine
Yen1	0.00001	0.99997	0.00001	0.00001	Pleistocene East Eurasian aurochs
Yer1	0.00001	0.00001	0.737308	0.262672	Ancient Central Asian domesticates
YoA	0.99997	0.00001	0.00001	0.00001	C. and N. European aurochs - Post-domestic contact