Herding Ancient Domesticates:
From Bones to Genomes

Victoria Elizabeth Mullin

A thesis submitted to the University of Dublin for the degree of Doctor of Philosophy

Department of Genetics

University of Dublin Trinity College

2018

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.

Signed…………………….............................. Date………………………

Acknowledgements

First and foremost I wish to thank Professor Daniel Bradley for providing me with the opportunity to work as part of the Codex project, contributing to the fantastic world class dataset that we have developed, and for his guidance and teaching over the last 5 years. Additionally, I would like to thank him for his generous philosophy of supporting his PhD students in their endeavors outside of their PhD, enabling me to combine my two passions, science and sport.

I'd like to thank all past and present Bradley lab members for making the lab such a pleasant place to work. In particular thank you Russell, Ian, Rui and Cristina for their generous bioinformatic help, advice, copious drinks and stimulating conversations. To the mother-in-chief of the lab, Valeria: thank you for all your help with the enormous amounts of lab work, for your MiSeq expertise and for your support throughout these five years.To Matthew who taught me the command line didn't need to be downloaded. Thank you for the years of bioinformatic guidance, the on-hand computer support and the many fruitful discussions which are just too numerous to list. To Amelie for being an awesome cattle collaborator throughout the codex project, thank you for imparting your cattle knowledge. To Pier, for your generous help and points of view over the last year whether it be helping me with bloomin mitochondria or my CV. To Kevin for parting with his knowledge on BEAST in the final sprint of this epic marathon and for generally being a good sport. To Andrew for lending me some of his sheep samples in my thesis, you really have a fantastic study ahead of you. To Lara for her friendship and long discussions in both the lab and the pub over the years. To Eppie for befriending me when I first moved to a new lab and a new country, and whom I can still count as an absolutely fabulous friend. Lastly, but by no means least, to my partner in cattle and cake eating crime, Marta, who has been a fantastic project colleague, stress eating participant and more importantly a friend.

To the departmental staff who have always been there to answer a question, sort out an order or solve the problem of the ceiling falling in (both figuratively and literally). Many thanks to Brenda, Dave, Paul, Rachel and Sue.

Thank you to the European Research Council for funding the Codex project that has enabled us to sequence so many new and exciting ancient genomes. A word of thanks to the codex collaboators, especially Matthew Collins for his enthusiasm and sharing of both knowledge and contacts. Archeologists and museums have been integral to this PhD and the codex project as a whole. My personal thanks must go to Jacqui Mulville, Finbar McCormick, National Museum of Ireland, Terry O’Connor, Mike Parker Pearson, Ingrid Mainland, David Allen (Hampshire Cultural Trust), Lisa Brown (Wiltshire Museum), Danielle Schreve, Umberto Albarella, Ros O Maolduin, Carlton Jones, Fiona Beglane, Andy Maxted (Royal Pavilion) and Richard Madgwick.

To the staff in the trinity sports department who have been so supportive of me throughout the last five years, in particular Caroline, Deirdre, Mark, Cathy and Michelle. To all my friends in Dublin University Rifle Club. In particular, much appreciation has to be expressed to Conor for his python teaching when I first started...and for a while after that. A gigantic thank you to Siobhan for the conversations, donuts, running our euromillions syndicate and for saving my sanity with her invaluable help with formatting this thesis.

Keeping friendships alive when being all consumed by PhD and sport can be exceedingly difficult. I must however thank a small group of friends for still being there even when I am the worst person for contact; Fiona, Eilis, Kendra, Geri, Gorgs, Elizabeth, Miranda and Rishika.

Thank you to the staff at my alma mater Royal Holloway for ensuring I could achieve whilst being an athlete and for delivering a course which only furthered my interest in biological sciences. To Dr Tony Stead for his encouraging words and to Prof Ian Barnes for introducing me to the world of ancient DNA. To Dr Selina Brace, thank you for the numerous pep talks, advice and emails.

To my sporting family especially Steve Pengelly, Margaret Thomas and Phoebe Sanders for being supportive of me as a person and not just an athlete.

And lastly and most importantly to my family. To Tess, Kika and Oscar, the past family dogs that have truly filled our lives with joy and laughter and cemented my love of man's best friend. To John for his never ending support, understanding and most importanly love through these tough few years. You saw the light at the end of the tunnel when I could not and ensured I kept moving forward towards it. I can’t imagine what the last few years would have been lilke without you there, thank you for everything. To my little sister Katherine who has put up with being Victoria's little sister her whole life, your passion for the world and environment is truly inspiring and I'm so grateful that you are my sister. To auntie Sue for your witty emails at a time when humour was truly lacking.

To my gran Betty and my grandma Avril for your unconditional love and support for me as a person in everything I do. For your belief in me from day one, thank you. To my parents, Lynne and Jon, who are truly phenomenal people. Thank you for your sacrifices to ensure a top class education both inside and outside the classroom, for your fight to ensure dyslexia and dyspraxia were not going to hold me back, for the somewhat harrowing weekend tutorials, for the copious amounts of proofreading over the years, the support for my sport and most importantly for your love and support from both near and afar.

Lastly to two men no longer with us, my grandads who taught me many a life lesson. To my grandad John Mullin who's scientific knowledge left me in awe as a child and whose dedication to scientific excellence continues to inspire me. To my grandad Colin Raper whose passion for nature, farming, people and sport was truly infectious and will forever remain with me.

Many heartfelt thanks to you all

#thisgirlIcan

A Sport England programme to encourage women in sport, no matter their expertise.

“Working hard is important.
But there is something that matters even more, believing in yourself”
Harry Potter – J.K. Rowling

Summary

The domestication of animals and plants is a landmark event in human prehistory warranted of the attention of many past population genetic studies. Utilising modern breeds to understand domestication has its limitations, as breed formation and historic population movement can confound conclusions. The use of ancient DNA recovered from archaeological material can enrich our understanding of the domestication process and subsequent migrations prior to the development of modern breeding. With improvements in sequencing technologies and ancient DNA recovery, the contribution of ancient genomes to our understanding of the process of domestication has only just begun. This thesis demonstrates the power of Next Generation Sequencing (NGS) of ancient domestic genomes in order to analyse past populations of two domesticate species, cattle (*Bos taurus*) and sheep (*Ovis aries)*.

Chapter 3 discusses the successes and failures of recovering ancient DNA from 183 archaeological specimens using NGS. The prowess of the petrous bone for the preservation of DNA is examined, while molecular sexing of a time series of cattle and sheep from Britain and Ireland are briefly examined in respect to herd management strategies.

Chapter 4 assesses the *Bos* phylogeny and geographical distribution of ancient *Bos taurus* and *Bos primigenius* haplogroups through the analysis of 127 ancient mitogenomes and previously published modern mitogenomes. These ancient mitogenomes were achieved through a mixture of target capture and shotgun sequencing. Included in this chapter is a Bayesian phylogenetic analysis of 24 radiocarbon dated ancient mitogenomes, along with modern genomes, to provide greater resolution of the timings of mtDNA haplogroup divergence. In addition, this chapter presents a preliminary analysis of a 200,000 year old British *Bos primigenius* sample.

Chapter 5 presents the analysis of 113 ancient cattle whole genomes from Anatolia to Ireland, spanning from the Neolithic to the Medieval period. The Neolithic migration of cattle is demonstrated by 42 Neolithic genomes from Anatolia to the North Atlantic Edge, gaining an insight into the genetic structure present in Neolithic Europe. Admixture with the wild European auroch, *Bos primigenius*, is demonstrated to have likely occurred several times in the Neolithic. While a times series of 77 genomes from the North Atlantic Edge enables the study of genetic structure through time, indicating that the population history of cattle does not mirror that of the contemporary human population. Additionally, preliminary results of allele frequencies suggest selection for increase milk yields occuring post the appearance of lactose tolerance in European humans.

Chapter 6 presents the analysis of whole genome data from 28 ancient Eurasian sheep, ranging temporally from the Neolithic to Medieval period, with a predominant focus on the 19 genomes from Britain and Ireland. The study aimed to gain insight into the genetic structure present in ancient samples prior to the formation of the breed books and its variation through time. Demonstration of a geographic structure similar to that of modern breeds existed in the ancient Eurasian sheep population and tentative conclusions of subsequent sheep migrations post-Bronze Age, help to provide a foundation for future work.

Table of Contents

List of Tables

List of Figures

Abbreviations

A - adenine

aDNA - ancient DNA

AD - Anno Domini

BP - Before Present

BC - Before Christ

C - cytosine

DNA - Deoxyribonucleic acid

G - guanine

mtDNA - mitochondrial DNA

nuDNA - nuclear DNA

NGS - Next Generation Sequencing

PCR - Polymerase Chain Reaction

Per. comm. - personal communication

PPNA - Pre-Pottery Neolithic A

PPNC - Pre-Pottery Neolithic C

T – thymine

Chapter 1. General Introduction

1.1. Ancient DNA

It is widely viewed that the field of ancient DNA (aDNA) has its origins in two publications; the sequencing of mitochondrial DNA (mtDNA) from the extinct Quagga (Higuchi et al. 1984) and from a mummified human brain (Pääbo 1985). Since the publications of the two original papers the field has proven itself as a credible source of valuable insight into the past (Shapiro et al. 2004; Svensson et al. 2007; Skoglund et al. 2012; Palkopoulou et al. 2013; Gamba et al. 2014; Olalde et al. 2014; Haak et al. 2015; Frantz et al. 2016).

However, working with aDNA brings a number of challenges that need to be overcome in order to secure valid data. The stereotypical chemical modifications present in aDNA are the deamination of unmethylated cytosine and depurination of the DNA backbone which create short damaged fragments, easily contaminated with contemporary DNA (Hansen et al. 2001; Hofreiter et al. 2001). The speed of degradation of the nucleic acids and thus its preservation is highly dependent on the post-mortem *in situ* environment of the sample, for example the freezing cold temperatures of permafrost are known to better preserve DNA than a humid tropical environment (Lindahl 1993; Smith et al. 2001; Allentoft et al. 2012; Campos et al. 2012). Deamination of unmethylated cytosine poses a particular problem as this process leads to the production of uracil, which during replication and hence sequencing, is replaced with thymine. To best ensure the validity of results, protocols for the practice of aDNA research have been developed (Cooper & Poinar 2000; Hofreiter et al. 2001). These include separate working laboratories for pre- and post- PCR and the use of protective clothing.

Prior to 2006 mtDNA was used to gain an understanding of the genetic past, at both a species and a population level (Troy et al. 2001; Shapiro et al. 2004). This was achieved utilising the combination of PCR amplification and Sanger sequencing. However, the lack of autosomal sequences restricted the scope of information.

A resurgence of aDNA research has been achieved due to the development of Next Generation Sequencing (NGS). With decreasing costs, NGS has enabled aDNA research to begin to develop past the restrictions of PCR. Within 6 months of introduction, NGS increased the amount of sequencing data obtained by three orders of magnitude; the previous increase of two orders of magnitude had taken 20 years to obtain (Shapiro & Hofreiter 2010). The first reported NGS sequences of aDNA were that from a frozen permafrost mammoth (Poinar et al. 2006). Subsequently there have been many publications of hominid genomes (Rasmussen et al. 2010; Meyer et al. 2012; Olalde et al. 2014; Gamba et al. 2014; Allentoft et al. 2015; Fu et al. 2015; Llorente et al. 2015; Cassidy et al. 2016; Jones et al. 2017) offering us insight into prehistoric population movements that were unattainable previously. Contamination was shown to remain an issue whilst working with NGS (Green et al. 2006). However the methodology used in Green and colleagues (2006) did not adhere to aDNA protocol guidelines (Cooper & Poinar 2000; Hofreiter et al. 2001).

Ancient DNA methodology papers have focused on improving NGS library protocols, as well as extraction protocols (Meyer & Kircher 2010; Meyer et al. 2012; Dabney et al. 2013; Seguin-Orlando et al. 2013; Damgaard et al. 2015; Gamba et al. 2015). These improvements have enabled DNA to be extracted and sequenced from samples as old as 560,000-780,000 yrBP (Orlando et al. 2013), substantially older than the previously proposed ceiling for aDNA recovery of 100,000 years age (Hofreiter et al. 2001). Not only has this limit been pushed backwards in time, but also the percentage of endogenous aDNA recovered from a sample has been increased. The Denisovan phalanx, from which a full genome of an archaic hominim was recovered, was reported to contain 70% endogenous DNA (Meyer et al. 2012). Moreover, Gamba and colleagues (2014) demonstrated that targeting robust bone elements, such as the petrosal portion of the temporal bone, increases the chance of recovering high endogenous percentages of DNA.

The challenges of NGS are not confined to the wet-lab; with a vast increase in data volume, the computational challenge facing NGS researchers is considerable. Consequently, numerous bioinformatic programs and tools have been developed to process and analyse NGS sequences (Li & Durbin 2009; Li et al. 2009; McKenna et al. 2010). Whilst the majority of NGS programs have not been designed for aDNA, specific tools have been developed to assess levels of authenticity via deamination patterns (Jónsson et al. 2013; Skoglund et al. 2014). Additionally, packages specifically for read processing and data analysis, such as PALEOMIX (Schubert et al. 2014), and aDNA aware software such as ANGSD (Korneliussen et al. 2014), have been developed.

1.2. Domestication

The process of domestication has been a focus for both aDNA and modern DNA publications (Troy et al. 2001; Larson et al. 2005; Edwards et al. 2007; Scheu et al. 2015; Frantz et al. 2016; Loog et al. 2017). However, few publications have applied NGS technology to analyse nuclear aDNA in order to further understand the process of domestication.

While dogs were the first domesticate prior to 15,000 yrBP (Frantz et al. 2016), domestication of plants and animals, along with the appearance of sedentism and pottery, marks the transition from the Mesolithic to the Neolithic period, a gradual shift away from the hunter-gatherer lifestyle to one of a growing reliance on managed food resources known as Neolithisation. This change in lifestyle is primarily proposed to have occurred due to the need to sustain an increasing human population, coinciding with the climate warming after the end of the cold spell of the Younger Dryas, approximately 11,100 years BP (Bruford et al. 2003). However, it is recognised that there are many reasons possible for this change in lifestyle, likely including anthropological and ecological reasoning (Vigne 2011; Vigne 2015).

The gradual and long-term nature of domestication presents many challenges for study. For example, defining a start date or concluding when it has stopped are both somewhat difficult**.** Within an archaeological context, sexual dimorphism can cause confusion when designating archaeological remains as wild or domesticated (Zeder & Hesse 2000). Domestication is not a process confined to history, with the recent domestication of fish occurring in the 20th century (as reviewed by Vigne 2015).

The first reliable date from the archaeological record of domesticated plant material, domestic emmer wheat (*Triticum dicoccum*) and einkorn wheat (*Triticum monococcum*), is from the early Pre-Pottery Neolithic between 11,500 and 11,000 years BP in the Upper Euphrates valley, modern-day Turkey (Zeder 2011).

Whereas plants such as wheat (*Triticum* sp) and barley (*Hordeum vulgare*) were domesticated in the well-known Fertile Crescent, plants such as rice (*Oryza* sp) and quinoa (*Chenopodium quinoa*) were domesticated in the two other major areas of domestication, China and the Andean region of South America. Many animals have been successfully domesticated, cattle (*Bos taurus* and *Bos indicus*) from the auroch (*Bos primigenius*), sheep (*Ovis aries*) from mouflon and dog (*Canis lupus familiaris*) from the wolf (*Canis lupus*). There is also evidence for failed attempts most notably of fallow deer (Vigne 2011).

As domestication of animals was an important component of Neolithisation, the study of ancient domesticate remains can act as a proxy to chart the spread of the Neolithic (Tresset et al. 2009). In addition, the study of ancient domesticate genomes may illuminate other parts of human history, such as subsequent human migrations and possible population collapses. Utilising modern breeds to understand domestication has limitations, as historic population movement and the formation of breeds can confound conclusions based solely on modern data (Kijas et al. 2012). By predating breed formation, the study of ancient genomes from a variety of ages can enrich our understanding (Loog et al. 2017).

1.2.2. Cattle

Domestic cattle (*Bos taurus* and *Bos indicus*) and their extinct wild progenitor, the auroch (*Bos primigenius*) are placed within the Bovini tribe (Sub family Bovinae, family Bovidae), with the other *Bos*, *Bison* and *Bubalus* species. Commonly inferred to as sub-species, *B. taurus* and *B. indicus* are inter-fertile, however they are morphologically dissimilar. The presence of a cerviothoracic hump in *B. indicus* and its absence in *B. taurus* is one marker of the sub-species different evolutionary histories. Phylogenies of modern cattle indicate a deep bifurcation of these two sub-species occurring hundreds of thousands of years BP and therefore separate domestications from distinct populations (Loftus et al. 1999; Bradley et al. 1996; MacHugh et al. 1997; Troy et al. 2001; Chen et al. 2010).

Indicine cattle are less studied than taurine cattle. Indicine cattle were domesticated in the Indus valley, India, (Lei et al. 2006; Chen et al. 2010) some 8,000 yrBP. Taurine cattle, to which modern day European cattle belong, were domesticated in the Near East (Southwest Asia) (Troy et al. 2001; Edwards et al. 2007) around 11,000 to 10,000 years BP in the Upper and Middle Euphrates Valley (Helmer et al. 2005) from a small effective number of female aurochs with migration into Europe occurring around 8,400 BP, eventually reaching northern Europe after 6,100 BP (Bollongino et al. 2012; Scheu et al. 2015).

Until relatively recently the understanding of cattle population histories has been based upon mtDNA, Y chromosome studies, and microsatellites data (Loftus et al. 1994; Bradley et al. 1996; MacHugh et al. 1997; Mannen et al. 1998; Troy et al. 2001; Bollongino et al. 2006; Edwards et al. 2007; Achilli et al. 2009; Speller et al. 2013; Scheu et al. 2015; Niemi et al. 2015). With advancement in sequencing technology, the publication of the bovine nuclear genome (Bovine Genome Sequencing and Analysis Consortium et al. 2009), and genome wide SNP survey (Bovine HapMap Consortium et al. 2009) have enabled studies to focus on autosomal variation, identifying patterns of admixture between taurine and indicine cattle and phylogenetic structure (Decker et al. 2014; Park et al. 2015; Upadhyay et al. 2016). The publication and ongoing sequencing of 1000 Bull whole genomes (Daetwyler et al. 2014; Chung et al. 2017) and the first ancient bovid genome from a

British Mesolithic *B. primigenius* (Park et al. 2015), present further opportunity to understand population histories of cattle at the whole genome level.

1.2.3. Sheep

Domestic sheep (*Ovis aries*) belong the subfamily Caprinae and in turn the family Bovidae. Archaeozoological evidence indicates the domestication of sheep occurred in the Fertile Crescent region of Southwest Asia, approximately 11,000 yrBP (Zeder 2008). Domesticated primarily for meat, wool and milk production are believed to have been selected for as secondary products (Chessa et al. 2009). Through human mediated movement sheep migrated into Europe, Asia and Africa, possibly through two large migration events, where more primitive initial populations were displaced by improved secondary product populations (Chessa et al. 2009).

Mitochondrial analyses have detailed the global dispersal of two major global haplogroups with the restriction of another three minor haplogroups discovered in the Southwest Asia. This mtDNA diversity in SW. Asia, replicated in microsatellite data, ties in with this region as a centre of origin and hence genomic diversity as seen in other domesticates (Troy et al. 2001; Meadows et al. 2005; Pereira et al. 2006; Tapio et al. 2006; Lawson Handley et al. 2007; Peter et al. 2007; Meadows et al. 2011; Lv et al. 2015). The first global genome-wide study of modern sheep breeds analysed 50,000 SNPs, demonstrating a phylogeographic structure in modern breeds. Echoing the phylogeographic clustering of the major mtDNA haplogroups, clear genetic divisions between European, Asian and African breeds were identified (Kijas et al. 2012). The study of ancient genomes provides the opportunity to understand the genetic variation in sheep and uncover any phylogeographic structure prior to the formation of breeds and the historic population movement (Kijas et al. 2012).

1.3. The Atlantic Edge: Britain and Ireland

On the Atlantic Edge of Europe lie a number of islands, the two largest are Britain and Ireland separated from each other by the Irish Sea. Evidence for modern anatomical humans in Britain stretches back into the Pleistocene epoch to 40,000 years ago (Higham et al. 2011). The Pleistocene began 1.8 million years ago and lasted until the beginning of the current epoch the Holocene 11,500 years ago (Cox & Moore 2005), during which Northern Europe was subjected to a number of glacial and inter-glacial events, causing both the expansion and contraction of many species ranges (Hewitt 2000; Lister & Stuart 2008). With the stable climate of the Holocene the recolonization of Britain by an early Mesolithic hunter-gatherer society occurred 12,500 yrBP (Tolan-Smith 1998; Barton &

Roberts 2004), however the Mesolithic did not occur in Ireland until approximately 9,000 years before present (Woodman 1986).

The initial Neolithisation of the British Isles remains a contentious topic in regards to the time, the route and the speed of expansion across the isles (Thomas 1991; Sheridan 2010; Whittle et al. 2011). Moreover, due to their geographic situation the isles are a western endpoint to the spread of the Neolithic and therefore an interesting population genetic study for both humans and domesticated animals.

Ferriter's cove, Co Kerry Ireland, is home to the oldest radiocarbon dates of cattle bones in Britain and Ireland, 4450-4270 calBC (Wooman et al. 1999). However it is unclear whether these bones signified a failed colonization (Sheridan 2010) or were just simply joints of meat traded from Britain or the continent (Whittle 2007). Various studies have attempted to date the initial colonisations; whilst each have their failings, approximate dates for the introduction of the Neolithic to Britain and Ireland are 4,000 yrBC and 3,800 yrBC respectively (Collard et al. 2010; Whittle et al. 2011; McClatchie et al. 2012; Whitehouse et al. 2014). During the Neolithic domesticated animals first entered Britain from the continent, perhaps progressing to Ireland or potentially with animals originating from a separate European source. Domestic populations may have persisted but there is a suggestion that Neolithic crop farming may have crashed in the Late Neolithic (Stevens & Fuller 2012; Whitehouse et al. 2014), while genetic evidence suggests mass human population movement and genetic replacement across Europe in the subsequent Bronze age (Allentoft et al. 2015; Haak et al. 2015; Cassidy et al. 2016; Olalde et al. 2017). One key question that ancient domesticate genomes can address, is whether there was genetic continuity of cattle and sheep from the Neolithic through tumultuous periods of prehistory until the introduction of breed books and selective breeding in the 18th Century.

1.4. References

Achilli, A. et al., 2009. The multifaceted origin of taurine cattle reflected by the mitochondrial genome. *PloS one*, 4(6), p.e5753.

Allentoft, M.E. et al., 2015. Population genomics of Bronze Age Eurasia. *Nature*, 522(7555), pp.167–172.

Allentoft, M.E. 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), pp.4724–4733.

Barton, N. & Roberts, A., 2004. The Mesolithic period in England: current perspectives and new research. *Mesolithic Scotland and its neighbours*, pp.339–358.

Bollongino, R. et al., 2006. Early history of European domestic cattle as revealed by ancient DNA. *Biology letters*, 2(1), pp.155–159.

Bollongino, R. et al., 2012. Modern taurine cattle descended from small number of near-eastern founders. *Molecular biology and evolution*, 29(9), pp.2101–2104.

Bovine Genome Sequencing and Analysis Consortium et al., 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, 324(5926), pp.522–528.

Bovine HapMap Consortium et al., 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science*, 324(5926), pp.528–532.

Bradley, D.G. et al., 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), pp.5131–5135.

Bruford, M.W., Bradley, D.G. & Luikart, G., 2003. DNA markers reveal the complexity of livestock domestication. *Nature reviews. Genetics*, 4(11), pp.900–910.

Campos, P.F. et al., 2012. DNA in ancient bone - where is it located and how should we extract it? *Annals of anatomy = Anatomischer Anzeiger: official organ of the Anatomische Gesellschaft*, 194(1), pp.7–16.

Cassidy, L.M. et al., 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), pp.368–373.

Chen, S. et al., 2010. Zebu cattle are an exclusive legacy of the South Asia neolithic. *Molecular biology and evolution*, 27(1), pp.1–6.

Chessa, B. et al., 2009. Revealing the history of sheep domestication using retrovirus integrations. *Science*, 324(5926), pp.532–536.

Chung, N.C. et al., 2017. Population Structure Analysis of Bull Genomes of European and Western Ancestry. *Scientific reports*, 7, p.40688.

Collard, M. et al., 2010. Radiocarbon evidence indicates that migrants introduced farming to Britain. *Journal of archaeological science*, 37(4), pp.866–870.

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

Cox, B.C. & Moore, P.D., 2005. *Biogeography--An Ecological and Evolutionary Approach*, Blackwell, Oxford, UK.

Dabney, J. et al., 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America*, 110(39), pp.15758–15763.

Daetwyler, H.D. et al., 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature genetics*, 46(8), pp.858–865.

Damgaard, P.B. et al., 2015. Improving access to endogenous DNA in ancient bones and teeth. *Scientific reports*, 5, p.11184.

Decker, J.E. et al., 2014. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS genetics*, 10(3), p.e1004254.

Edwards, C.J. et al., 2007. Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1616), pp.1377–1385.

Frantz, L.A.F. et al., 2016. Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science*, 352(6290), pp.1228–1231.

Fu, Q. et al., 2015. An early modern human from Romania with a recent Neanderthal ancestor. *Nature*. Available at: http://dx.doi.org/10.1038/nature14558.

Gamba, C. et al., 2015. Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Molecular ecology resources*. Available at: http://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12470/full.

Gamba, C. et al., 2014. Genome flux and stasis in a five millennium transect of European prehistory. Nature communications, 5, p.5257.

Green, R.E. et al., 2006. Analysis of one million base pairs of Neanderthal DNA. *Nature*, 444(7117), pp.330–336.

Haak, W. et al., 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. Available at: http://dx.doi.org/10.1038/nature14317.

Hansen, A. et al., 2001. Statistical evidence for miscoding lesions in ancient DNA templates. *Molecular biology and evolution*, 18(2), pp.262–265.

Helmer, D. et al., 2005. Identifying early domestic cattle from Pre-Pottery Neolithic sites on the Middle Euphrates using sexual dimorphism. In J. D. Vigne, J. Peter, & D. Helmer, eds. *The First Steps of Animal Domestication*. Oxbow Books, Oxford, pp. 86–95.

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789), pp.907–913.

Higham, T. et al., 2011. The earliest evidence for anatomically modern humans in northwestern Europe. *Nature*, 479(7374), pp.521–524.

Higuchi, R. et al., 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature*, 312(5991), pp.282–284.

Hofreiter, M., Serre, D., et al., 2001. Ancient DNA. *Nature reviews. Genetics*, 2(5), pp.353–359.

Hofreiter, M., Jaenicke, V., et al., 2001. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic acids research*, 29(23), pp.4793–4799.

Jones, E.R. et al., 2017. The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European Farmers. *Current biology: CB*, 27(4), pp.576–582.

Jónsson, H. et al., 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* , 29(13), pp.1682–1684.

Kijas, J.W. et al., 2012. Genome-wide analysis of the world’s sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS biology*, 10(2), p.e1001258.

Korneliussen, T.S., Albrechtsen, A. & Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC bioinformatics*, 15, p.356.

Larson, G. et al., 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science*, 307(5715), pp.1618–1621.

Lawson Handley, L.-J. et al., 2007. Genetic structure of European sheep breeds. *Heredity*, 99(6), pp.620–631.

Lei, C.Z. et al., 2006. Origin and phylogeographical structure of Chinese cattle. *Animal genetics*, 37(6), pp.579–582.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , 25(16), pp.2078–2079.

Li, H. & Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* , 25(14), pp.1754–1760.

Lindahl, T., 1993. Instability and decay of the primary structure of DNA. *Nature*, 362(6422), pp.709–715.

Lister, A.M. & Stuart, A.J., 2008. The impact of climate change on large mammal distribution and extinction: Evidence from the last glacial/interglacial transition. *Comptes rendus: Geoscience*, 340(9), pp.615–620.

Llorente, M.G. et al., 2015. Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science*, 350(6262), pp.820–822.

Loftus, R.T. et al., 1999. A microsatellite survey of cattle from a centre of origin: the Near East. *Molecular ecology*, 8(12), pp.2015–2022.

Loftus, R.T. et al., 1994. Mitochondrial genetic variation in European, African and Indian cattle populations. *Animal genetics*, 25(4), pp.265–271.

Loog, L. et al., 2017. Inferring allele frequency trajectories from ancient DNA indicates that selection on a chicken gene coincided with changes in medieval husbandry practices. *Molecular biology and evolution*. Available at: http://dx.doi.org/10.1093/molbev/msx142.

Lv, F.-H. et al., 2015. Mitogenomic Meta-Analysis Identifies Two Phases of Migration in the History of Eastern Eurasian Sheep. *Molecular biology and evolution*, 32(10), pp.2515–2533.

MacHugh, D.E. et al., 1997. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (Bos taurus and Bos indicus). *Genetics*, 146(3), pp.1071–1086.

Mannen, H. et al., 1998. Mitochondrial DNA variation and evolution of Japanese black cattle (Bos taurus). *Genetics*, 150(3), pp.1169–1175.

McClatchie, M. et al., 2012. Neolithic farming in north-western Europe: archaeobotanical evidence fromIreland. *Journal of archaeological science*, 51(0), pp.206–215.

McKenna, A. et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), pp.1297–1303.

Meadows, J.R.S. et al., 2005. Mitochondrial sequence reveals high levels of gene flow between breeds of domestic sheep from Asia and Europe. *The Journal of heredity*, 96(5), pp.494–501.

Meadows, J.R.S., Hiendleder, S. & Kijas, J.W., 2011. Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel. *Heredity*, 106(4), pp.700–706.

Meyer, M. et al., 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science*, 338(6104), pp.222–226.

Meyer, M. & Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor protocols*, 2010(6), p.db.prot5448.

Niemi, M. et al., 2015. Temporal fluctuation in North East Baltic Sea region cattle population revealed by mitochondrial and Y-chromosomal DNA analyses. *PloS one*, 10(5), p.e0123821.

Olalde, I. et al., 2014. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature*, 507(7491), pp.225–228.

Olalde, I. et al., 2017. The Beaker phenomenon and the genomic transformation of northwest Europe. *bioRxiv*, p.28. Available at: http://repository.nms.ac.uk/1864/ [Accessed October 17, 2017].

Orlando, L. et al., 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, 499(7456), pp.74–78.

Pääbo, S., 1985. Molecular cloning of Ancient Egyptian mummy DNA. *Nature*, 314(6012), pp.644–645.

Palkopoulou, E. et al., 2013. Holarctic genetic structure and range dynamics in the woolly mammoth. *Proceedings. Biological sciences / The Royal Society*, 280(1770), p.20131910.

Park, S. et al., 2015. Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and evolution of cattle. *Genome Biology*, 16(1), p.234. Available at: http://genomebiology.com/2015/16/1/234.

Pereira, F. et al., 2006. Genetic signatures of a Mediterranean influence in Iberian Peninsula sheep husbandry. *Molecular biology and evolution*, 23(7), pp.1420–1426.

Peter, C. et al., 2007. Genetic diversity and subdivision of 57 European and Middle-Eastern sheep breeds. *Genetics*. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2052.2007.01561.x/full.

Poinar, H.N. et al., 2006. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science*, 311(5759), pp.392–394.

Rasmussen, M. et al., 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*, 463(7282), pp.757–762.

Scheu, A. et al., 2015. The genetic prehistory of domesticated cattle from their origin to the spread across Europe. *BMC genetics*, 16, p.54.

Schubert, M. et al., 2014. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature protocols*, 9(5), pp.1056–1082.

Seguin-Orlando, A. et al., 2013. Ligation bias in illumina next-generation DNA libraries: implications for sequencing ancient genomes. *PloS one*, 8(10), p.e78575.

Shapiro, B. et al., 2004. Rise and fall of the Beringian steppe bison. *Science*, 306(5701), pp.1561–1565.

Shapiro, B. & Hofreiter, M., 2010. Analysis of ancient human genomes. *BioEssays: news and reviews in molecular, cellular and developmental biology*, 32(5), pp.388–391.

Sheridan, J.A., 2010. The Neolithisation of Britain and Ireland: the big picture. In B. Finlayson & G. Warren, eds. *Landscapes in transition*. Levant Supplementary Series. Oxford: Oxbow, pp. 89–105.

Skoglund, P. et al., 2012. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science*, 336(6080), pp.466–469.

Skoglund, P. et al., 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), pp.2229–2234.

Smith, C.I. et al., 2001. Neanderthal DNA. Not just old but old and cold? *Nature*, 410(6830), pp.771–772.

Speller, C.F. et al., 2013. Ancient mtDNA Analysis of Early 16th Century Caribbean Cattle Provides Insight into Founding Populations of New World Creole Cattle Breeds M. T. P. Gilbert, ed. *PloS one*, 8(7), p.e69584.

Stevens, C.J. & Fuller, D.Q., 2012. Did Neolithic farming fail? The case for a Bronze Age agricultural revolution in the British Isles. *Antiquity*, 86(333), pp.707–722.

Svensson, E.M. et al., 2007. Tracing genetic change over time using nuclear SNPs in ancient and modern cattle. *Animal genetics*, 38(4), pp.378–383.

Tapio, M. et al., 2006. Sheep mitochondrial DNA variation in European, Caucasian, and Central Asian areas. *Molecular biology and evolution*, 23(9), pp.1776–1783.

Thomas, J., 1991. *Rethinking the neolithic*, CUP Archive.

Tolan-Smith, C., 1998. RADIOCARBON CHRONOLOGY AND THE LATEGLACIAL AND EARLY POSTGLACIAL RESETTLEMENT OF THE BRITISH ISLES. *Quaternary International*, 49-50, pp.21–27.

Tresset, A. et al., 2009. Early diffusion of domestic bovids in Europe. *Becoming eloquent: advances in the emergence of language, human cognition, and modern cultures*. Available at: https://books.google.com/books?hl=en&lr=&id=X_cac38pfmAC&oi=fnd&pg=PT75&dq=Tresset+2009+neolithic&ots=K3AMeAG_GL&sig=b4sgwm0DXFYVJIJR6UHqJnswMYQ.

Troy, C.S. et al., 2001. Genetic evidence for Near-Eastern origins of European cattle. *Nature*, 410(6832), pp.1088–1091.

Troy, C.S. et al., 2001. Genetic evidence for Near-Eastern origins of European cattle. *Nature*, 410(6832), pp.1088–1091.

Upadhyay, M.R. et al., 2016. Genetic origin, admixture and population history of aurochs (Bos primigenius) and primitive European cattle. *Heredity*. Available at: http://dx.doi.org/10.1038/hdy.2016.79.

Vigne, J.-D., 2015. Early domestication and farming: what should we know or do for a better understanding? *Anthropozoologica*, 50(2), pp.123–150.

Vigne, J.-D., 2011. The origins of animal domestication and husbandry: a major change in the history of humanity and the biosphere. *Comptes rendus biologies*, 334(3), pp.171–181.

Whitehouse, N.J. et al., 2014. Neolithic agriculture on the European western frontier: the boom and bust of early farming in Ireland. *Journal of archaeological science*, (0). Available at: http://dx.doi.org/10.1016/j.jas.2013.08.009.

Whittle, A., 2007. Going over: people and their times. In *PROCEEDINGS-BRITISH ACADEMY*. OXFORD UNIVERSITY PRESS INC., p. 617.

Whittle, A.W.R., Healy, F.M.A. & Bayliss, A., 2011. *Gathering Time: Dating the Early Neolithic Enclosures of Southern Britain and Ireland*, Oxford: Oxbow Books.

Woodman, P.C., 1986. Problems in the Colonisation of Ireland. *Ulster Journal of Archaeology*, 49, pp.7–17.

Wooman, P.C., Anderson, E. & Finlay, N., 1999. *Excavations at Ferriter's Cove, 1983-95: Last Foragers, First Farmers in the Dingle Peninsula*, Dublin, Ireland: Wordwell.

Zeder, M.A., 2008. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the National Academy of Sciences of the United States of America*, 105(33), pp.11597–11604.

Zeder, M.A., 2011. The Origins of Agriculture in the Near East. *Current anthropology*, 52(S4), pp.S221–S235.

Zeder, M.A. & Hesse, B., 2000. The initial domestication of goats (Capra hircus) in the Zagros mountains 10,000 years ago. *Science*, 287(5461), pp.2254–2257.

Chapter 2. General Materials and Methods

2.1. Materials

2.1.1. Samples

Predominantly cattle and sheep, were screened for DNA preservation; ranging geographically from the Near East to the Atlantic Edge and temporally from 200,000 BP to approximately 800 BP (Appendix Table. 1). The majority of samples discussed in this thesis have been sourced, and permission granted, from Britain and Ireland. The analyses presented are supplemented with other unpublished ancient samples processed by Marta Verdugo, Amelie Scheu, Andrew Hare and Kevin Daly.

2.2. Method

All sample preparation, extraction and library preparation were performed in a dedicated Ancient DNA laboratory in the Smurfit Institute of Genetics, Trinity College Dublin. Standard aDNA protocols were used throughout the process (Cooper & Poinar 2000; Pääbo et al. 2004) including the use of extraction, library and PCR controls.

2.2.1. Sample Screening

2.2.1.1. Sample Preparation

Samples were decontaminated via UV exposure, the removal of surface layers using a dental drill and further UV exposure. Using a dremel diamond wheel, a piece of bone from the densest area of the sample was then cut out and sub-sampled with the use of a mixer mill (MM 400, Retsch). The mill was used in intervals of between 30 seconds to a minute, dependent on bone quality, removing the powder after every interval until the bone had all become bone powder. All tools used in this process were cleaned with bleach, DNA-ExitusPlus™ and ethanol before UV exposure for a total of 60 minutes.

2.2.1.2. Extraction

Extractions of between 0.09g and 0.3g of bone powder were performed based on (Gamba et al. 2014) with the majority using a two stage extraction first described by (Orlando et al. 2011). Briefly, the bone powder was incubated at 37°C using a thermo mixer at 700rpm with 1ml of lysis buffer (Tris HCl pH 7.4 - 20 mM; Proteinase K, recombinant, PCR Grade – 0.65 U/ml; Sarkosyl® NL 30 – 0.7 %; EDTA pH 8 – 47.5 mM) for 24 hours. Samples were then centrifuged for 10 minutes at 10,000rpm, supernatant removed, fresh lysis buffer added and the incubation and centrifugation steps were repeated. As contamination is more likely to be accessed during the first lysis stage (Vilstrup et al. 2013) supernatant from either the second or third lysis stage was taken to the next stage. Modifications to the above method include; only one extraction step with incubation of 55°C for 24 hours followed by 24 hours at 37°C 2 and an additional third extraction step temperatures originally mentioned (Appendix Table 1).

Extracts were subjected to a two step purification; first, the centrifugation of the supernatant in an Amicon® Ultra4 Centrifugal Filter Unit 30K with 3ml of 10 mM Tris-EDTA Buffer for approximately 20 minutes at 2,700rpm or until the volume obtained within the filter was 250µl. The flow through was then discarded, fresh buffer added and the centrifugation repeated until the volume obtained in the filter was 100µl. The final purification of the 100µl was then achieved using a silica column MinElute PCR Purification Kit , (QIAGEN), following manufacturer's instructions until the final step when TWEEN® 20 (0.05% final concentration, SigmaAldrich) was added to the Elution Buffer. The purified extracts were put on ice until the next step.

2.2.1.3. Library Preparation and Endogenous Screening

Double stranded libraries were constructed from 16.25ul - 30ul of DNA extract based on the protocol of (Meyer & Kircher 2010) with modifications as reported in (Gamba et al. 2014). Briefly, blunt end repair was performed using EBNext® End Repair Module (New England BioLabs Inc.) and during the final step of adapter fill-in *Bst* activity was arrested through heat inactivation (20 min at 80°C) .

Indexing PCRs were performed using Accuprime™ Pfx Supermix (Life Technology), primer IS4 (0.2 µM) and indexing primers (0.2 µM) (Meyer and Kircher 2010). 3µl of library was added to a total volume of 25µl of PCR mix and amplified using 5 min at 95°C, 10-12 cycles of 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 (Gamba et al 2014). PCR products were purified using MinElute MinElute PCR Purification Kit, QIAGEN.

Assessment of amplified libraries was achieved via either the Agilent 2100 Bioanalyzer or the Agilent 2200 TapeStation, following manufacturer's instructions. The majority of libraries were first screened on an Illumina MiSeq™ platform at TrinSeq (Trinity Genome Sequencing Laboratory, Trinity College Dublin, Ireland), using 50 bp single-end sequencing and a PhiX control at 1%. A small percentage of libraries were screened on a Illumina HiSeq™ by Macrogen Inc, Seoul, Republic of Korea.

2.2.1.4. Read Processing

Raw reads were processed through a pipeline prior to analysis. Sequencing read quality was first assessed with FastQC (Andrews & Others 2010). Cutadapt (Martin 2011) was used to remove adapter sequences from reads, with a final minimum read length of 30 base employed, allowing for a one base overlap between the adapter and the read. Sequence reads were aligned using the Burrows-Wheeler Algorithm (BWA) (version 0.7.5ar405) (Li & Durbin 2009) to the *Bos taurus* genome, build bosTau6 (downloaded from https://genome-euro.ucsc.edu/), or the *Ovis aries* genome, build Oar_v3.1 (downloaded from https://genome-euro.ucsc.edu/), run using default parameters apart from disabling the seed option ("l 1024") to improve alignment accuracy (Schubert et al. 2012).

Where species ID proved difficult to obtain morphologically, the use of FastQ Screen (https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/) prior to BWA enabled quick alignment to multiple genomes to obtain a species ID. Post BWA alignment, SAMtools (version 0.1.1996b5f2294a) (Li et al. 2009) was utilised to remove duplicates and exclude reads with a mapping quality lower than 30. Endogenous percentages were then calculated by the fraction of reads that aligned to the total number of sequencing reads for that barcode.

2.2.1.5. Deamination Patterns

Ancient DNA carries signatures of post-mortem deamination, characterised by an increase of cytosine to thymine misincorporation. These substitutions are clustered toward the termini of the read due to the acceleration of deamination of cytosine at single-stranded overhangs, creating the pattern of cytosine to thymine at the 5' termini of the reads, and therefore a reciprocal pattern of guanine to adenine misincorporation at 3' end (Brotherton et al. 2007; Ginolhac et al. 2011). This chemical signature of aDNA can be utilised to assess the degree of deamination present and therefore judge the authenticity of the aDNA. The programme MapDamage has been developed to assess the degree of misincorporation at the 5' and 3' read ends (Jónsson et al. 2013). MapDamage(2.0) was therefore performed to access the fragmentation patterns present in the samples.

2.2.1.6. Sexing

Sexing was performed using an in house script (M.Teasdale *per. comm.*), that examines the number of reads aligned per chromosomes relative to length of the chromosome.

2.2.2. Whole Genome Sequencing

2.2.2.1. Materials

Extract was used as from section 2.2.1.2.

2.2.2.2. USER treatment

To alleviate downstream analysis problems caused by the deamination of aDNA, samples were treated with a combination of uracil-DNA-glycosylase (UDG) and endonuclease VIII known as USER enzyme (Briggs et al. 2010). UDG removes uracil residues from the aDNA leaving abasic sites, which are then cleaved by endonuclease VIII enabling the undamaged DNA fragments intact and sequenceable (Briggs et al. 2010).

5µl of USER enzyme (1µ/µl) was added to 16.25µl of DNA extract and incubated at 37°C for 3 hours. Once finished the first stage of library preparation followed immediately.

2.2.2.2. Library formation and sequencing

Library formation was performed as in section 2.2.1.3. Calculations were performed to estimate the number of sequencing lanes required to achieve the desired genome coverage for each sample, average coverages ranged from 0.2 to 15X. PCRs were pooled by sample and sent for 100bp single-strand sequencing on an Illumina HiSeq™ by Macrogen Inc, Seoul, Republic of Korea.

2.2.2.3. Read Processing

Raw reads were processed through a pipeline similar to that reported in section 2.1.2.4 with modifications as follows: to allay the possibility of post-PCR contamination, raw reads were filtered for exact matches to the indices used in library preparation. Indices were processed separately through Cutadapt and BWA (section 2.1.24) (Li & Durbin 2009; Martin 2011). The resulting BAM files were merged using the *MergeSamFiles* tool from Picard v1.129 (http://broadinstitute.github.io/picard/) and further duplicates removed via SAMtools (version 0.1.1996b5f2294a) (Li et al. 2009) followed by indel realignment using GATK's *RealignerTargetCreator* and *Indel Realigner* (v3.3-0-g37228af) (McKenna et al. 2010) and then filtered via SAMtools for a mapping quality of 25. Genome coverage was calculated post mapping quality via QualiMap v2.13 (Okonechnikov et al. 2015). MapDamage (v2) (Jónsson et al. 2013) was used to assess the success of USER treatment where applicable, and a custom script used to softclip the first and last 2 base pairs of each read.

2.2.2.4. Modern Dataset Curation

2.2.2.4.1. Illumina Bovine HD 770K - Cattle

This dataset was curated by M.Teasdale containing locally held samples and the publicly available WIDDE dataset (Sempéré et al. 2015)(Appendix Table 2). Datasets were merged and filtered using PLINK (Purcell et al. 2007) as follows: autosomal sites only, removal of A/T and G/C SNPs, removal of sites with a genotyping rate less than 90%, removal of individuals with a genotype missingness greater than 10%, and the removal of previously identified problem SNPs. This left a dataset of 665,252 SNPs and 824 individuals from 45 breeds and two outgroups - Yak and Gaur. This dataset was then filtered for a maximum of 20 animals per breed, a minimum minor allele frequency of 1% and removal of sites with a genotyping rate less than 99%, leaving 620,025 sites in 512 individuals from 45 breeds and two outgroups.

2.2.2.4.2. 1000 Bulls Genomes - Cattle

Variant calls generated from Run6 of the 1K Bull run were previously filtered using the open source python parser PyVCF (https://github.com/jamescasbon/PyVCF/) (1K Bull Genomes lead group *per. comm.*), filtering options were as follows: removal of variants with 2 or more alternate alleles, alternate allele must be present on both the forward and reverse strands, overall quality of genotype - phred score QUAL ≥ 20, mapping quality - phred score MQ ≥ 30, minimum depth of coverage of 10 across all animals and a maximum depth of the median read depth +3* standard deviation read depth, opposing homozygotes filter - used a threshold of 10% to filter variants that produce too many opposing homozygotes, removed variants with the same basepair position, removed the lower QUAL INDEL when two were closer than 10 base pairs, remove lower QUAL variants if closer than 3 base pairs and removed SNPs closer than 5 base pairs to INDELs.

The dataset was then filtered using PLINK (Purcell et al. 2007) for autosomal biallelic sites posterior probability of 0.99, removal of individuals with a missingness greater than 10%, removal of one animal from a pair with a relatedness greater than 0.3 Pihat, maximum number of individuals per breed of 10. To help to counteract any possible reference bias the dataset was split into European taurine, Asian taurine, African taurine, indicine, African indicine admix and European taurine indicine admix. The European taurine dataset was subjected to a minimum allele frequency of 20%, whilst in the remaining five groups polymorphic sites were retained. If a site that had been lost in the European taurine dataset was a polymorphic site in the other groups then this site was retained. Sites with a genotyping rate less than 99% and a minor allele frequency less than 1% were then removed. After filtering the dataset comprised 13,046,412 SNPs and 390 individuals from 70 breeds as well as two outgroups - Yak and Gaur.

2.2.2.4.3. Ovine HapMap50K - Sheep

This dataset was curated by M.Teasdale containing the publicly available SNP50 HapMap dataset (http://www.sheephapmap.org/) (Appendix Table 3). The dataset was filtered as follows; autosomal sites with secure positions on the sheep genome OviAri 3.1, breeds were filtered for a maximum of 25 animals per breed and relatedness. The dataset was filtered for a minimum minor allele frequency of 1%, leaving 44,223 sites and 1499 individuals from 81 breeds.

2.2.2.5. SNP Calling

A distinction between low and high coverage samples was required when SNP calling, any sample with an average coverage below 8X was considered low coverage, anything above was considered high coverage.

2.2.2.5.1. Low Coverage

Many of the ancient genomes sequenced have a low average coverage, with the potential for many sites to be only covered by one read. Therefore, diploid calls can not be called with the necessary confidence. To circumvent this issue Skoglund and colleagues (2012) developed the pseudodiploid call; the calling of a site, picking one allele and duplicating it to produce a diploid call (pseudodiploid). If more than one read covers the base pair then the allele is chosen by consensus or at random, depending on which technique is chosen. The consensus call method is likely to bias towards the reference genome, while the random call method may select variation due to damage or sequencing error.

SNP calls were obtained using the pileup command in GATK (McKenna et al. 2010) on the q25 BAM files. The resulting pileup files were transformed into pseudodiploid PLINK (Purcell et al. 2007) ped and map files using pileupTools.py (https://github.com/teasdalm/pileupTools), filtering for a minimum base quality score of 30, calling and duplicating the consensus allele with ties solved by random pick and discarding triallelic sites.

2.2.2.5.2. High Coverage

SNP calls were obtained using SAMtools (v1.3.1) (Li et al. 2009) mpileup, options -s, -u, base quality 20 and mapping quality 30, and bcftools (v1.3.1) (http://SAMtools.github.io/bcftools/bcftools.html), options -m O.
Sites were then filtered using vcftools, bcftools and in-house custom scripts (P Delser *per. comm.*). Filtering was as follows:

1. Sites within 3bp of indels were removed
2. Indels were removed
3. Triallelic sites were removed
4. Minimum genotype quality of 30
5. Depth of coverage; minimum depth of 6 and maximum depth equal to the 95% quantile range
6. Strand bias - filtered out genotype calls with SP>13. This filter is to ensure that the variant allele is not disproportionately represented on one strand and therefore indicative of an error.
7. Heterozygosity ratio - filtered out genotype calls outside of the range 0.3-0.7. This filter is to ensure only genuine heterozygous calls are kept and potential spurious calls are removed.

2.2.3. General Population Genetic Analysis programmes

2.2.3.1. Principal Components Analysis

The PCA package LASER 2.0 (Wang et al. 2015) was utilised to visualise ancient data in the context of modern population variation, with the resulting output drawn in (R Core Team 2015).

The LASER 2.0 package employs PCA in conjunction with a Procrustes analysis. A reference panel of highly confident genotype calls, in this case the various modern datasets, is used to draw the initial PCA. The reference samples are then down sampled via the simulation of sequence data matching the coverage of the study sample. A ‘PCA ancestry map’ is then calculated using the simulated sequences for the reference samples and the genuine sequence data for the study sample. This ‘ancestry map’ is then projected using Procrustes analysis onto the initial PCA space, the down sampled reference samples are then utilised to place the study sample in its own position on the initial PCA (Wang et al. 2015). When multiple study samples are visualised together their position is calculated separately, therefore the study samples do not affect the placement of one another.

Modern datasets were prepared for LASER by M. Teasdale. The preparation of ancient samples for LASER followed the suggested pipeline; SAMtools *mpileup* using options of minimum mapping quality 30, minimum base quality 20 and the -B to reduce the effect of reference bias in the calls. Input files for LASER were generated by running the pileup2seq.py script that is available with the package.

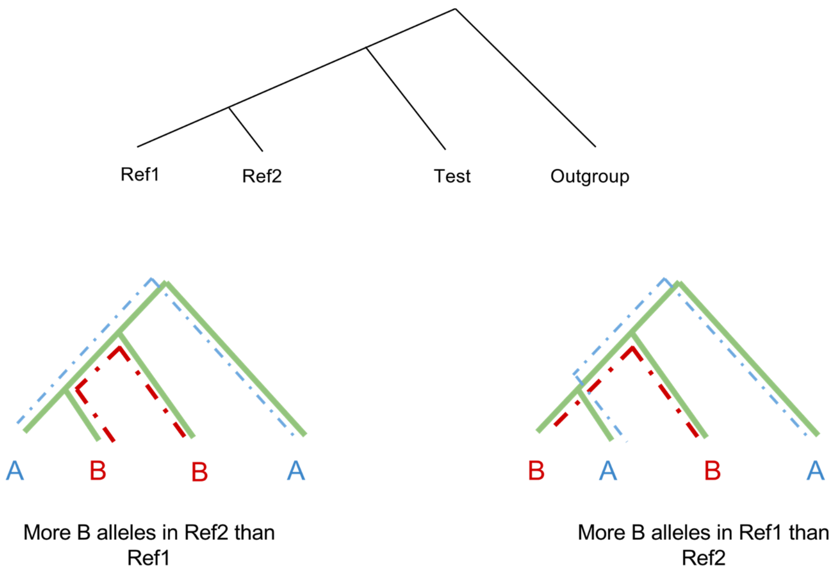
2.2.3.2. Ancestry Estimation

ADMXITURE (Alexander et al. 2009), the maximum-likelihood model-based ancestry estimation programme, was used to estimate the ancestry fractions of the ancient samples. This programme performs an unsupervised model clustering based upon SNP data, to ask what population an individual sample belongs to by estimating the maximum likelihood of an individuals ancestries. Individuals are then clustered together based upon genetic similarity dependent of the user-assigned number of clusters (K).

Prior to the running of admixture the merged ancient/modern datasets were subjected to linkage disequilibrium pruning, this was run in PLINK (Purcell et al. 2007), options for pruning are mentioned in individual chapters. ADMIXTURE was run with default settings, each value of K was run ten times with a randomly generated seed. For each K the replicate with the lowest likelihood was chosen to be visualized in R (R Core Team 2015).

2.2.3.3. D-statistic

The *D*-statistic is a four population test that can calculate the genetic closeness of an individual to two reference populations (Green et al. 2010) . The *D*-statistic, can be utilised to test for clades between populations as well a formal test for admixture (Fig 2.1). If gene flow has not occurred between the test population and the two references populations there will be no significant difference in the number of derived alleles in the two reference populations. However if there is a significant difference in number of derived alleles between the two reference populations then either an ABBA or a BABA tree is returned; ABBA indicating gene flow between Ref2 and Test, while BABA indicates gene flow between Ref1 and Test (Fig 2.1).



More B alleles in Ref2 than Ref1

More B alleles in Ref1 than Ref2

Fig 2.1. *D-statistic (Ref1, Ref2, Test, Outgroup). Shows the two trees (ABBA, BABA) that are expected when there is a significant deviation away from the null hypothesis of no gene flow. Ancestral alleles are symbolised by A, while derived alleles are symbolised by B.*

To compute the *D*-statistic, the *qpDstat* test from the AdmixTools package (Patterson et al. 2012) was used. *qpDstat* calculates the *D*-statistic for any four populations or sample frequencies defined as (w,x,y.z) where:

$$D = numerator/denominator$$

With:

$$num = (w - x)(y - z)$$

$$den = (w + x - 2wx)(y + z - 2yz)$$

If the *Z*-score is positive then gene flow has occurred between either W and Y or X and Z. While is the *Z*-score is negative then gene flow has occurred between either W and Z or X and Y.

AdmixTools was run with default settings. A weighted block jackknife was used to obtain standard errors on the statistic and a *Z*-score for high statistical significance ($0.001 \geq p$) was set at 3 (Green et al. 2010; Reich et al. 2010). However results with lower Z-scores are still discussed within this thesis.

2.2.3.4. F-statistics

2.2.3.4.1. f_3-test for Admixture

A formal test of admixture, the three population test, is known as the f_3-test (Reich et al. 2009; Patterson et al. 2012). This assesses whether admixture between two source populations took place in the ancestry of a target population by utilising allele frequency ratios (Fig 2.2).

This calculates the expected value of:

$$f3(C;A,B)$$

or

$$F3 = E(fC - fA, fC - fB)$$

Where f_c,f_a,f_B are equal to the allele frequencies in C,A&B (Reich et al 2009).

If the result of f_3(C;A,B) is negative this indicates admixture between A and B did occur in the ancestry of C, if it is positive this suggests that C is un-admixed. However unlike the *D*-statistic, the f_3 is not impervious to drift, therefore inability to detect admixture does not discount it from having occurred.

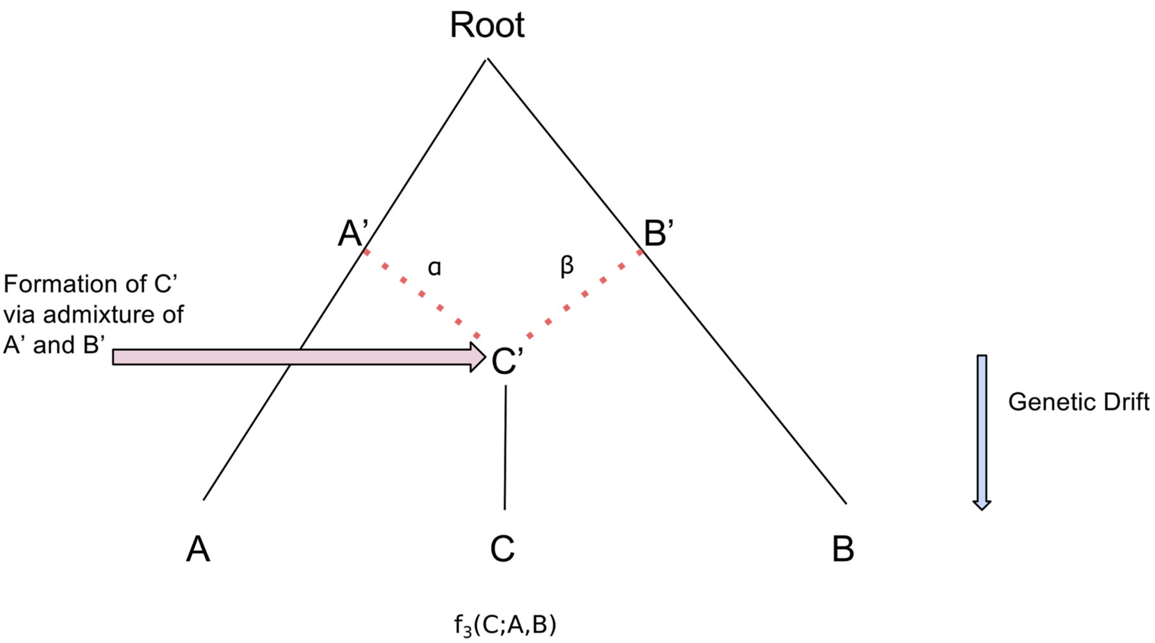


Fig 2.2. *f_3 statistic tree adapted from Reich et al., 2009. Populations A’ and B’ split from the root population, population C’ is then formed via the admixture of A’ and B’. The modern populations of A,B&C are then formed via drift.*

The f_3-statistics were calculated using the *qp3pop* package from ADMIXTOOLS (Patterson et al., 2012), using default settings. Where the target population was pseudodiploid the inbreed:YES option was applied. Significance was determined by a *Z*-score of -3 or lower.

2.2.3.4.2. *f_3-test, The Outgroup Case*

The *f_3*-statistic can also be utilised to assess the shared drift between populations, known as the outgroup *f_3*-statistic (C; A, B) (Patterson et al., 2012) (Fig 2.3). A higher *f_3* value suggests a greater proportion of shared drift between the two test populations of A and B. When a number of samples are rotated through A and B in a pairwise fashion, the phylogenetic relationship between a number of populations can be determined by utilising allele frequency ratios.

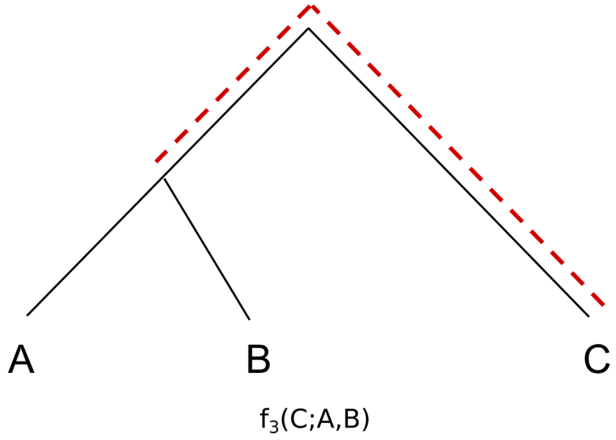


Fig 2.3. *Outgroup f_3 statistic tree as described by Patterson et al., 2012. The red line indicates shared drift of A and B from the outgroup C.*

The outgroup *f_3* statistics were performed using the *qp3pop* package from ADMIXTOOLS (Patterson et al., 2012), using default settings.

2.2.3.4.3. *f_4-ratio estimation*

The *f_4*-ratio statistic is similar to the *D*-statistic, however unlike the *D*, it can estimate the size of ancestry proportions (Reich et al. 2009). It is possible to calculate these proportion even if the true ancestral populations are not accessible (Patterson et al., 2012).

The *qpF4ratio* option in the package AdmixTools (Patterson et al., 2012) was run on default settings, using a significance of 3 for the *Z*-score.

2.2.4. Radiocarbon dating

Selected samples (Appendix Table 1) were radiocarbon dated at the Chrono Centre, Queens University Belfast.

2.3. References

Alexander, D.H., Novembre, J. & Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, 19(9), pp.1655–1664.

Andrews, S. & Others, 2010. FastQC: a quality control tool for high throughput sequence data.

Briggs, A.W. et al., 2010. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic acids research*, 38(6), p.e87.

Brotherton, P. et al., 2007. Novel high-resolution characterization of ancient DNA reveals C> U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic acids research*, 35(17), pp.5717–5728.

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

Gamba, C. et al., 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nature communications*, 5, p.5257.

Ginolhac, A. et al., 2011. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* , 27(15), pp.2153–2155.

Green, R.E. et al., 2010. A draft sequence of the Neandertal genome. *Science*, 328(5979), pp.710–722.

Jónsson, H. et al., 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* , 29(13), pp.1682–1684.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , 25(16), pp.2078–2079.

Li, H. & Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* , 25(14), pp.1754–1760.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1), pp.10–12.

McKenna, A. et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), pp.1297–1303.

Meyer, M. & Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor protocols*, 2010(6), p.db.prot5448.

Okonechnikov, K., Conesa, A. & García-Alcalde, F., 2015. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* . Available at: http://dx.doi.org/10.1093/bioinformatics/btv566.

Orlando, L. et al., 2011. True single-molecule DNA sequencing of a pleistocene horse bone. *Genome research*, 21(10), pp.1705–1719.

Pääbo, S. et al., 2004. Genetic analyses from ancient DNA. *Annual review of genetics*, 38, pp.645–679.

Patterson, N. et al., 2012. Ancient admixture in human history. *Genetics*, 192(3), pp.1065–1093.

Purcell, S. et al., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*, 81(3), pp.559–575.

R Core Team, 2015. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Reich, D. et al., 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, 468(7327), pp.1053–1060.

Reich, D. et al., 2009. Reconstructing Indian population history. *Nature*, 461(7263), pp.489–494.

Schubert, M. et al., 2012. Improving ancient DNA read mapping against modern reference genomes. *BMC genomics*, 13, p.178.

Sempéré, G. et al., 2015. WIDDE: a Web-Interfaced next generation database for genetic diversity exploration, with a first application in cattle. *BMC genomics*, 16, p.940.

Skoglund, P. et al., 2012. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. Science, 336(6080), pp.466–469.

Vilstrup, J.T. et al., 2013. Mitochondrial phylogenomics of modern and ancient equids. *PloS one*, 8(2), p.e55950.

Wang, C. et al., 2015. Improved ancestry estimation for both genotyping and sequencing data using projection procrustes analysis and genotype imputation. *American journal of human genetics*, 96(6), pp.926–937.

Chapter 3. The Screening of Archaeological Material for DNA Preservation

3.1. Introduction

The screening of archaeological and palaeontological bones dates back to the beginnings of the ancient DNA (aDNA) field in the 1980s. Inferring past genetic variation and demographic events from modern day can be challenging, especially for domesticates which have within the last several hundred years undergone breed formation and human mediated migrations (Kijas et al. 2009; Speller et al. 2013; Decker et al. 2014; Lenstra et al. 2014). Therefore ancient DNA offers the scientific community a chance to directly study the genetic variation of the past. Unfortunately the study of aDNA is complicated by the post-mortem degradation of DNA (Lindahl 1993). aDNA is characterised by short, fragmented, damaged and easily contaminated sequences (Hofreiter et al. 2001). The breakdown by microorganisms and nucleases help to reduce the size of the DNA sequences, while chemical degradation such as deamination and depurination causes changes in the coding sequences (Pääbo et al. 2004). These chemical processes result in the characteristic aDNA damage pattern of cytosine to thymine and guanine to adenine changes, primarily at the ends of sequencing reads.

Damage to aDNA varies in different climates due to the post-mortem *in situ* environment of a bone. Due to the DNA preserving nature of the freezing temperatures of permafrost, aDNA is more readily accessible in samples originating from such regions than from more southern latitudes (Smith et al. 2001; Smith et al. 2003; Brace et al. 2012; Allentoft et al. 2012; Orlando et al. 2013). However sampling only from environments more amenable to DNA preservation is a limited strategy. Therefore improvements in DNA extraction protocols have been developed (Orlando et al. 2011; Dabney et al. 2013; Damgaard et al. 2015; Gamba et al. 2015), while other have suggested targeting specific skeletal elements (Gamba et al. 2014).

In this chapter the aDNA screening results for 183 cattle and sheep bones are presented and the DNA preservation levels discussed in relation to previously published data and our understanding of DNA degradation. Molecular sexing results for 90 animals from a timeseries of Britain and Ireland are briefly examined in respect to herd management strategies.

3.2. Materials

DNA content was screened in 183 bones morphologically identified as cattle or sheep. These samples originated from sites with a geographic range from the Near East to the Atlantic Edge and temporally from approximately 250,000 BP to approximately 800 BP (Appendix Table.1, Table 3.1 & Fig 3.1). The majority of the samples were sourced from Museums and zooarchaeologists across Britain and Ireland.

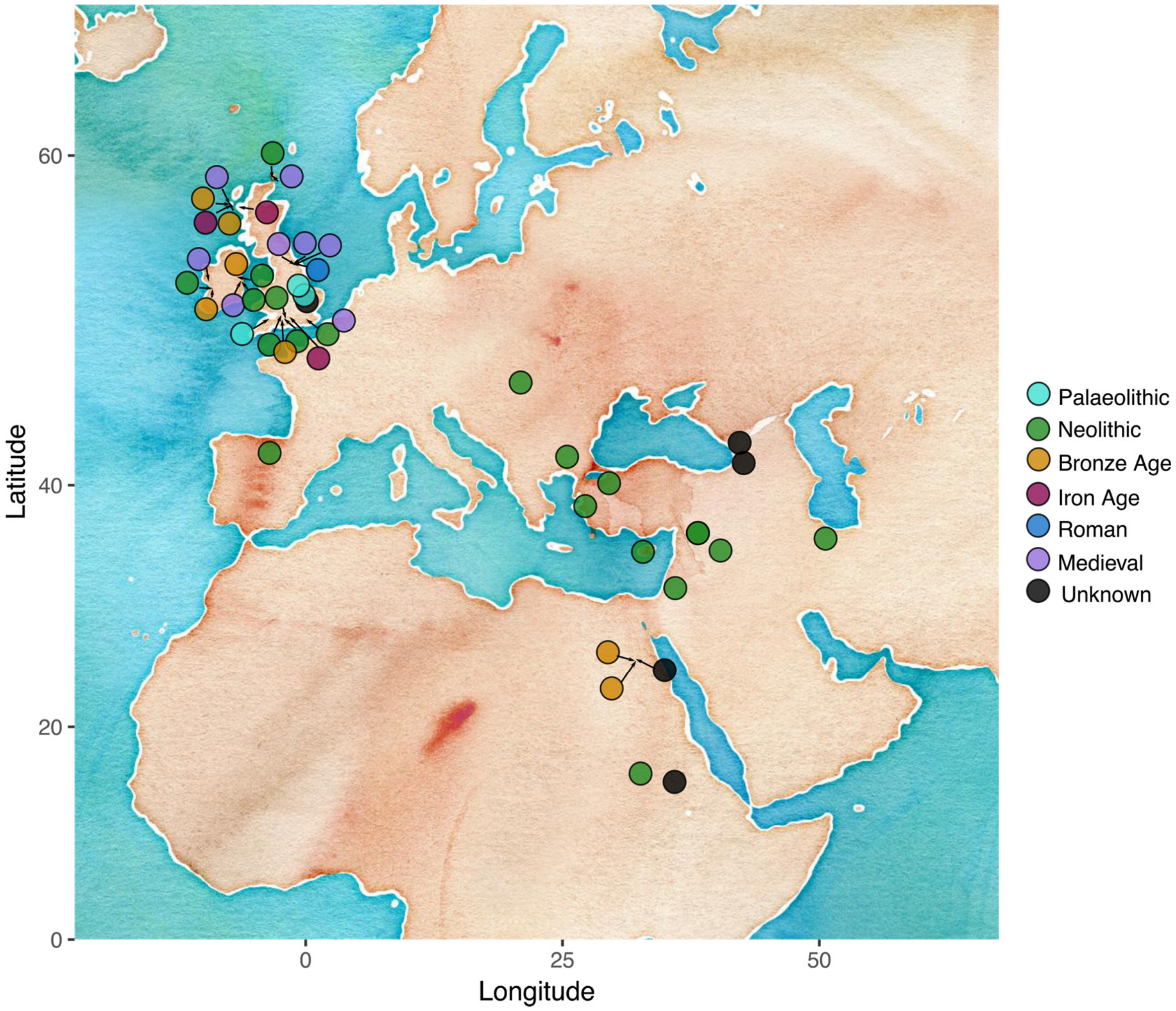


Fig 3.1. *Distribution of sites from which samples were screened for DNA survival. The predominant cultural period of the site is indicated by the colour of the circle.*

3.3. Methods

Samples were extracted and prepared for NGS sequencing in an ancient DNA laboratory, as detailed in section 2.1. These were sequenced at TrinSeq (Trinity Genome Sequencing Laboratory, Trinity College Dublin, Ireland), using 50 bp single-end sequencing and a PhiX control at 1%. A small percentage of libraries were screened on a Illumina HiSeq™ 2500 by Macrogen Inc, Seoul, Republic of Korea.

Sequencing reads were processed as given in section 2.2.1.4. Where necessary damage patterns were accessed with MapDamage (section 2.2.1.5) and when read count allowed, molecular sexing performed (2.2.1.6). Generally, samples with endogenous percentages over 10% were then USER treated and sent for further sequencing on an Illumina HiSeq 2000 or HiSeq 25000 at Beckman Coulter Genomics (USA), BGI Next Generation Sequencing Services (CHN) or Macrogen (KOR).

3.4. Results and Discussion

Results and analysis of post USER treated samples can be found in Chapters 4, 5 and 6.

3.4.1. Overall

Species identified through DNA included *Bos Taurus* (cattle), *Ovis aries* (sheep), *Capra hircus* (goat), *Cervus elaphus* (red deer) as well as *Sus* (pig) species (Table 3.1). Percentage endogenous DNA varied across the dataset from essentially no endogenous DNA survival (0.01%) to a high of 71.12% in an approximately 2,940 year old *Bos taurus* sample from the Outer Hebrides, UK.

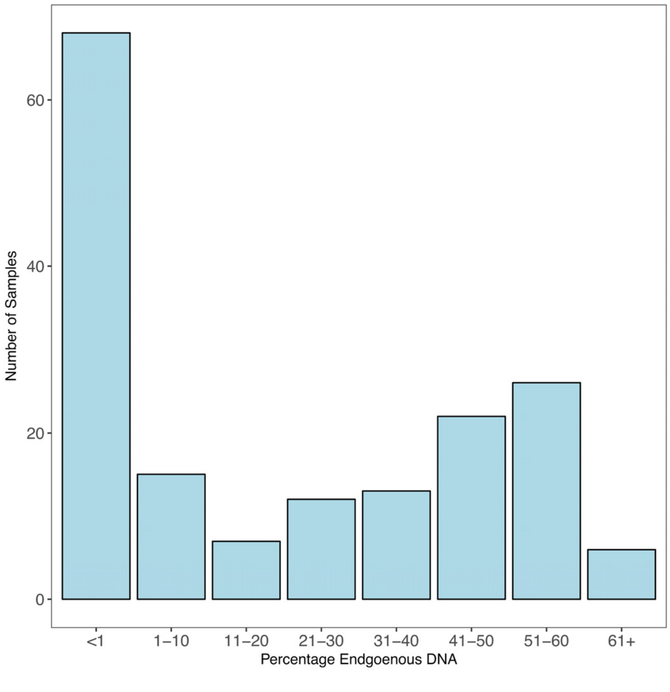


Fig 3.2. *Number of samples binned by percentage endogenous DNA survival for Bos sp and ovid/caprine samples. The percentage endogenous presented was calculated post removal of duplicates and filtering of mapping quality of 25.*

Usually, aDNA samples have low levels of fragmented and poorly preserved endogenous DNA (Hofreiter et al. 2001). As expected DNA was not preserved in a large proportion of *Bos* species and ovid/caprine samples (Figure 3.2). However, a substantial number of the samples had endogenous DNA percentages of 10 or more. These high percentage endogenous are not unique. Previously published studies describe high endogenous samples, notably a 70% Denisovan sample from a cave site (Meyer et al. 2012) and an 85.4% Hungarian Neolithic sample (Gamba et al. 2014). With the expense of sequencing, targeting high endogenous DNA samples is advantageous when attempting to develop an ancient dataset.

3.4.2. Petrous vs Non Petrous

A variety of skeletal elements were sampled, however one bone element in particular has proven useful in regards to DNA survival; the petrous bone (Figure 3.3). Relatively recently the petrous portion of the temporal bone was identified as the bone element of choice for aDNA studies (Gamba et al. 2014). Several bones from the same skeleton were screened for percentage endogenous DNA and the petrous portion excelled. Since then a number of papers have utilised this bone including the publication of the first complete ancient dog genome from a Neolithic Irish sample (Frantz et al. 2016).

***Fig 3.3.** Sample New2/VEM123. A Disarticulated Neolithic Bos taurus Petrous Bone from Newgrange, Ireland (Sample New2). Scale in centimeters.*

The targeting of the petrous bone, stemming from the Latin *petrosus* - ‘stone-like’, revolutionised this thesis. This bone element had previously been identified as very dense, approximately 10% denser than the middle shafts of long bones of wildebeest, Zebra and Reindeer (Lam et al. 1999). Here, percentage endogenous DNA results obtained from petrous bones were far higher than endogenous percentages from any other bone element (Figure 3.4). The highest percentage endogenous, post quality filtering, from a non petrous bone was 9.48% from a Neolithic Anatolian *Bos taurus* bone, Turk11, in comparison to the 71.12% achieved from a petrous bone. Out of 55 non petrous samples, only 5 samples contained endogenous percentages greater than 1. While the petrous bone is not impervious to failure, the number of petrous samples producing less than 1% endogenous was 33 compared to 99 petrous samples that contained greater than 1% endogenous. Targeting the petrous bones increased the chances of achieving an endogenous percentage greater than 1. However it should be noted that the majority of the non-petrous bone samples were older than the majority of the petrous bones (Figure 3.4) and were more likely to have originated from latitudes closer to the equator. Therefore DNA survival in these samples was already less likely to occur. While DNA has been shown to survive in a 560-780 kyrBP Equid persevered in the permafrost of Alaska (Orlando et al. 2013), non of these samples presented here have had the benefit of preservation in the permafrost.

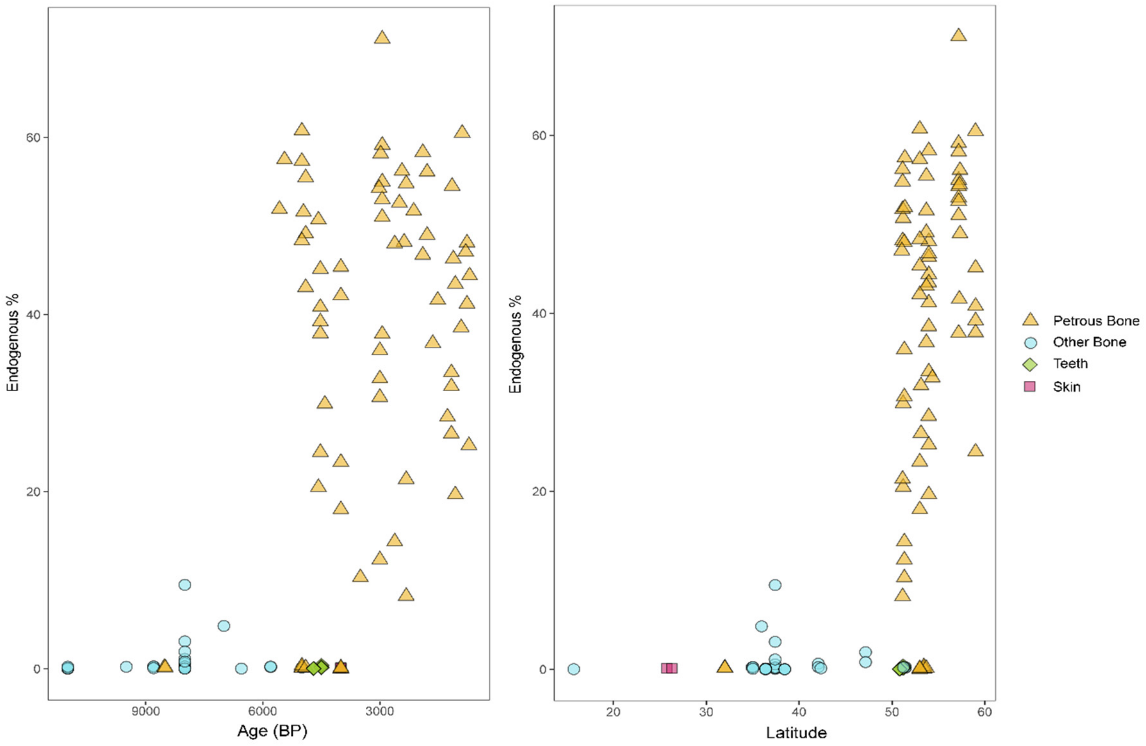


Fig 3.4. *Percentage endogenous DNA survival post removal of duplicates and filtering of mapping quality of 25 for Bos specimens plotted, against age and latitude. Type of skeletal element is denoted by shape. The two Pleistocene aurochs were not included in this graph.*

Post-mortem DNA decays over time, additionally, the rate of this decay has been demonstrated to vary depending on the temperature of the post-mortem environment; known as ‘thermal age’, the warmer the environment, the faster the decay (Lindahl 1993; Smith et al. 2003; Allentoft et al. 2012). Therefore post-mortem DNA survival is less likely to occur in warmer environments such as in Anatolia, the Levant and Africa.

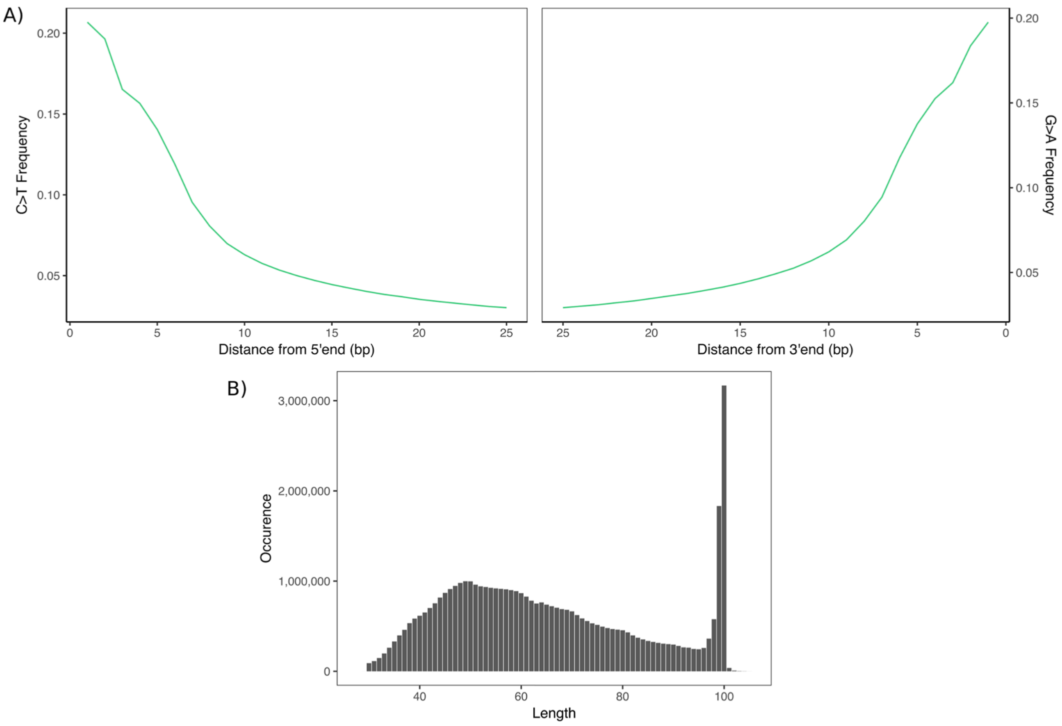


Fig 3.5. *DNA damage pattern results for sample a Medieval Irish petrous bone sample Dub2. Checking for these damage patterns in ancient DNA sequencing reads is a widely used method for validation of genuine ancient DNA. Sequencing reads were produced on an Illumina HiSeq 2000 by Beckmann Coulter genomics. A) Demonstrates the increase in deamination at towards the end of the sequencing. B) The read length distribution pattern; displaying a fragmented DNA pattern caused by the breakdown of DNA sequences via nucleases and microorganisms.*

When both teeth and petrous from the same Neolithic/Bronze Age site of Durrington Walls (UK) were screened for endogenous DNA percentage the difference between the two elements was stark. Non of the five teeth contained greater than 1% endogenous DNA, while all three petrous bone contained over 20% endogenous DNA. Little is known about the microscopic structure of the petrous bone, perhaps its dense nature and composition, such as pore size, enables it to more readily withstand microbial and fungal attack, known to be part of the post-mortem bone and DNA degradation process (Lindahl 1993; Collins et al. 2002; Jans et al. 2004; Gamba et al. 2014). If exogenous microbial attack is limited, then once damage from the nucleases released post cell death as well as putrefaction of soft tissue has passed (Darzynkiewicz et al. 1997; Campos et al. 2012), the main mode of DNA degradation would be chemical such as the processes of deamination and depurination; responsible for the damage patterns associated with authenticity of aDNA (Fig3.5). It is also possible that the distance from the intestinal system may help to decrease the amount microbial attack from endogenous gut bacteria (Jans et al. 2004). While future study on the survival of DNA in the petrous bone is necessary, the data here does supports previous claims, that targeting the petrous bone increases the chance of high endogenous DNA percentages (Gamba et al. 2014).

When considering only petrous samples, no trend is visible in these data between age or latitude and percentage endogenous DNA retrieved. The *in situ* post-mortem environment is important for DNA survival. While temperature, and temperature stability, are key to the longevity of DNA (Lindahl 1993; Smith et al. 2003), there are other factors that may affect the rate of decay including the species of microorganism in the soil, salt content, pH and additionally the availability of oxygen and free water (Lindahl 1993; Campos et al. 2012). Therefore between sites there will be differences in the preservation of DNA within petrous bones from high latitudes. DNA survival is also known to vary within a single assemblage (Ottoni et al. 2009). Rapid loss of DNA has been shown to occur in the first year of death (Campos et al. 2012), while this loss may be associated with putrefaction of soft tissue, it may also be associated with the taphonomic environment of the individual sample. Microbial analysis of human and animal bone highlighted differences in microbial attack between the two. This is suggested to be due to differences in burial environments, animals buried in waste pits, humans in single graves, exposing the samples to different bacterial populations and humic factors that may act as inhibitors of bacterial attack on bone (Jans et al. 2004). Therefore differences in burial environments within a single archaeological site, and even butchery practices, will likely have influenced bone preservation and therefore DNA preservation in ancient samples.

3.4.3. Molecular Determination of the Sex of Cattle and Sheep within a British and Irish Transect

Herd management strategies have traditionally been studied by zooarchaeologists via kill-off patterns. This entails calculating the relative representation of different age-groups and the sex of the bones. Age is usually assessed via wear of mandibular teeth, while sex is determined by the shape of metacarpals or more recently via aDNA using a single polymorphism in the *ZFXY* gene (Payne 1973; Svensson et al. 2008; McGrory et al. 2012). While age could not be assessed for this study, the sex of 90 animals, 65 cattle and 25 sheep was assessed (Figure 3.6).

In both species there are more females than males. In cattle the number of females (43) is over double the numbers of males (20), while in sheep the number of females (16) is just shy of double the number of males (9). The kill-off pattern for a beef herd would be the slaughter of cattle of both sexes at 19-36 month age, while a dairy herd would see a slaughter of surplus calves under six months with a sex ratio of two males for every female in the the slaughtered calves (Payne 1973; Svensson et al. 2008; McGrory et al.

2012). Within the dairy herd there would then be a predominance of females over males over the 6-month age group.

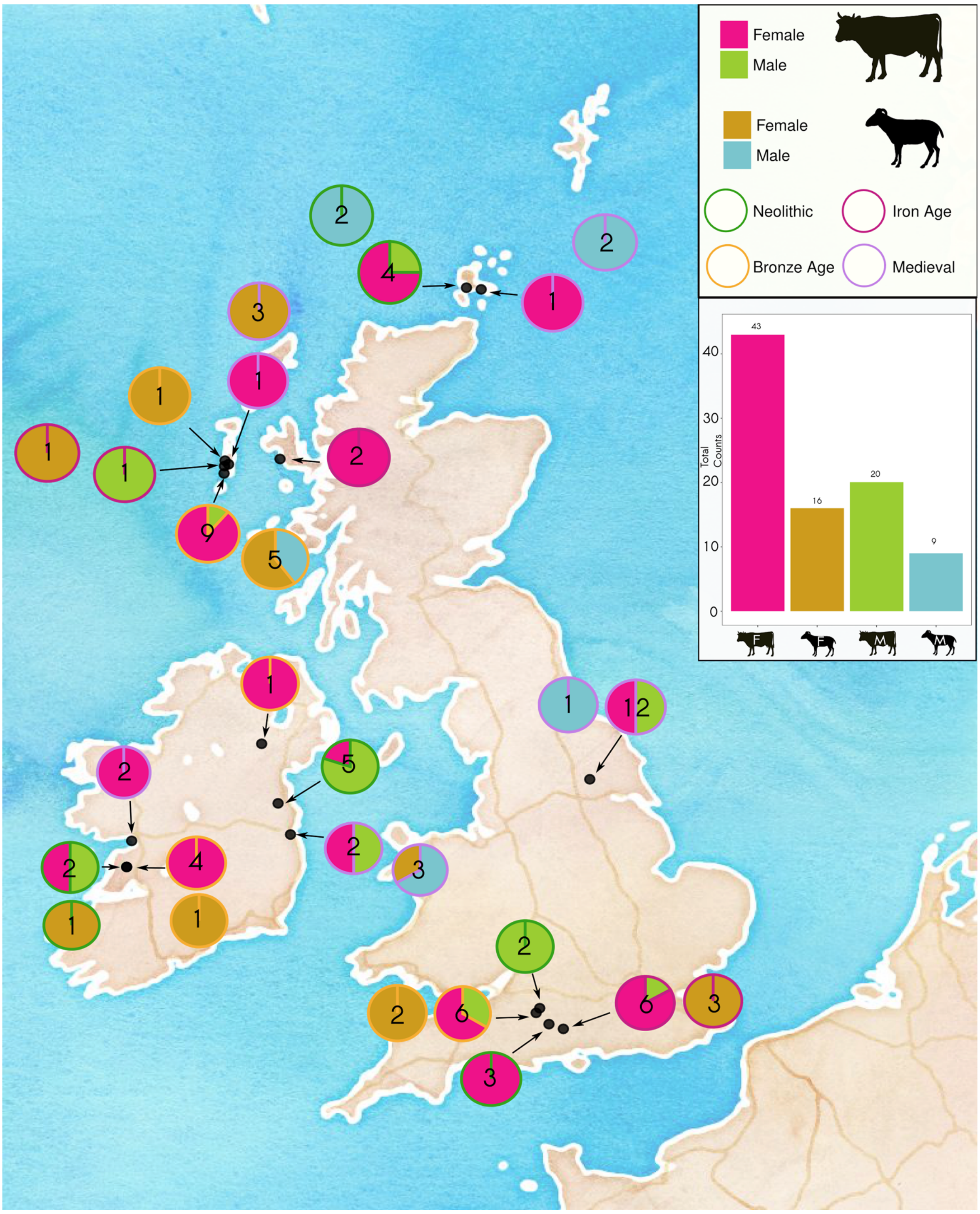


__Fig 3.6.__ Molecular sexing of domestic cattle and sheep from archaeological samples from Britain and Ireland grouped by site. When sheep and cattle are present at both sites two pie charts sit next to each other, with the colour of the fill denoting cattle or sheep. The number of individuals sampled sits inside the pie, with the outside of the pie coloured as to the predominate cultural time period at the site. The several sites sampled from York are combined into one site. Female animals are more common at the majority of the sites than males in both species. Individual sample results are listed in Table 3.2.

While age profiles for these samples could not be calculated, it is interesting that overall there is a majority of females within the herds of cattle and flocks of sheep. The predominance of either sex differs between sites. The South British Early Neolithic site of Bishops Cannings produced two male animals sampled from partially intact skulls. Very little information exists on this site, perhaps these animals were load bearing oxen or perhaps, as the skulls were intact they were some sort of ceremonial offering. The famous Neolithic ceremonial site of Newgrange (Ireland) produced 3 Neolithic males, 1 Neolithic female and 1 Iron Age male. Additionally, both a male and a female were sequenced from the court tomb Parknabinia (Co. Clare Ireland). In contrast more females than males were sequenced from the Neolithic ceremonial site of The Ness of Brodgar, Orkney. These animals are believed to have been slaughtered at the same time during the abandonment of the temple like structure in which they were found. Isotopic analysis suggests that a number of these animals were local to Orkney (Ingrid Mainland *per. comm.*). No pattern to the sex of the cattle found at these Neolithic ceremonial/burial sites exists.

Intriguingly, at two sites suggested to have dairying herd structure there is a prevalence of female animals; the Bronze Age site of Cladh Hallan (Payne 1973; Svensson et al. 2008; McGrory et al. 2012) and the Iron Age site of Danebury (Copley et al. 2005). While the samples sizes at each site are small, it is curious that the petrous to survive at these sites were female. Perhaps smaller younger samples, more likely to be male in a dairy herd, did not survive the in archaeological record as well. The sites at the city of York yielded equal numbers of females and males; perhaps as a market town animals were brought from the fields to the city to be traded and slaughtered and therefore the ratio of the two sexes can not be used to theorise herd structure.

Overall the sex ratio of sheep indicates a predominance of females within flocks. In general, without age information for these samples and greater numbers no strong conclusions can be drawn from these results. However, these results do highlight that it is likely that female domesticate remains form a greater proportion of the zooarchaeological record. This result in itself is suggestive of some form of management strategy, as it would be expected that equal numbers of males and females to be born. For some reason, perhaps lack of preservation of younger petrous bones, less males have been sampled. The petrous bones selected for aDNA anaylsis were primarily judged on outward appearance, if the bone did not have a reasonable outward preservation they were not removed from the assemblage, perhaps biasing this result.

3.5. Conclusion

In conclusion, the screening results of over 183 samples have shown a variety of aDNA preservation levels. High percentages of endogenous DNA have been demonstrated in petrous bones from Northern Europe, helping to support previous claims of potentially superior DNA preservation within this skeletal element when compared to others. When numerous samples from the same site are processed, molecular sexing can offer an insight into ancient herd structures. While age cannot be currently determined from the petrous bone and therefore full kill-off patterns established, the results presented indicate a higher number of females of both cattle and sheep were present in Britain and Ireland, itself suggestive of some form of herd management. Further sequencing and analysis of a number of these cattle and sheep are presented in Chapters 4,5&6 of this thesis (Appendix Table1).

Table 3.1. *The endogenous percentage of samples screened for DNA survival. Where known, site and country, cultural era, bone element, direct radiocarbon date and endogenous % are given. A more comprehensive table is given in Appendix (Table 1).*

Lab Name	Thesis Name	Site	Country	Era	Speces	Bone element	Direct Radiocarbon cal BC/AD	% Endogenous
VEM001	VEM001	Ais Yiorkis	Cyprus	Neolithic	Bos sp	longbone shaft		0.23
VEM002	VEM002	Bouqras	Syria	early Neolithic	Bos sp	fragment		0.24
VEM004	VEM004	Bouqras	Syria	early Neolithic	Bos sp	fragment		0.27
VEM006	VEM006	Bouqras	Syria	early Neolithic	Bos sp	fragment		0.07
VEM007	VEM007	Mentese	Turkey	Neolithic	B. taurus	mandible		0.06
VEM008	VEM008	Mentese	Turkey	Neolithic	B. taurus	Phalanx		0.13
VEM009	Turk11	Mentese	Turkey	Neolithic	B. taurus	Femur		9.48
VEM010	VEM010	Mentese	Turkey	Neolithic	B. taurus	Carpal		0.10
VEM011	Turk12	Mentese	Turkey	Neolithic	B. taurus	Calcaneum		0.57
VEM012	VEM012	Mentese	Turkey	Neolithic	B. taurus	Phalanx		3.10
VEM013	Turk13	Mentese	Turkey	Neolithic	B. taurus	Carpal		0.22
VEM014	Turk06	Mentese	Turkey	Neolithic	B. taurus	Phalanx		1.12
VEM016	VEM016	unknown	unknown	Paelolithic	?	Bone tool		0.05
VEM017	VEM017	unknown	unknown	Paelolithic	?	Bone tool		0.02
VEM018	VEM018	unknown	unknown	Paelolithic	?	Bone tool		0.05
VEM030	VEM030	Yabalkovo	Bulgaria	Neolithic	Bos sp	bone		0.65
VEM031	VEM031	Yabalkovo	Bulgaria	Neolithic	Bos sp	bone		0.26
VEM063	VEM063	Khattara	Egypt	unkown	Bos sp	distal humerus		0.03
VEM065	VEM065	Kadero	Sudan	Neolithic	Bos sp	bone		0.03
VEM066	VEM066	Khasm-el-Girba	Sudan	unknown	Bos sp	distal fragment metatarsal		0.03
VEM068	VEM068	Dja'de	Syria	PPNA	B. primigenius	tarsel		0.10
VEM069	VEM069	Dja'de	Syria	PPNA	B. primigenius	tarsel		0.09

VEM070	VEM070	Dja'de	Syria	PPNA	B. primigenius	phalanx		0.09
VEM071	VEM071	Jerf El Ahmar	Syria	PPNA	B. primigenius	metatarsal		0.09
VEM072	VEM072	Jerf El Ahmar	Syria	PPNA	B. primigenius	phalanx		0.06
VEM074	VEM074	Dja'de	Syria	PPNA	B. primigenius	phalanx		0.10
VEM075	VEM075	Jerf El Ahmar	Syria	PPNA	B. primigenius	phalanx		0.06
VEM076	VEM076	Jerf El Ahmar	Syria	PPNA	B. primigenius	tarsel		0.04
VEM077	VEM077	Jerf El Ahmar	Syria	PPNA	B. primigenius	phalanx		0.01
VEM078	VEM078	Jerf El Ahmar	Syria	PPNA	B. primigenius	bone		0.05
VEM079	VEM079	Jerf El Ahmar	Syria	PPNA	B. primigenius	phalanx		0.05
VEM080	VEM080	Jerf El Ahmar	Syria	Neolithic	B. primigenius	bone		0.04
VEM081	VEM081	Yeşilova Höyüğü	Turkey	Neolithic	Bos sp	bone		0.04
VEM082	VEM082	Yeşilova Höyüğü	Turkey	Neolithic	Bos sp	bone		0.02
VEM083	VEM083	Yeşilova Höyüğü	Turkey	Neolithic	Bos sp	bone		0.02
VEM084	VEM084	Maral Tappeh	Iran	Chalcolithic/ Bronze Age	Bos sp	Scapula		4.84
VEM085	VEM085	Ecsegfalva	Hungary	early Neolithic	B. primigenius	carpal		1.94
VEM086	VEM086	Ecsegfalva	Hungary	early Neolithic	B. primigenius	distal phalanx		0.82
VEM087	VEM087	Cueva del Mirair	Spain	Late Neolithic	Bos sp	bone		0.13
VEM088	VEM088	Windmill Hill	UK	Neolithic	B. primigenius	femur		0.19
VEM089	VEM089	Windmill Hill	UK	Neolithic	B. primigenius	proximal phalange		0.25
VEM090	VEM090	Thoth Hill	Egypt	XI Dynasty - Bronze Age	B. taurus	soft tissue		0.13
VEM091	VEM091	Sakajia	Georgia	unknown	Bison bonasus	bone		0.15
VEM092	VEM092	Sakajia	Georgia	unknown	Bison bonasus	bone		0.12
VEM093	VEM093	Mgvimari	Georgia	unknown	Bison bonasus	bone		0.04
VEM095	Dub2	Dublin	Ireland	Medieval - Viking	B. taurus	Petrous		26.54
VEM096	Dub1	Dublin	Ireland	Medieval - Viking	B. taurus	Petrous		31.89
VEM097	VEM097	Kilshane	Ireland	Neolithic	B. taurus	Petrous		0.11
VEM098	VEM098	Kilshane	Ireland	Neolithic	B. taurus	Petrous		0.10
VEM099	VEM099	Kilshane	Ireland	Neolithic	B. taurus	Petrous		0.34
VEM100	Yor1	Tanner Row	UK	Roman	B. taurus	Petrous	2 BC - 135 AD	58.32
VEM101	Yor7	Coppergate	UK	Medieval - 9th Century	B. taurus	Petrous		33.46
VEM102	Yor2	Hungate	UK	Medieval -13th Century	B. taurus	Petrous		41.21
VEM103	Yor8	Hungate	UK	Medieval - 13-14th Century	B. taurus	Petrous		25.23

VEM104	VEM104	Deir el Bahri	Egypt	XI Dynasty - Bronze Age	B. taurus	soft tissue		0.10
VEM105	Yor5	Hungate	UK	Roman/Medieval	B. taurus	Petrous		44.40
VEM106	Yor9	Fishergate	UK	Medieval - 8th Century	B. taurus	Petrous		28.45
VEM107	VEM107	Hungate	UK	Medieval - 10th Century	Equus f. caballus	Petrous		63.91
VEM108	Yor6	Hungate	UK	Medieval - 10th Century	B. taurus	Petrous		43.45
VEM109	Yor3	Coppergate	UK	Medieval - 13th Century	B. taurus	Petrous		48.11
VEM110	Yor13	Hungate	UK	Medieval - 11th Century	O. aries	Petrous		17.71
VEM111	Yor4	Coppergate	UK	Medieval- 10th Century	B. taurus	Petrous		46.31
VEM112	Yor10	Hungate	UK	Medieval - 10th Century	B. taurus	Petrous		19.67
VEM113	Yor11	Tanner Row	UK	Roman	B. taurus	Petrous		46.73
VEM115	New2	Newgrange	Ireland	Iron Age	B. taurus	Petrous	321-425 AD	36.76
VEM116	New3	Newgrange	Ireland	Neolithic	B. taurus	Petrous		49.12
VEM117	New4	Newgrange	Ireland	Neolithic	B. taurus	Petrous		43.08
VEM118	VEM118	unknown	Israel	unknown	ovid/caprid	Petrous		0.05
VEM119	VEM119	Hovk Cave	Armenia	Paleolithic	Caprid sp	Petrous	too old to date	28.28
VEM121	Yor12	Hungate	UK	11-12th Century	B. taurus	Petrous		38.54
VEM122	VEM122	Newgrange	Ireland	Neolithic	B. taurus	Petrous		0.14
VEM123	New1	Newgrange	Ireland	Neolithic	B. taurus	Petrous	2795-2580 BC	55.47
VEM124	New5	Newgrange	Ireland	Neolithic	B. taurus	Petrous	3024-2894 BC	51.57
VEM125	HF1	Haughey Fort	UK	Bronze Age	B. taurus	Petrous		32.77
VEM126	VEM126	unknown	Israel	unknown	O. aries	Petrous		1.11
VEM127	VEM127	unknown	Israel	unknown	O. aries	Petrous		0.12
VEM129	Dub5	Dublin	Ireland	Viking	O. aries	Petrous		42.63
VEM130	Dub3	Dublin	Ireland	Viking	O. aries	Petrous		55.73
VEM131	Dub4	Dublin	Ireland	Viking	O. aries	Petrous		58.43
VEM132	Dur4	Durrington Walls	UK	Neolithic	B. taurus	Teeth		0.27
VEM133	VEM133	Durrington Walls	UK	Neolithic	B. taurus	Teeth		0.12
VEM134	VEM134	Durrington Walls	UK	Neolithic	B. taurus	Teeth		0.41
VEM136	VEM136	Durrington Walls	UK	Neolithic	B. taurus	Teeth		0.19
VEM137	VEM137	Durrington Walls	UK	Neolithic	B. taurus	Teeth		0.13
VEM139	VEM139	Ain Ghazal	Jordan	PPNC	B. taurus	Petrous		0.18
VEM140	VEM140	Ain Ghazal	Jordan	Neolithic	C. hircus	Petrous		3.57
VEM141	VEM141	Ain Ghazal	Jordan	Neolithic	C. hircus	Petrous		0.32
VEM142	VEM142	Ain Ghazal	Jordan	Neolithic	C. hircus	Petrous		1.85
VEM143	VEM143	Ain Ghazal	Jordan	Neolithic	O. aries	Petrous		0.27
VEM144	VEM144	Ain Ghazal	Jordan	Neolithic	O. aries	Petrous		0.32
VEM145	VEM145	Ain Ghazal	Jordan	Neolithic	ovid/caprid	Petrous		0.30

VEM146	Cla1	Cladh Hallan	UK	Bronze Age	B. taurus	Petrous	1054-826 BC	59.14
VEM147	Cla10	Cladh Hallan	UK	Bronze Age	O. aries	Petrous		44.53
VEM148	Cla3	Cladh Hallan	UK	Bronze Age	B. taurus	Petrous		71.12
VEM149	Cla2	Cladh Hallan	UK	Bronze Age	B. taurus	Petrous		53.01
VEM150	Cla11	Cladh Hallan	UK	Bronze Age	O. aries	Petrous		61.84
VEM151	Cla14	Cladh Hallan	UK	Iron Age	O. aries	Petrous		56.52
VEM152	VEM152	Ain Ghazal	Jordan	mid PPNC	ovid/caprid	Petrous		0.06
VEM153	VEM153	Ain Ghazal	Jordan	Neolithic	B. taurus	Petrous		0.14
VEM154	VEM154	Ain Ghazal	Jordan	Neolithic	ovid/caprid	Petrous		0.08
VEM155	VEM155	Ain Ghazal	Jordan	Neolithic	ovid/caprid	Petrous		0.11
VEM157	VEM157	Ain Ghazal	Jordan	Neolithic	C. hircus	Petrous		2.66
VEM158	VEM158	Ain Ghazal	Jordan	Neolithic	C. hircus	Petrous		0.19
VEM159a	VEM159a	Ain Ghazal	Jordan	Neolithic	C. hircus	Petrous		0.93
VEM159b	VEM159b	Cladh Hallan	UK	Late Bronze Age	Cervus elaphus	Petrous		9.70
VEM160	Cla5	Cladh Hallan	UK	Late Bronze Age	B. taurus	Petrous		55.00
VEM161	Cla4	Cladh Hallan	UK	Late Bronze Age	B. taurus	Petrous		37.80
VEM162	Dur1	Durrington Walls	UK	Neolithic	B. taurus	Petrous	2410-2199 BC	29.90
VEM163	Dur2	Durrington Walls	UK	Neolithic	B. taurus	Petrous	2695-2457 BC	50.70
VEM164	Dur3	Durrington Walls	UK	Neolithic	B. taurus	Petrous		20.50
VEM165	Da6	Danebury	UK	Iron Age	B. taurus	Petrous	203-43BC	51.70
VEM166	Da1	Danebury	UK	Iron Age	B. taurus	Petrous		54.80
VEM167	Da2	Danebury	UK	Iron Age	B. taurus	Petrous		8.21
VEM168	Cla6	Cladh Hallan	UK	Early Iron Age	B. taurus	Petrous		52.62
VEM169	Cla7	Cladh Hallan	UK	Late Bronze Age	B. taurus	Petrous		51.01
VEM170	Fisk1	Fiskavik	UK	Iron Age	B. taurus	Petrous		56.10
VEM172	Fisk2	Fiskavik	UK	Iron Age	B. taurus	Petrous		48.97
VEM173	Bor1	Bornais	UK	Iron Age	B. taurus	Petrous		41.64
VEM174	Cla12	Cladh Hallan	UK	Late Bronze Age	O. aries	Petrous		21.33
VEM175	Cla13	Cladh Hallan	UK	Late Bronze Age	O. aries	Petrous		22.65
VEM176	Bis1	Bishop Cannings	UK	Neolithic	B. taurus	Petrous	3529-3362BC	57.50
VEM177	Bis2	Bishop Cannings	UK	Neolithic	B. taurus	Petrous	3638-3512BC	51.90
VEM178	Cla8	Cladh Hallan	UK	Bronze Age	B. taurus	Petrous	1060-899BC	58.16
VEM179	Cla9	Cladh Hallan	UK	Bronze Age	B. taurus	Petrous	1118-924BC	54.26
VEM180	Da3	Danebury	UK	Iron Age	B. taurus	Petrous	404-354BC	48.20
VEM181	Da4	Danebury	UK	Iron Age	B. taurus	Petrous		21.40
VEM182	Da5	Danebury	UK	Iron Age	B. taurus	Petrous	511-357BC	56.20
VEM183	VEM183	Durrington Walls	UK	Neolithic	Sus sp	Petrous		Fastqscreen 60%
VEM184	VEM184	Durrington Walls	UK	Neolithic	Sus sp	Petrous		Fastqscreen 70%
VEM185	VEM185	Durrington Walls	UK	Neolithic	Sus sp	Petrous		Fastqscreen 65%
VEM186	Da7	Danebury	UK	Iron Age	O. aries	Petrous		37.62

VEM187	VEM187	Danebury	UK	Iron Age	O. aries	Petrous		9.48
VEM188	Da8	Danebury	UK	Iron Age	O. aries	Petrous		21.95
VEM189	VEM189	Cambridge	UK	unknown	Bos sp	Petrous		0.15
VEM190	VEM190	Whitehawk Enclosure	UK	Neolithic	B. taurus	teeth		0.03
VEM191	VEM191	Whitehawk Enclosure	UK	Neolithic	B. taurus	teeth		0.05
VEM192	VEM192	Whitehawk Enclosure	UK	Neolithic	B. taurus	teeth		0.03
VEM193	Pot1	Potterne	UK	Bronze Age	B. taurus	Petrous		35.95
VEM194	Pot2	Potterne	UK	Bronze Age	B. taurus	Petrous		30.67
VEM195	Pot3	Potterne	UK	Bronze Age	B. taurus	Petrous		12.33
VEM196	Pot4	Potterne	UK	Bronze Age	B. taurus	Petrous	769-472BC	48.01
VEM197	Pot5	Potterne	UK	Bronze Age	B. taurus	Petrous		14.37
VEM198	Pot6	Potterne	UK	Bronze Age	B. taurus	Petrous		10.32
VEM199	VEM199	Potterne	UK	Bronze Age	C. hircus	Petrous		Fastqscreen 88%
VEM200	VEM200	Potterne	UK	Bronze Age	O. aries	Petrous		1.13
VEM201	Pot7	Potterne	UK	Bronze Age	O. aries	Petrous		35.83
VEM202	Ness1	Ness of Brodgar	UK	Neolithic	B. taurus	Petrous		37.86
VEM203	Ness2	Ness of Brodgar	UK	Neolithic	B. taurus	Petrous		45.13
VEM204	VEM204	Ness of Brodgar	UK	Neolithic	Cervus elaphus	Petrous		16.19
VEM205	Ness3	Ness of Brodgar	UK	Neolithic	B. taurus	Petrous		24.46
VEM206	Ness4	Ness of Brodgar	UK	Neolithic	B. taurus	Petrous		40.84
VEM207	Ness6	Ness of Brodgar	UK	Neolithic	O. aries	Petrous		47.11
VEM208	Ness7	Ness of Brodgar	UK	Neolithic	O. aries	Petrous		43.55
VEM209	Bor2	Bornais	UK	Iron Age	O. aries	Petrous		55.05
VEM210	Sil1	Silgenach	UK	Bronze Age	O. aries	Petrous		50.49
VEM211	VEM211	Bradley Fen	UK	Palaeolithic	B. primigenius	Petrous		0.07
VEM212	VEM212	Norton Bottoms	UK	Palaeolithic	B. primigenius	Petrous		0.25
VEM213	VEM213	Gully Cave, Ebbor gorge	UK	Holocene	B. primigenius	R. astragalus		0.24
VEM214	Ness5	Ness of Brodgar	UK	Neolithic	B. taurus	Petrous	2588-2464BC	39.20
VEM215	VEM215	Roughan Hill	Ireland	Bronze Age	B. taurus	Petrous		0.02
VEM216	Rou1	Roughan Hill	Ireland	Bronze Age	B. taurus	Petrous		42.14
VEM217	Rou2	Roughan Hill	Ireland	Bronze Age	B. taurus	Petrous		45.36
VEM218	VEM218	Roughan Hill	Ireland	Bronze Age	B. taurus	Petrous		0.03
VEM219	Rou3	Roughan Hill	Ireland	Bronze Age	B. taurus	Petrous		23.34
VEM220	VEM220	Roughan Hill	Ireland	Bronze Age	O. aries	Petrous		0.02
VEM221	VEM221	Roughan Hill	Ireland	Bronze Age	O. aries	Petrous		1.61
VEM222	VEM222	Roughan Hill	Ireland	Bronze Age	Sus sp	Petrous		Fastqscreen 12%
VEM223	Gen1	Gent	Belgium	Medieval -12-13th	B. taurus	Petrous		47.06

				Century				
VEM224	VEM224	Kilpheder	UK	Medieval/Norse 10-14th C	O. aries	Petrous		4.43
VEM225	VEM225	Kilpheder	UK	Medieval/Norse	O. aries	Petrous		11.24
VEM226	Kil1	Kilpheder	UK	Medieval/Norse	B. taurus	Petrous		54.50
VEM227	Kil2	Kilpheder	UK	Medieval/Norse	O. aries	Petrous		65.64
VEM228	Ork1	Snusgar	UK	Medieval/Norse - 11th century	O. aries	Petrous		62.66
VEM229	Ork2	Snusgar	UK	Medieval/Norse - 12th century	O. aries	Petrous		33.27
VEM230	Sir1	Snusgar	UK	Medieval/Norse - 12th century	B. taurus	Petrous	1039-1215AD	60.49
VEM231	Par1	Parknabinia	Ireland	Neolithic	B. taurus	Petrous		48.34
VEM232	Par2	Parknabinia	Ireland	Neolithic	B. taurus	Petrous		60.74
VEM233	Par3	Parknabinia	Ireland	Neolithic	B. taurus	Petrous		57.32
VEM234	Par4	Parknabinia	Ireland	Neolithic	O. aries	Petrous		1.71
VEM235	Bal1	Ballybane	Ireland	Medieval	B. taurus	Petrous		57.81
VEM236	Rou4	Roughan Hill	Ireland	Bronze Age	B. taurus	Petrous		17.98
VEM237	VEM237	Roughan Hill	Ireland	Bronze Age	B. taurus	Petrous		0.13
VEM238	Bal2	Ballybane	Ireland	Medieval	B. taurus	Petrous		55.56

Table 3.2. *The results of molecular sexing for cattle and sheep from the Britain and Ireland.*

Species	Sample	Time Period	Site	Sex
Cattle	Bal1	Medieval	Ballybane	Female
	Bal2	Medieval	Ballybane	Female
	Bis1	Neolithic	Bishops Cannings	Male
	Bis2	Neolithic	Bishops Cannings	Male
	Bor1	Iron	Bornais	Male
	Cla1	Bronze	Cladh Hallan	Female
	Cla2	Bronze	Cladh Hallan	Male
	Cla3	Bronze	Cladh Hallan	Female
	Cla4	Bronze	Cladh Hallan	Female
	Cla5	Bronze	Cladh Hallan	Female
	Cla6	Iron	Cladh Hallan	Female
	Cla7	Bronze	Cladh Hallan	Female
	Cla8	Bronze	Cladh Hallan	Female
	Cla9	Bronze	Cladh Hallan	Female
	Da1	Iron	Danebury	Female
	Da2	Iron	Danebury	Male
	Da3	Iron	Danebury	Female
	Da4	Iron	Danebury	Female
	Da5	Iron	Danebury	Female
	Da6	Iron	Danebury	Female
	Dub1	Medieval	Dubin	Female
	Dub2	Medieval	Dubin	Male
	Dur1	Neolithic	Durrington Walls	Female
	Dur2	Neolithic	Durrington Walls	Female
	Dur3	Neolithic	Durrington Walls	Female
	Fis1	Iron	Fiskavaig	Female
	Fis2	Iron	Fiskavaig	Female
	HF1	Brozne	Haughey's Fort	Female
	Kil1	Medieval	Kilpheder	Female
	Ness2	Neolithic	Ness of Brodgar	Female

	Ness3	Neolithic	Ness of Brodgar	Female
	Ness4	Neolithic	Ness of Brodgar	Female
	Ness5	Neolithic	Ness of Brodgar	Male
	New1	Neolithic	Newgrange	Male
	New2	Iron	Newgrange	Male
	New3	Neolithic	Newgrange	Male
	New4	Neolithic	Newgrange	Female
	New5	Neolithic	Newgrange	Male
	Par1	Neolithic	Parknabinia	Female
	Par2	Neolithic	Parknabinia	Male
	Pot1	Bronze	Potterne	Female
	Pot2	Bronze	Potterne	Female
	Pot3	Bronze	Potterne	Male
	Pot4	Iron	Potterne	Female
	Pot5	Bronze	Potterne	Female
	Pot6	Bronze	Potterne	Male
	Rou1	Bronze	Roughan Hill	Female
	Rou2	Bronze	Roughan Hill	Female
	Rou3	Bronze	Roughan Hill	Female
	Rou4	Bronze	Roughan Hill	Female
	Sir1	Medieval	Snusgar	Female
	Yor1	Roman	York	Female
	Yor10	Medieval	York	Female
	Yor11	Roman	York	Female
	Yor12	Medieval	York	Male
	Yor2	Medieval	York	Male
	Yor3	Medieval	York	Male
	Yor4	Medieval	York	Female
	Yor5	Medieval	York	Male
	Yor6	Medieval	York	Female
	Yor7	Medieval	York	Male
	Yor8	Medieval	York	Male

	Yor9	Medieval	York	Female
Sheep	Bor2	Iron	Bornais	Female
	Cla10	Bronze	Cladh Hallan	Female
	Cla11	Bronze	Cladh Hallan	Male
	Cla12	Bronze	Cladh Hallan	Male
	Cla13	Bronze	Cladh Hallan	Female
	Cla14	Iron	Cladh Hallan	Female
	Da7	Iron	Danebury	Female
	Da8	Iron	Danebury	Female
	Dub3	Medieval	Dublin	Female
	Dub4	Medieval	Dublin	Male
	Dub5	Medieval	Dublin	Male
	Kil2	Norse	Kilpheder	Female
	Ness6	Neolithic	Ness Brodgar	Male
	Ness7	Neolithic	Ness Brodgar	Male
	Ork1	Norse	Orkney	Male
	Ork2	Norse	Orkney	Male
	Par4	Neolithic	Parknibina	Female
	Pot7	Bronze/Iron	Potterne	Female
	Sil1	Bronze	Silgenach	Female
	VEM187	Iron	Danebury	Female
	VEM200	Bronze	Potterne	Female
	VEM221	Roughan Hill	Roughan Hill	Female
	VEM224	Norse	Kilpheder	Female
	VEM225	Norse	Kilpheder	Female
	Yor13	Medieval	York	Male

3.6. References

Allentoft, M.E. 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), pp.4724–4733.

Brace, S. et al., 2012. Population history of the Hispaniolan hutia Plagiodontia aedium (Rodentia: Capromyidae): testing the model of ancient differentiation on a geotectonically complex Caribbean island. *Molecular ecology*, 21(9), pp.2239–2253.

Campos, P.F. et al., 2012. DNA in ancient bone - where is it located and how should we extract it? *Annals of anatomy = Anatomischer Anzeiger: official organ of the Anatomische Gesellschaft*, 194(1), pp.7–16.

Collins, M.J. et al., 2002. The survival of organic matter in bone: a review. Available at: http://onlinelibrary.wiley.com/doi/10.1111/1475-4754.t01-1-00071/full.

Copley, M.S. et al., 2005. Dairying in antiquity. I. Evidence from absorbed lipid residues dating to the British Iron Age. *Journal of archaeological science*, 32(4), pp.485–503.

Dabney, J. et al., 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America*, 110(39), pp.15758–15763.

Damgaard, P.B. et al., 2015. Improving access to endogenous DNA in ancient bones and teeth. *Scientific reports*, 5, p.11184.

Darzynkiewicz, Z. et al., 1997. Cytometry in cell necrobiology: analysis of apoptosis and accidental cell death (necrosis). *Cytometry*, 27(1), pp.1–20.

Decker, J.E. et al., 2014. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS genetics*, 10(3), p.e1004254.

Frantz, L.A.F. et al., 2016. Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science*, 352(6290), pp.1228–1231.

Gamba, C. et al., 2015. Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Molecular ecology resources*. Available at: http://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12470/full.

Gamba, C. et al., 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nature communications*, 5, p.5257.

Hofreiter, M. et al., 2001. Ancient DNA. *Nature reviews. Genetics*, 2(5), pp.353–359.

Jans, M.M.E. et al., 2004. Characterisation of microbial attack on archaeological bone. *Journal of archaeological science*, 31(1), pp.87–95.

Kijas, J.W. et al., 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PloS one*, 4(3), p.e4668.

Lam, Y.M., Chen, X. & Pearson, O.M., 1999. Intertaxonomic Variability in Patterns of Bone Density and the Differential Representation of Bovid, Cervid, and Equid Elements in the Archaeological Record. *American antiquity*, 64(2), pp.343–362.

Lenstra, J.A. et al., 2014. Meta-Analysis of Mitochondrial DNA Reveals Several Population Bottlenecks during Worldwide Migrations of Cattle. *Diversity*, 6(1), pp.178–187.

Lindahl, T., 1993. Instability and decay of the primary structure of DNA. *Nature*, 362(6422), pp.709–715.

McGrory, S. et al., 2012. A novel method for integrated age and sex determination from archaeological cattle mandibles. *Journal of archaeological science*, 39(10), pp.3324–3330.

Meyer, M. et al., 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science*, 338(6104), pp.222–226.

Orlando, L. et al., 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, 499(7456), pp.74–78.

Orlando, L. et al., 2011. True single-molecule DNA sequencing of a pleistocene horse bone. *Genome research*, 21(10), pp.1705–1719.

Ottoni, C. et al., 2009. Preservation of ancient DNA in thermally damaged archaeological bone. *Die Naturwissenschaften*, 96(2), pp.267–278.

Pääbo, S. et al., 2004. Genetic analyses from ancient DNA. *Annual review of genetics*, 38, pp.645–679.

Payne, S., 1973. Kill-off Patterns in Sheep and Goats: the Mandibles from Aşvan Kale. *Anatolian Studies*, 23, pp.281–303.

Smith, C.I. et al., 2001. Neanderthal DNA. Not just old but old and cold? *Nature*, 410(6830), pp.771–772.

Smith, C.I. et al., 2003. The thermal history of human fossils and the likelihood of successful DNA amplification. *Journal of human evolution*, 45(3), pp.203–217.

Speller, C.F. et al., 2013. Ancient mtDNA Analysis of Early 16th Century Caribbean Cattle Provides Insight into Founding Populations of New World Creole Cattle Breeds M. T. P. Gilbert, ed. *PloS one*, 8(7), p.e69584.

Svensson, E.M., Götherström, A. & Vretemark, M., 2008. A DNA test for sex identification in cattle confirms osteometric results. *Journal of archaeological science*, 35(4), pp.942–946.

Chapter 4. Mitogenome analysis of ancient *Bos* samples

4.1. Introduction

Archaeozoological evidence indicates the domestication of taurine cattle occurred in the Upper and Middle Euphrates Valley of the Near East approximately 9,000 BC from the wild progenitor *Bos primigenius* (Helmer et al. 2005; Arbuckle 2014). The partial mitochondrial DNA study of Troy et al. (2001) showed the highest mtDNA haplogroup diversity of modern taurine cattle is in Southwest Asia, indicating that this area was likely the centre of domestication and therefore supporting archaeological evidence. Utilising modern data to understand domestication has limitations, due to past migrations, selection events, population turnover and in the case of domestic animals, breed formation. However the pattern described by Troy and colleagues (2001) has since been confirmed by partial mitochondrial DNA analysis of ancient samples (Bollongino et al. 2006; Scheu et al. 2015). Independent domestication events across Europe are, in general, not supported (Edwards et al. 2007; Scheu et al. 2008). However the presence of divergent mtDNA haplotypes, mainly in Italian animals, has lead some to argue for seperate domestication processes (Beja-Pereira et al. 2006; Achilli et al. 2008; Bonfiglio et al. 2010).

The geographic distribution of cattle mtDNA haplogroups has long been a research focus, utilising both partial and full modern and ancient mtDNA. The haplogroup definitions for cattle were initially assigned based on variation within a 240bp region of the D-loop of the mitochondria; the majority of these assignments have since been scaled to the mitogenome level (Loftus et al. 1994; Bailey et al. 1996; Bradley et al. 1996; Mannen et al. 1998; Troy et al. 2001; Magee et al. 2002; Edwards et al. 2007; Achilli et al. 2008; Achilli et al. 2009; Bonfiglio et al. 2012; Olivieri et al. 2015). The majority of taurine cattle belong the macro haplogroup T. Most modern and ancient domestic European cattle belong the large haplogroup T3, African cattle belong almost exclusively to T1, while T4 has only been identified in Japanese cattle and T5 in Italian and Croatian animals (Achilli et al. 2008; Ivankovic et al. 2014). The greatest diversity is seen in the Near East with haplogroups T,T2,T3 and Q present (Mannen et al. 1998; Troy et al. 2001; Olivieri et al. 2015). The haplogroup R has only been published in modern Italian cattle (Bonfiglio et al. 2010). Three haplogroups identified primarily in *Bos primigenius* are published, P in Europe, E in one German individual and C from a Chinese Mesolithic auroch (Bailey et al. 1996; Troy et al. 2001; Edwards et al. 2007; Stock et al. 2009; Zhang et al. 2013). While very rare in moderns, P has been found to have survived in a modern Korean beef animal (DQ124389.1 unpublished).

Evidence from the modelling of ancient mtDNA and modern mtDNA has suggested a small effective population of 80 female aurochsen were domesticated (Bollongino et al. 2012) in the Fertile Crescent. A migration of cattle then followed; from Southern Anatolia to Western Anatolia and the Aegean post 7,000BC. Introduction into Europe then followed approximately 6,400 BC, eventually reaching the North of Europe by 4,100 BC (Scheu et al. 2015). As these domesticated cattle moved across the European continent, contact with wild populations would have increased. Contradicting results from Y chromosome studies have left inconclusive understanding of male mediated introgression. Conflicting results for introgression on the maternal line have also been published (Beja-Pereira et al. 2006; Edwards et al. 2007; Achilli et al. 2008; Scheu et al. 2015). However, the sequencing and publication of the first whole genome *Bos primigenius* from a Mesolithic context in Britain, in conjunction with genome-wide SNP data of modern breeds, indicated European auroch introgression did indeed occur (Park et al. 2015; Upadhyay et al. 2016).

Recent developments in sequencing technology has enabled the sequencing of whole nuclear genomes of different species, including woolly mammoth, domestic dog, horse and hominid genomes (Poinar et al. 2006; Rasmussen et al. 2010; Meyer et al. 2012; Gamba et al. 2014; Schubert et al. 2014; Frantz et al. 2016). However prior to this technology the traditional focus of ancient DNA, as well as many modern studies, was the maternal mitochondrial locus. This matrilineal locus has enabled the development of an understanding of the evolutionary histories of many different species utilising both modern and ancient samples, including cattle (Loftus et al. 1994; Bradley et al. 1996; Troy et al. 2001), pigs (Larson et al. 2005), Brown bears (Barnes et al. 2002), Arctic fox (Dalén et al. 2005; Dalén et al. 2007) and bison (Shapiro et al. 2004) to name a few.

The nature of aDNA; low concentrations of fragmented, easily contaminated damaged DNA (Hansen et al. 2001; Hofreiter et al. 2001), is a hurdle one must overcome to extract and sequence ancient genomes, nuclear or mitochondrial. The endogenous percentages of DNA in many ancient samples is between 0.1 to 5%, however there are exceptions to this rule as discussed in Chapter 3. The high copy number of mitochondrial DNA in cells, eg. 3,000-6,000 per diploid nuclear genome in muscle cells (Miller et al. 2003), aids the probability of recovering mtDNA from archaeological and paleontological specimens, especially when endogenous percentages are low.

Earlier mitochondrial studies focused on targeting regions of the mtDNA with PCR amplification, such as the hypervariable region or cytochrome B. Only a handful of whole aDNA mitochondria were sequenced using PCR amplification and sanger sequencing, for example the Moa (Cooper et al. 2001) and Woolly mammoth (Rogaev et al. 2006).

However with improvements in technology, such as the introduction of Next Generation Sequencing (NGS), the sequencing of mitogenomes became more easily accessible to the aDNA community. Whole mitogenomes of many different species have been published, including; Woolly mammoth, cattle, horse, rhino, Neanderthal, Denisovan, and anatomically modern humans (Gilbert et al. 2007; Gilbert et al. 2008; Green et al. 2008; Briggs et al. 2009; Willerslev et al. 2009; Edwards et al. 2010; Reich et al. 2010; Vilstrup et al. 2013).

Focusing on mitochondrial DNA and the use of NGS technologies does not fully solve the problems posed by low endogenous DNA samples. To sequence the mitogenome of such a sample via the widely used shotgun method is an expensive exercise. One solution is to use a target enrichment strategy to capture the mitochondria (Briggs et al. 2009; Krause et al. 2010; Reich et al. 2010), reviewed by (Paijmans et al. 2013). These hybridisation capture methods enrich for a predetermined portion of the genome, often the whole mitochondria or nuclear exons, using homologous single-stranded DNA or RNA baits. These baits are exposed to NGS libraries, with hybridisation either occurring in solution with biotinylated baits and target DNA immobilised by streptavidin-covered beads or with the baits immobilised on a solid phase (Briggs et al. 2009; Burbano et al. 2010; Reich et al. 2010). Sequences not hybridised to baits, including non targeted regions of the genome and exogenous DNA, are washed away and the resulting library is enriched for the targeted region. This method enables the capture of genetic information from samples with low endogenous content, for which whole genome sequencing would not be viable at this present time. This method has also been used to target SNPs in the nuclear genome of ancient human populations (Rohland & Reich 2012; Carpenter et al. 2013; Haak et al. 2015).

The aim of this chapter is to co-analyse 127 ancient *Bos* mitogenomes, achieved through a mixture of target capture and shotgun sequencing, with published modern mitogenomes, to assess the *Bos* phylogeny and geographical distribution of ancient *Bos taurus* (domestic cattle) and *Bos primigenius* (auroch) haplogroups. A number of these samples are then utilised to provide molecular dating for the divergences of the major haplogroups of cattle using a Bayesian phylogenetic framework. Finally, the preliminary analysis of a 200,000 year old *Bos primigenius* is presented to help to illuminate the evolutionary history of the *Bos* lineage.

4.2. Materials

4.2.1. Ancient

The samples in this chapter were screened as in 2.2.1 and results presented in Chapter 3. This chapter is also supplemented with samples prepared by Marta Verdugo and Amelie Scheu. Ancient samples range temporally from the Paleolithic to Medieval and geographically from Egypt to Ireland (Table 4.1). Additionally two previously published ancient mitogenomes the divergent Chinese C sequence (Zhang et al 2013) and the European auroch CPC (Edwards et al. 2010) were downloaded from NCBI.

4.2.2. Modern

NCBI Accession numbers for the 183 modern cattle used in this analysis are given in Appendix Table 4. The majority of these samples are part of the dataset used for haplotype determination inMitoToolPy (https://github.com/kizcas/MitoToolPy), supplemented with other published mitochondria (Appendix Table 5).

4.3. Methods

Mitochondria were sequenced following two different pipelines:

1. Target Capture
2. Shotgun Whole Genome Sequencing

4.3.1. Capture Array

In total four mitochondrial target captures were performed using two different capture array methods, see Appendix Table 6 for details of which capture was performed on what samples.

4.3.1.1. Agilent Capture - SureSelect Target Enrichment protocol

The DNA extraction and library building protocol (non USER libraries) were followed as described in 2.2.1. Between one and two indexing PCRs of 12 cycles per library were performed prior to pooling and enrichment using a previously designed custom in solution RNA capture (Agilent) following the SureSelect Target Enrichment protocol. Post enrichment a final PCR amplification step of 18 cycles was performed prior to sequencing on the Illumina MiSeq^{TM} platform at TrinSeq St James's Hospital Dublin.

4.3.1.2. MYcroarray® – Mybaits protocol

The DNA extraction and library building protocol (non USER libraries) were followed as described in 2.2.1. Between one and eight indexing PCRs of 12 to 16 cycles per library were performed prior to pooling and enrichment using a previously designed custom in solution RNA capture (Mycroarray, 5692 Plymouth Road, Ann Arbor, MI 48105) following the manufacturer's instructions. Post enrichment a final PCR amplification step of 15 cycles was performed prior to sequencing on the Illumina MiSeq^{TM} platform at TrinSeq St James's Hospital Dublin.

4.3.2. Read Processing

4.3.2.1. Read Processing for Capture Array

Sequencing reads were aligned as previously discussed (2.2.1.4) to the *Bos taurus* mitochondrial reference genome (V00654.1) (Anderson et al. 1982). Reads were filtered for PCR duplicates, mapping quality of 25, unique (XT flag) and removal of suboptimal hits (X1).

4.3.2.2. Shotgun Whole Genome Sequencing

Sequencing reads from USER and non USER libraries were processed as previously discussed (2.2.1-2.2.2). Using SAMtools (Li et al. 2009) mitochondrial reads were removed from BAM files of whole genome alignment previously filtered for PCR duplicates. Reads were then filtered for unique (XT flag) and removal of suboptimal hits (X1).

4.3.3. Coverage

Coverage of the mitochondria was calculated using Qualimap (V2.1.3) running default options (Okonechnikov et al. 2015).

4.3.4. Fasta formation

4.3.4.1. General

Pileup files were produced by GATK(v3.3-0-g37228af) (McKenna et al. 2010) using the mpileup option with default settings. The resulting files were filtered using a custom script for a minimum base quality of 20, minimum depth of coverage of 2, and calling the consensus allele. Where necessary ties were symbolised by ambiguity characters and later replaced by the character N. All generated fasta files were visually inspected against visualized bam files, and where necessary altered using SEAVIEW (Gouy et al. 2010).

4.3.5. Multiple Sequence Alignment

Multiple sequence alignment was performed using the MUSCLE (Edgar 2004) algorithm in SEAVIEW (Gouy et al. 2010).

4.3.6. Tree

4.3.6.1. Maximum Likelihood

Multiple Sequence Alignment files were uploaded onto the PhyML 3.0 online execution platform (Guindon et al. 2010) and maximum likelihood trees drawn using the following options:

- Automatic substitution model selection using the AIC (Akaike Information Criterion) (Lefort et al. 2017). The model selected was the GTR substitution model with a gamma-distributed rate of variation among sites (+G) and invariant sites (+I) (ie. GTR +G+I model).
- Starting tree with BioNJ, using NNI for tree topology search.
- Branch support using aBayes method (Anisimova et al. 2011), a bayseian-like transformation of an approximate likelihood ratio test [aLRT].

Trees were then visualized and altered using FigTree (Rambaut 2007).

4.3.6.2. Bayesian Analysis - BEAST

Phylogenetic analysis conducted under a Bayesian framework was performed using the program BEAST (version 2.4.7). BEAST allows for the use of temporally different samples, such as timestamped ancient and modern samples, to estimate the mutation rate simultaneously with population history parameters, such as branch time depth (Drummond & Rambaut 2007).

A multiple alignment of 183 modern samples, 26 radiocarbon dated ancients including 2 previously published (Edwards et al. 2010 & Zhang et al. 2013) was performed in SEAVIEW. Additionally a second alignment including non radiocarbon dated samples was performed. Due to different mutation rates of different regions of the mitochondrial genome, the alignment was partitioned into regions using the partitions denoted for the *Bos* reference mitochondrial sequence (V00654.1)(Anderson et al. 1982). Partitionfinder was run on default settings to check if partitions could be merged, as well as to decide the substitution model per partition (Lanfear et al. 2012). Five partitions were used; D-loop, 1st and 2nd coding positions within the coding regions, 3rd coding position, RNA genes, and tRNA + remainder of positions. Models for each of the partitions were as follows; D-loop = GTR+I+G, C1C2 = TRN+I+G, C3 =TRN+I, RNA = HKY+I+G, and tRNA = HKY+G. The demographic Bayesian coalescent skyline model was implemented with the tree model partitions linked and a strict molecular clock with a lognormal prior for the substitution rates of each partition. Substitution rates were as follows; D-loop mean -14.955, stdev 1.4 equivalent to a median 3.20e-07 [95% HPD 320e-08 - 3.20e-06] substitutions per site per year (Finlay et al. 2007). As executed by (Massilani et al. 2016) estimates of substitution rates of the human mitogenome were used to set the priors: 1st & 2nd coding positions, mean=-18.5 stdev=2.0, corresponding to a median of 9.24e-9 [95%HPD 3.44e-10 - 2.48e-7], 3rd coding position mean=-17.7 stdev=2.0, corresponding to a median of 2.06e-8 [95%HPD 7.66e-10 - 5.52e-7], RNA and tRNA+remainder partitions, mean=-18.65 stdev=2.0, corresponding to a median of 7.95e-8 [2.96e-10 - 2.13e-7].

The Markov chain Monte Carlo analysis was implemented in BEAST, running two independent chains of 200 million iterations each, samples drawn every 20,000 iterations and 10% burnin discarded. Log files were visually inspected using TRACER (v.1.6) (Rambaut et al. 2014) to ensure MCMC convergence, the two runs combined using LogCombiner (v.2.4.7) and the maximum clade credibility tree with the median height of the nodes was calculated using TreeAnnotator (v. 2.4.7) (Drummond & Rambaut 2007). Visualization of tree was achieved through FigTree (v. 1.4.3) (Rambaut 2007).

4.3.7. VEM212 - a 200,000 yrBP *Bos primigenius*

Due to the age of this sample the number of unique reads obtained, from both capture and shotgun sequencing, was low. Preliminary analysis suggested that this sample is an outgroup to other *B. primigenius*. Therefore rather than align the reads to the *B. taurus* reference genome, reads from both the capture and shotgun sequencing were aligned to the published Cpc *B. primigenius* mtDNA genome (Edwards et al. 2010) in a bid to improve the alignment.

Reads were filtered as explained above and GATK (McKenna et al. 2010) *mpileup* used to produce pileup files. The custom made script was utilised to create a fasta file with a minimum of base quality of 20, however minimum coverage was set at one and the consensus further altered with the use of SEAVIEW (Gouy et al. 2010). Variance away from CPC was generally accepted as follows:

- Transversions at a coverage of one.
- Transitions in the middle of reads at a coverage of 2-3 depending on the placement along all of the reads.
- Transitions at the ends of reads at a coverage of 3, however variance was not taken if all of the varying bases were always the first/last base of the reads.

Multiple alignment of the consensus sequence and a dataset of ancient samples and a reduced number of moderns, was performed using SEAVIEW (Gouy et al. 2010). A Neighbour joining tree was drawn using the BioNJ option in SEAVIEW and altered using Figtree (Rambaut 2007).

Table 4.1. *Ancient samples analysed in this chapter. The individual responsible for the laboratory work for these samples is denoted in the Lab Work column - AS Amelie Scheu, MV - Marta Verdugo, VEM - Victoria Mullin. It is noted if these samples were USER treated and if the sample was shotgun sequenced or target enrichment in the form of mtDNA capture was performed.*

Name	Cultural Period	Radiocarbon Date calBC (2 Sigma)	Site	Country	Lab Work	USER	Shotgun/Capture
Gen1	Medieval		Gent	Belgium	VEM	Yes	Shotgun
Dzh1	Neolithic		Dzulhunitsa	Bulgaria	AS	Yes	Shotgun
Dzh2	Neolithic		Dzulhunitsa	Bulgaria	AS	Yes	Shotgun
VEM030	Unknown		Yabalkovo	Bulgaria	VEM	No	Capture
VEM031	Unknown		Yabalkovo	Bulgaria	VEM	No	Capture
VEM090	Bronze Age		Thoth Hill	Egypt	VEM	No	Capture
VEM104	Bronze Age		Deir el Bahri	Egypt	VEM	No	Capture
Da1	Iron Age		Danebury	England	VEM	Yes	Shotgun
Da2	Iron Age		Danebury	England	VEM	Yes	Shotgun
Da3	Iron Age	404-354BC	Danebury	England	VEM	Yes	Shotgun
Da4	Iron Age		Danebury	England	VEM	Yes	Shotgun
Da5	Iron Age	511-357BC	Danebury	England	VEM	Yes	Shotgun
Da6	Iron Age	203-43BC	Danebury	England	VEM	Yes	Shotgun
Pot4	Iron Age	769-472BC	Potterne	England	VEM	Yes	Shotgun
Pot1	Late Bronze Age		Potterne	England	VEM	Yes	Shotgun
Pot2	Late Bronze Age		Potterne	England	VEM	Yes	Shotgun
Pot3	Late Bronze Age		Potterne	England	VEM	Yes	Shotgun
Pot5	Late Bronze Age		Potterne	England	VEM	Yes	Shotgun
Pot6	Late Bronze Age		Potterne	England	VEM	Yes	Shotgun
Yor10	Medieval		York	England	VEM	No	Shotgun
Yor12	Medieval		York	England	VEM	No	Shotgun
Yor2	Medieval		York	England	VEM	Yes	Shotgun
Yor3	Medieval		York	England	VEM	Yes	Shotgun
Yor4	Medieval		York	England	VEM	Yes	Shotgun
Yor5	Medieval		York	England	VEM	No	Shotgun
Yor6	Medieval		York	England	VEM	No	Shotgun
Yor7	Medieval		York	England	VEM	No	Shotgun
Yor8	Medieval		York	England	VEM	No	Shotgun
Yor9	Medieval		York	England	VEM	No	Shotgun
Bis1	Neolithic	3529-3362BC	Bishops Cannings	England	VEM	Yes	Shotgun
Bis2	Neolithic	3638-3512BC	Bishops Cannings	England	VEM	Yes	Shotgun
Dur1	Neolithic	2410-2199BC	Durrington Walls	England	VEM	Yes	Shotgun
Dur2	Neolithic	2695-2457BC	Durrington Walls	England	VEM	Yes	Shotgun
Dur3	Neolithic		Durrington Walls	England	VEM	Yes	Shotgun
VEM089	Neolithic		Windmill Hill	England	VEM	No	Capture
VEM132	Neolithic		Durrington Walls	England	VEM	No	Capture
VEM133	Neolithic		Durrington Walls	England	VEM	No	Capture
VEM134	Neolithic		Durrington Walls	England	VEM	No	Capture
VEM136	Neolithic		Durrington Walls	England	VEM	No	Capture
VEM137	Neolithic		Durrington Walls	England	VEM	No	Capture
VEM190	Neolithic		Whitehawk Enclosure	England	VEM	No	Capture
VEM191	Neolithic		Whitehawk Enclosure	England	VEM	No	Capture

VEM192	Neolithic		Whitehawk Enclosure	England	VEM	No	Capture
VEM212	Paleolithic		Norton Bottoms	England	VEM	No	Capture
Yor1	Roman	2BC-135AD	York	England	VEM	Yes	Shotgun
Yor11	Roman		York	England	VEM	No	Shotgun
VEM189	Unknown		Cambridge Fens	England	VEM	No	Capture
Dyr1	Medieval		Dyrotz	Germany	AS	Yes	Shotgun
Fal7	Medieval		Falkenwalde	Germany	AS	Yes	Shotgun
Bed3	Mesolithic	9805-9385BC	Bedburg-Königshoven	Germany	AS	Yes	Shotgun
Bed4	Mesolithic	10036-9647BC	Bedburg-Königshoven	Germany	AS	Yes	Shotgun
Tri1	Mesolithic	8261-7960BC	Trier-Brüderkrankenhaus	Germany	AS	Yes	Shotgun
Hxh1	Neolithic		Herxheim	Germany	AS	Yes	Shotgun
Hxh2	Neolithic		Herxheim	Germany	AS	Yes	Shotgun
Kir18	Neolithic	3798-3641BC	Kirschbaumhöhle	Germany	AS	Yes	Shotgun
Kir4	Neolithic	2816-2671BC	Kirschbaumhöhle	Germany	MV	Yes	Shotgun
VEM085	Neolithic		Ecsegfalva	Hungry	VEM	No	Capture
VEM086	Neolithic		Ecsegfalva	Hungry	VEM	No	Capture
VEM084	Chalcolithic /Bronze Age		Maral Tappeh	Iran	VEM	No	Capture
Rou1	Bronze Age		Roughan Hill	Ireland	VEM	Yes	Shotgun
Rou2	Bronze Age		Roughan Hill	Ireland	VEM	Yes	Shotgun
Rou3	Bronze Age		Roughan Hill	Ireland	VEM	Yes	Shotgun
Rou4	Bronze Age		Roughan Hill	Ireland	VEM	Yes	Shotgun
New2	Iron Age	321-425AD	Newgrange	Ireland	VEM	Yes	Shotgun
Bal1	Medieval		Ballybane	Ireland	VEM	Yes	Shotgun
Bal2	Medieval		Ballybane	Ireland	VEM	Yes	Shotgun
Dub1	Medieval		Dublin	Ireland	VEM	Yes	Shotgun
Dub2	Medieval		Dublin	Ireland	VEM	No	Capture
New1	Neolithic	2795-2580BC	Newgrange	Ireland	VEM	Yes	Shotgun
New3	Neolithic		Newgrange	Ireland	VEM	Yes	Shotgun
New4	Neolithic		Newgrange	Ireland	VEM	Yes	Shotgun
New5	Neolithic	3024-2894BC	Newgrange	Ireland	VEM	Yes	Shotgun
Par1	Neolithic		Parknabinia	Ireland	VEM	Yes	Shotgun
Par2	Neolithic		Parknabinia	Ireland	VEM	Yes	Shotgun
VEM097	Neolithic		Kilshane	Ireland	VEM	No	Capture
VEM098	Neolithic		Kilshane	Ireland	VEM	No	Capture
VEM099	Neolithic		Kilshane	Ireland	VEM	No	Capture
VEM139	Neolithic		Ain Ghazal	Jordan	VEM	No	Capture
VEM153	Neolithic		Ain Ghazal	Jordan	VEM	No	Capture
HF1	Bronze Age		Haughey's Fort	N. Ireland	VEM	No	Shotgun
Viz1	Bronze Age		Vianen-Zijdervel	Netherlands	MV	Yes	Shotgun
Viz2	Bronze Age		Vianen-Zijdervel	Netherlands	MV	Yes	Shotgun
Bun2	Iron Age		Bunnik	Netherlands	MV	Yes	Shotgun
Ste1	Iron Age/ Roman		Stevenshofjesp	Netherlands	MV	Yes	Shotgun
Bun1	Iron Age/ Roman		Bunnik	Netherlands	MV	Yes	Shotgun
Fir1	Medieval		Firdgum	Netherlands	MV	Yes	Shotgun
Hou1	Medieval		Houten-Loerik	Netherlands	MV	Yes	Shotgun
Hou2	Medieval		Houten-Hoogdij	Netherlands	MV	Yes	Shotgun
Sch1	Neolithic		Schipluiden	Netherlands	MV	Yes	Shotgun
Bri1	Roman		Britsum	Netherlands	MV	Yes	Shotgun
Dro1	Roman		Dronrijp-zuid	Netherlands	MV	Yes	Shotgun
Els1	Roman		Elst	Netherlands	MV	Yes	Shotgun

Otb1	Roman		Oosterbeintum	Netherlands	MV	Yes	Shotgun
Win1	Roman	417-561AD	Winsum	Netherlands	MV	Yes	Shotgun
Kie1	Neolithic		Kierzkowo	Poland	MV	Yes	Shotgun
Kie2	Neolithic		Kierzkowo	Poland	MV	Yes	Shotgun
Lud1	Neolithic		Ludwinowo	Poland	MV	Yes	Shotgun
Lud2	Neolithic		Ludwinowo	Poland	MV	Yes	Shotgun
Lud3	Neolithic		Ludwinowo	Poland	MV	Yes	Shotgun
Lud4	Neolithic		Ludwinowo	Poland	MV	Yes	Shotgun
Kalisz2	no info		Kalisz	Poland	MV	Yes	Shotgun
Kalisz3	no info		Kalisz	Poland	MV	Yes	Shotgun
Cla1	Bronze Age	1054-826BC	Cladh Hallan	Scotland	VEM	Yes	Shotgun
Cla2	Bronze Age		Cladh Hallan	Scotland	VEM	Yes	Shotgun
Cla3	Bronze Age		Cladh Hallan	Scotland	VEM	Yes	Shotgun
Cla4	Bronze Age		Cladh Hallan	Scotland	VEM	Yes	Shotgun
Cla5	Bronze Age		Cladh Hallan	Scotland	VEM	Yes	Shotgun
Cla7	Bronze Age		Cladh Hallan	Scotland	VEM	Yes	Shotgun
Cla8	Bronze Age	1060-899BC	Cladh Hallan	Scotland	VEM	Yes	Shotgun
Cla9	Bronze Age	1118-924BC	Cladh Hallan	Scotland	VEM	Yes	Shotgun
Bor1	Iron Age		Bornais	Scotland	VEM	Yes	Shotgun
Cla6	Iron Age		Cladh Hallan	Scotland	VEM	Yes	Shotgun
Fis1	Iron Age		Fiskavaig	Scotland	VEM	Yes	Shotgun
Fis2	Iron Age		Fiskavaig	Scotland	VEM	Yes	Shotgun
Ness2	Late Neolithic		Ness of Brodgar	Scotland	VEM	Yes	Shotgun
Ness3	Late Neolithic		Ness of Brodgar	Scotland	VEM	Yes	Shotgun
Ness4	Late Neolithic		Ness of Brodgar	Scotland	VEM	Yes	Shotgun
Ness5	Late Neolithic	2588-2464BC	Ness of Brodgar	Scotland	VEM	Yes	Shotgun
Kil1	Medieval - Norse		Kilpheder	Scotland	VEM	Yes	Shotgun
Sir1	Medieval - Norse	1039-1215AD	Snusgar	Scotland	VEM	Yes	Shotgun
Bel1	Neolithic		Belovode- Veliko	Serbia	MV	Yes	Shotgun
Bel2	Neolithic		Belovode- Veliko	Serbia	MV	Yes	Shotgun
Bub1	Neolithic		Bubanj	Serbia	MV	No	Shotgun
Plo1	Neolithic		Plocnik	Serbia	MV	Yes	Shotgun
Plo2	Neolithic		Plocnik	Serbia	MV	Yes	Shotgun
Plo3	Neolithic		Plocnik	Serbia	MV	Yes	Shotgun
Plo4	Neolithic		Plocnik	Serbia	MV	Yes	Shotgun
Plo5	Neolithic		Plocnik	Serbia	MV	Yes	Shotgun
Plo6	Neolithic		Plocnik	Serbia	MV	Yes	Shotgun
Plo7	Neolithic		Plocnik	Serbia	MV	Yes	Shotgun
Plo8	Neolithic		Plocnik	Serbia	MV	Yes	Shotgun
Stu1	Neolithic		Stubline	Serbia	MV	Yes	Shotgun
Ch22	Neolithic		Catal Hoyuk	Turkey	AS	Yes	Shotgun
Men1	Neolithic		Mentese	Turkey	MV	Yes	Shotgun
Men2	Neolithic	6048-5893BC	Mentese	Turkey	MV	Yes	Shotgun
Sub1	Neolithic	6221-6024BC	Suberde	Turkey	MV	Yes	Shotgun
Turk11	Neolithic		Mentese	Turkey	VEM	No	Capture
Turk12	Neolithic		Mentese	Turkey	VEM	No	Capture
Turk13	Neolithic	6078-5986	Mentese	Turkey	VEM	No	Capture
Turk6	Neolithic		Mentese	Turkey	VEM	No	Capture

4.4. Results

4.4.1. Overview

28 samples were captured for *Bos* mtDNA (Appendix Table 6 and Fig 4.1), however only 15 will be used in the phylogenetic analysis here due to incomplete coverage in the remaining samples (Table 4.2). A further 112 mitochondrial genomes were extracted from shotgun sequencing reads (Table 4.3). Coverage for the captured samples analysed in this chapter, range from 1.86 to 134.08X with an average 29.67X. Three of the captured samples, Turk11, VEM084 and VEM212, were supplemented by shotgun reads. The coverages of mtDNA produced via shotgun sequencing are higher than the target capture samples, with a range of 5.52X to 3532.02X, average 266.60X.

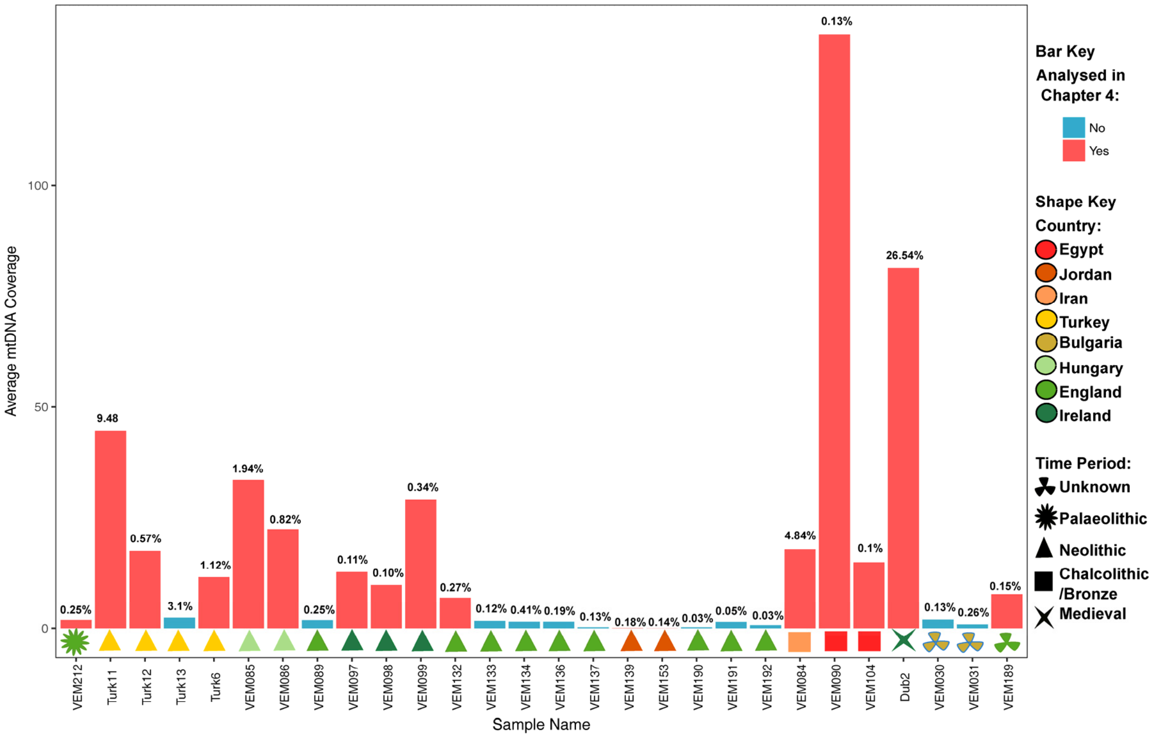


Fig 4.1. *Sequencing success for captured samples is demonstrated in this graph. Colour of bar indicates whether or not a sample is analysed in this chapter, average mtDNA coverage is on the x-axis with each bar labelled with % endogenous result from shotgun sequencing (Chapter 3). Time period and geography of the sample is denoted by the symbol underneath each bar.*

4.4.2. Preservation of Captured Samples

The majority of samples captured had low endogenous DNA percentages, below 1%, when screened via the shotgun method. Capture was attempted to enable sequencing of the mitochondria. Despite low percentages of endogenous DNA, it was still possible to capture and sequence mtDNA from some of these samples. Three samples that were successfully captured (VEM097,VEM098, VEM099), originate from a Neolithic site in

Kilshane, Ireland (Table 4.3). Attempts to radiocarbon date this site have failed due to poor collagen preservation (Finbar McCormick *per. comm.*). The two Egyptian Bronze Age samples, VEM090 and VEM104, that were successfully captured were pieces of tendon and skin that had been mummified and desiccated, with very low endogenous DNA content of 0.1% each. These two samples have a somewhat different damage pattern result to the damage pattern expected from ancient DNA pattern (Fig 4.2), with a lack of C>T and G>A substitutions in comparison to ancient bone samples.

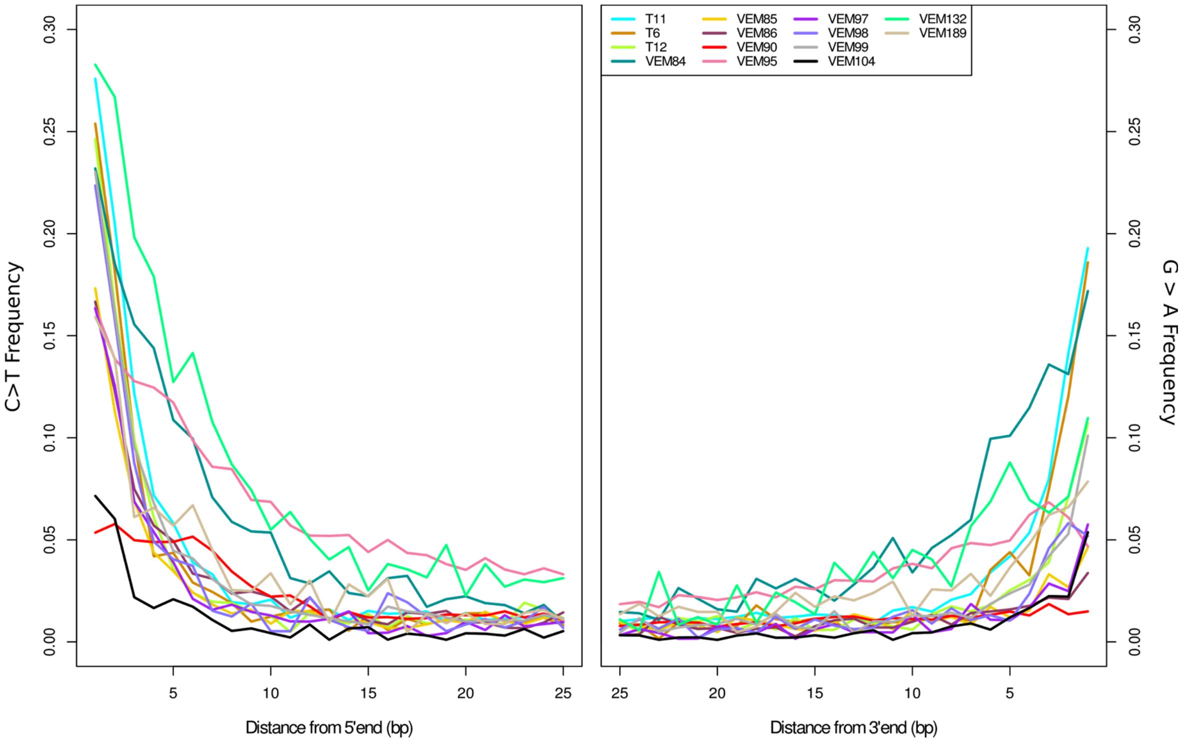


Fig 4.2. *Damage pattern results for the target capture samples anaylsed in this Chapter (bar VEM212). Note the different damage pattern of VEM090 (red) and VEM104 (black), especially the C>T frequency, in comparison to the other ancient samples. These two samples are the skin samples from Egyptian Bronze Age cattle.*

4.4.3. Overall Phylogenetic Structure of *Bos* mtDNA

The prior view of mtDNA phylogenetic structure of *Bos* (Fig 4.3) holds with both ancient and modern whole mtDNA data. The broomstick pattern of the unrooted tree contains deep internal branch lengths indicative of large divergence times between many of the haplogroups. The T haplogroups clusters to one end and *B. indicus* acts as the root for the tree (Fig 4.3a), a trifurcation in the maximum likelihood tree then splits the published haplogroups of R and C from the rest of the tree. These outgroup the European auroch haplogroup P, as well as Q and T haplogroups.

The trifurcation present in the maximum likelihood trees, is presented differently in the bayesian tree and BioNJ trees (Fig 4.5 & 4.7), with C acting as an outgroup to R and the ancient sample Kir4. However bootstrapping and posterior probabilities show low values, indicating low confidence in solving the phylogeny of this section of the tree.

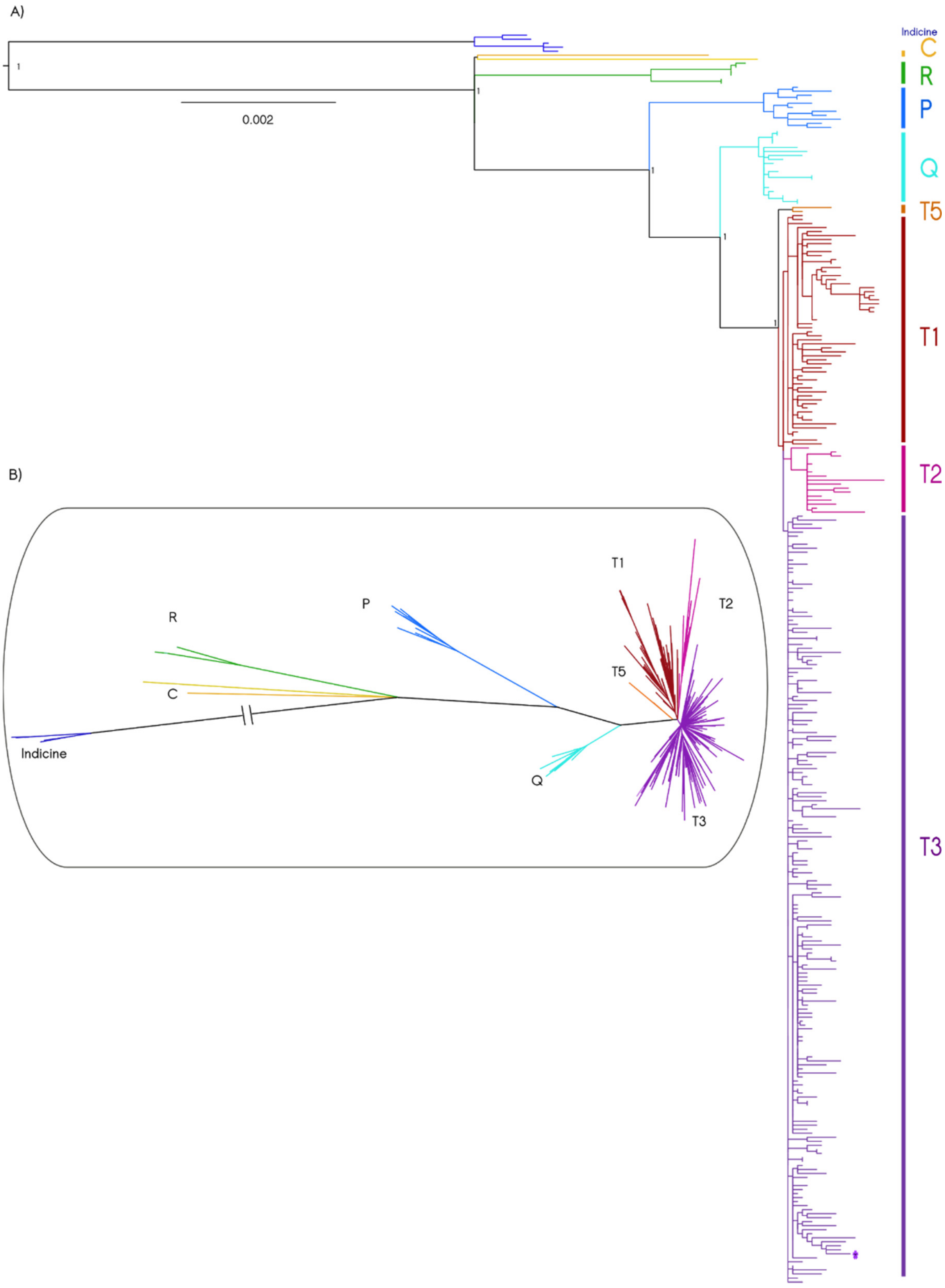


Fig 4.3. *Maximum likelihood tree of ancient and modern Bos species to demonstrate the overview of the Bos genus phylogeny. A) Tree rooted by Bos indicus. Published haplogroups are noted on the right hand side, * sample indicates the T4 haplogroup. Main branches with high support are indicated by 1. B) An unrooted tree displays the broomstick phylogeny of Bos, first identified by Troy et al 2001. The long branch lengths between the Indicine samples and the T haplogroups, indicative of highly divergent sequences, while the starburst of the T haplogroup characteristic of a rapid population expansion, such as domestication.*

4.4.3.1. Haplotype distribution and phylogeny of ancient Bos samples using Maximum Likelihood analysis

Out of the 127 ancient samples presented here, 96 belong to the T3 haplogroup, 3 belong to T2, 1 is T1, 7 are Q, 6 are P, and 2 are possible new haplogroups (Fig 4.4 & Fig 4.7).

The Mesolithic and Neolithic auroch samples from Germany, Poland and Hungary belong to the haplogroup P, however one morphologically auroch sample, Kalisz 3, has a Q mitochondria. Unfortunately this sample does not have temporal information. The possible auroch sample from Catal Hoyuk in Anatolia, Ch22, has a T3 haplotype. Sample VEM189, a very low endogenous DNA sample from the Cambridge Fens of an unknown age, is a T3 haplogroup despite its morphological designation as auroch, similar to the possible Roman auroch Bri1. Interestingly, the Medieval auroch sample, Fal7, despite younger age possess a P mtDNA haplogroup. Within the P haplogroup structure is suggested (Fig 4.4). The haplogroup splits into sub-haplogroups, there is no ascertainable temporal structure to these sub-haplogroups. However, there may be a geographical pattern emerging with two geographical clades; a primarily Eastern European clade and Western European clade.

The Neolithic domestic samples from Southwest Asia possess haplogroups T2,T3 and Q. This diversity is also present in Southeast Europe, however the diversity of haplogroups decreases through Germany and into Britain and Ireland as T3 becomes the dominant haplogroup of Northern Europe. Post-Neolithic the lack of haplogroup diversity within Northern Europe continues through to the Medieval period, with only T3 present in domestic cattle from these areas. In the middle East and North Africa diversity in haplogroups post-Neolithic is suggested as the Bronze Age sample from Iran contains a Q haplogroup while the two Bronze Age Egyptian samples are T1 and T2.

Sample VEM090 is the only ancient T1 sample in this dataset. This Bronze Age Egyptian, from the XI Dynasty, sample clusters with modern Ethiopian and Egyptian T1 samples, creating an outgroup to the Ethiopian samples within the T1b sub-haplogroup. It possess the T1b diagnostic mutations of 7542 and 16022 (Achilli et al. 2009). The Bronze Age Egyptian VEM104 belongs the T2 haplogroup, forming an outgroup to the Neolithic T2 samples (Fig 4.4) from Iran, Anatolia and South-East Europe.

The T3 haplogroup shows a comb like pattern of shallow branch lengths. Within this haplogroup, there is a lack of geographic or temporal clustering. Samples from the same sites with more than one individual are spread across the T3 haplogroup, including small island sites such as Ness of Brodgar (Orkney,UK) and Cladh Hallan (South Uist, Outer Hebrides, UK). All four samples from Ness of Brodgar (NessX) contain different haplotypes, while all nine samples from Cladh Hallan (ClaX) possess different haplotypes. In a larger context the ancient samples from Ireland and Britain do not specifically cluster with their geographic contemporaries. The only Northern European sample which is not T3 is the sample German Late Neolithic Kir4 (Fig 4.4). This mtDNA sequence is divergent, with placement within the tree as a sister clade to the Mesolithic Chinese C (Zhang et al. 2013) haplogroup or as an outgroup to R (Fig 4.5). When compared with the partial sequence of E haplogroup (Edwards et al. 2007), it only shares two of the eight published D-Loop mutations (Achilli et al. 2009) away from the *B. taurus* reference sequence (Anderson et al. 1982), mutations 16133 and 161388; a recurrent back mutation (Achilli et al. 2009). This sample is likely an example of a new unpublished haplogroup.

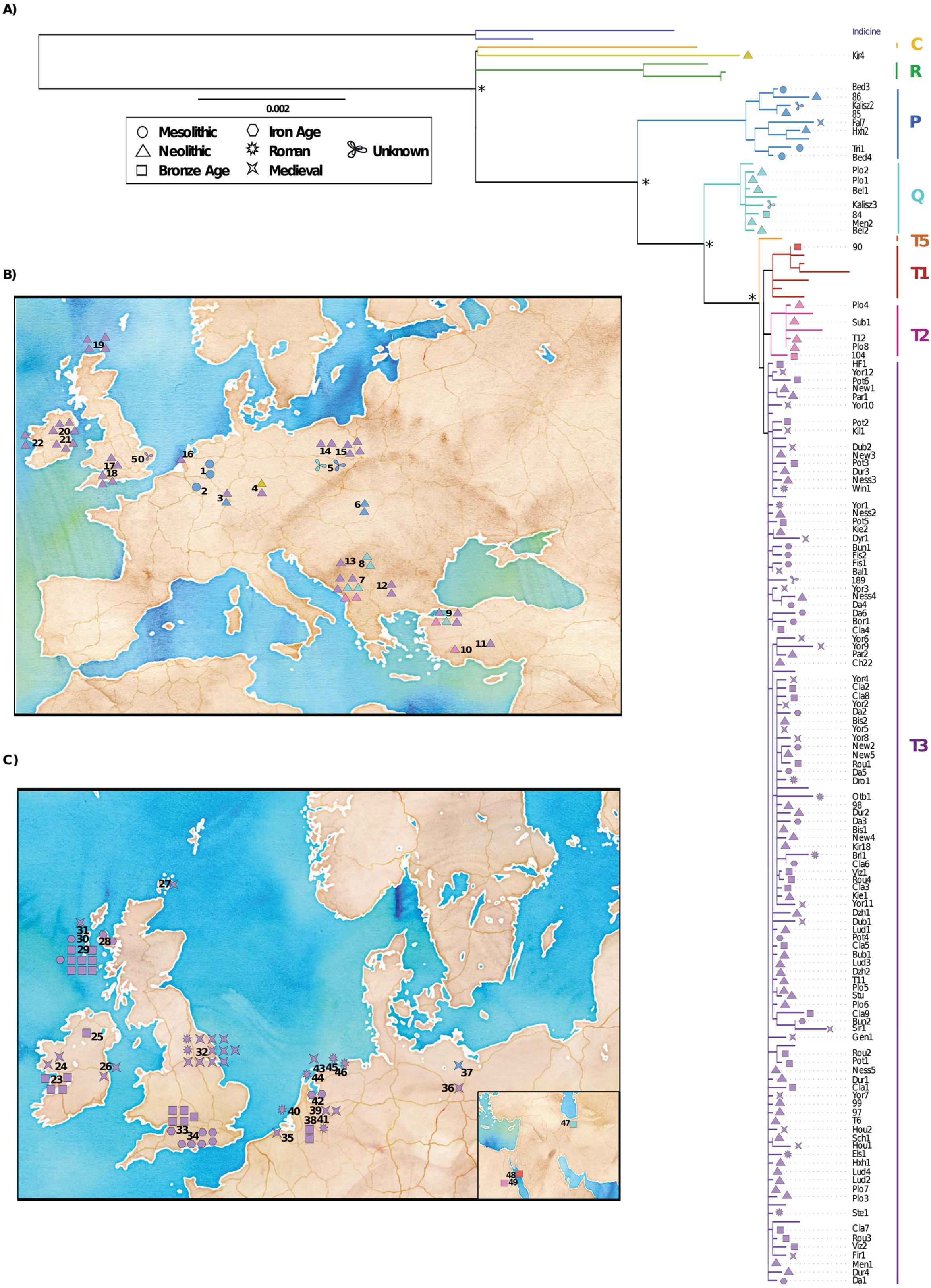


Fig 4.4. *Maximum likelihood tree of predominantly ancient samples and maps to demonstrate the geographic and temporal distribution of Bos haplogroups. A) Maximum likelihood tree, main branches with high support (1) are indicated by *. Haplogroups are denoted by colour, with the cultural period of a sample is indicated by a symbol at the end of the branch. Only ancient samples sequenced for this Chapter are denoted. B) Map of Europe with the distribution of haplogroups for the Mesolithic and Neolithic cultural periods. C) Map of Northern Europe, with an inset of Iran and Egypt, for the post-Neolithic period. Chalcolithic and Bronze Age are treated here as the Bronze Age for simplicity of labelling. See Table 4.4 for the archaeological sites.*

4.4.3.2. Timing the Divergence of Bos Haplogroups - A Bayesian Phylogeny of Cattle

The tree topology of the two Bayesian phylogenies are comparable (Fig 4.5) and are similar to the maximum likelihood tree (Fig 4.3). The branch split times between the two trees differ slightly, with the tree formed of modern and radiocarbon dated samples only containing slightly younger split times to the splits shown on the tree formed of modern samples, radiocarbon dated and temporally well defined samples (Fig 4.5a and Fig 4.5b). The splits times are given for the main bifurcations with high posterior probabilities from the analysis that used modern and radiocarbon dated samples only; the larger analysis did not have convergence of the MCMC chains for some of the priors. The median time point for the indicine vs taurine is dated to 127,446 yrBP (95% HPD 105,425-149,080 yrBP). The clade containing the R, C and Kir4 haplogroups splits from the rest of the tree 60,055 yrBP (median) (95% HPD 50,357-69,734 yrBP). The next split occurred 33,292 (median) (95% HDP 27,475-39,255 yrBP), which has the European auroch P haplogroup split from the Q/T clades. The Q haplogroup branches off from the T haplogroups 19,259 yearBP (median) (95% HDP 15,478-23,685 yrBP). The only split within the T haplogroups that has a posterior probability of 1 is the T5 vs T2/T1/T3 split occurring at a median value of 10,704 yrBP (95% HDP 9,182-12,565 yrBP). While at low posterior probabilities the split times from the T haplogroups are; T2 vs T1/T3 9,973 yrBP (median) (95% HDP 8,838-11,417 yrBP) and T1 vs T3 8,836 yrBP (median) (95% HDP 7,266 -10,487 yrBP).

The posterior probabilities for the branch split of C vs Kir4/R and Kir4 vs R are low 0.3167 and 0.6044 respectively, and are not denoted on the tree (Fig 4.5). These timings are; split of C vs Kir4/R is dated to 47,594-66,761 yrBP (95% HDP), while the split of Kir4 vs R dates to 44,920-63,831 yrBP (95% HDP). The splits within the P haplogroup are well supported with posterior values of 0.9998 or above. The first split between Bed3 and the rest of the samples occurred between 14,650 – 19,914 yrBP (95% HDP), with the next split occurring between two sister clades occurring 12,828-16,898 yrBP (95% HDP). The split between the two Mesolithic samples Bed4 and Tri1 occurred between 11,843 and 14,529 yrBP (95% HDP), while the split between Mesolithic CPC and modern day DQ124389.1 occurred between 9,372 yrBP and 14,179 yrBP (95% HDP). The split between the ancient Q Men2 and the modern Qs has a high posterior probability of 1, as does the split of the T2 Sub1 and the modern T2s, occurring 7,839-9,433 yrBP (95% HDP) and 8,020- 8,794 yrBP (95% HDP), respectively.

The mutation rate (Table 4.5) generated for the D-loop is similar to that generated by Ho and colleagues (2008) when invariant sites were assumed, $5.89X10^{-7}$ substitutions per site per year (95% HPD: $4.66x10^{-7}$ - $7.10X10^{-7}$) and similar to the $4.6X10^{-7}$ calculated for Bison Bayesian phylogeny (Massilani et al. 2016). The mutation rates for the other partitions are also similar to those calculated by Massilani and colleagues (2016).

Table 4.5. *Mutation rates for the mtDNA partitions as calculated by BEAST. Results are equivalent to substitutions per site per year.*

Mutation Rate	**mtDNA partitions**				
	C1C2	**C3**	**D**	**rRNA**	**tRNA + reminder**
mean	3.24E-08	1.08E-07	4.90E-07	3.21E-08	3.11E-08
stderr of mean	9.78E-11	3.05E-10	1.60E-09	9.51E-11	9.41E-11
median	3.22E-08	1.08E-07	4.96E-07	3.19E-08	3.08E-08
95% HPD	2.6808E-8 - 3.84E-8	9.2663E-8 - 1.2503E-7	3.7859E-7 - 6.0585E-7	2.4584E-8 - 4.0034E-8	2.2401E-8 - 4.0288E-8

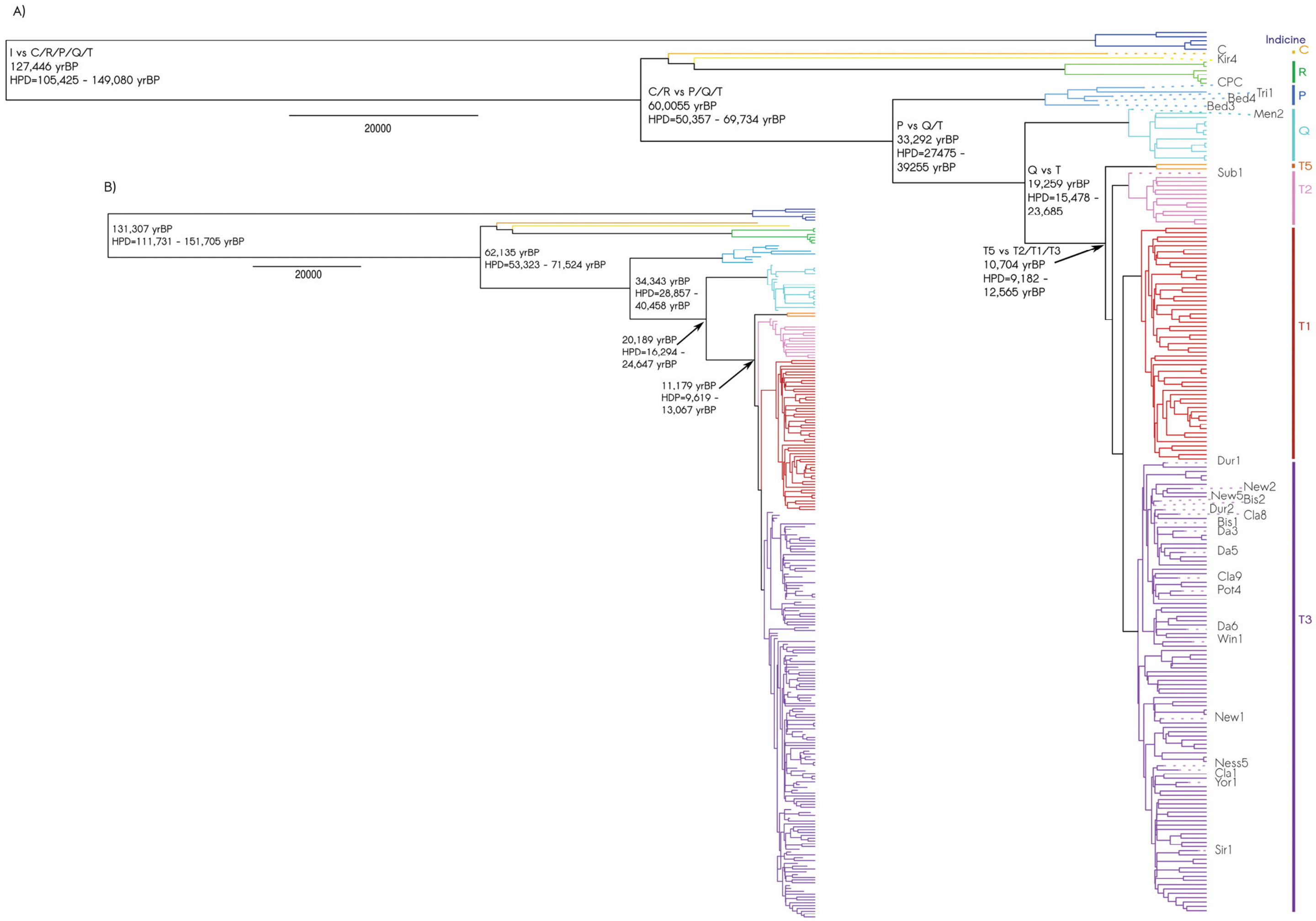


Fig 4.5. *Bayesian estimate of maximum clade credibility tree using 1modern and ancient Bos samples. A) Analysis run with 183 modern samples and 26 radiocarbon dated ancient samples only, including previously published ancient mitogenomes of CPC and C. Branch lengths have been extended and samples jittered in order to read names. Ancient samples are named while modern samples used are given in Appendix Table 2. B) Analysis run with modern and radiocarbon dated and non radiocarbon dated ancient samples. For both A) and B) the median and 95% highest probability density (HPD) split dates are provided for branches with Bayesian posterior probabilities of 1.*

4.4.2.3. VEM212 - A 200,000 yrBP Bos primigenius

This sample originates from Norton Bottoms Quarry, North-East of Newark-on-Trent in Lincolnshire (UK). The skull of an auroch was uncovered in 2008 in the Balderton Sand and Gravel of the river Trent terrace sequence; a sequence that spans the Marine Oxgygen Isotope Stages 8-7-6. The interglacial sediment in which this skull was found correlates with MIS 7 (243,000-191,000 yrBP). Mammalian biostratigraphical evidence suggests that the assemblage belonged to second part of this interglacial, MIS7a, which correlates to approximately 200,000 yrBP (Danielle Schreve *per. comm.* & Bridgland et al. 2014).

The endogenous percentage of this sample was very low (Chapter 3), however after shotgun sequencing on one lane of Illumina HiSeq^{TM} 2000 and two attempts at mitochondrial target capture, a mtDNA consensus was achieved (section 4.2.7). When aligned to the European auroch sample CPC (GU985279.1) the average coverage is 1.86X with 674 reads passing filtering, covering 72.71% of the mitogenome once and only 47.97% twice. When aligned to the *B. taurus* reference genome (V00654.1) (Anderson et al. 1982) the average coverage is lower at 1.77X with 644 reads passing filtering, covering 71.2% of the mitogenome once and 45.87% twice. The majority of the D-loop is not covered, likely due to high divergence, as well as damage, preventing alignment due to too many mismatches within a read. The reads have damage typical of ancient DNA (Hofreiter et al. 2001), with 44% of the first base pair of the reads showing deamination on the 5'prime end (Fig 4.6a). The read length distribution shows read length shifted towards the 30-50 base pair length (Fig 4.6b). When aligned to the *Bos taurus* mtDNA reference genome 14 of the 91 (Appendix Table 4) mutations are common to those described by (Achilli et al. 2009) that postdate the taurine/indicine split but predate the REPQT split, while a further two mutations are believed to have occured on the indicine lineage post the taurine/indicine split.The BioNJ tree (Fig 4.7) indicates this sample is basal to the taurine and primigenius haplogroups.

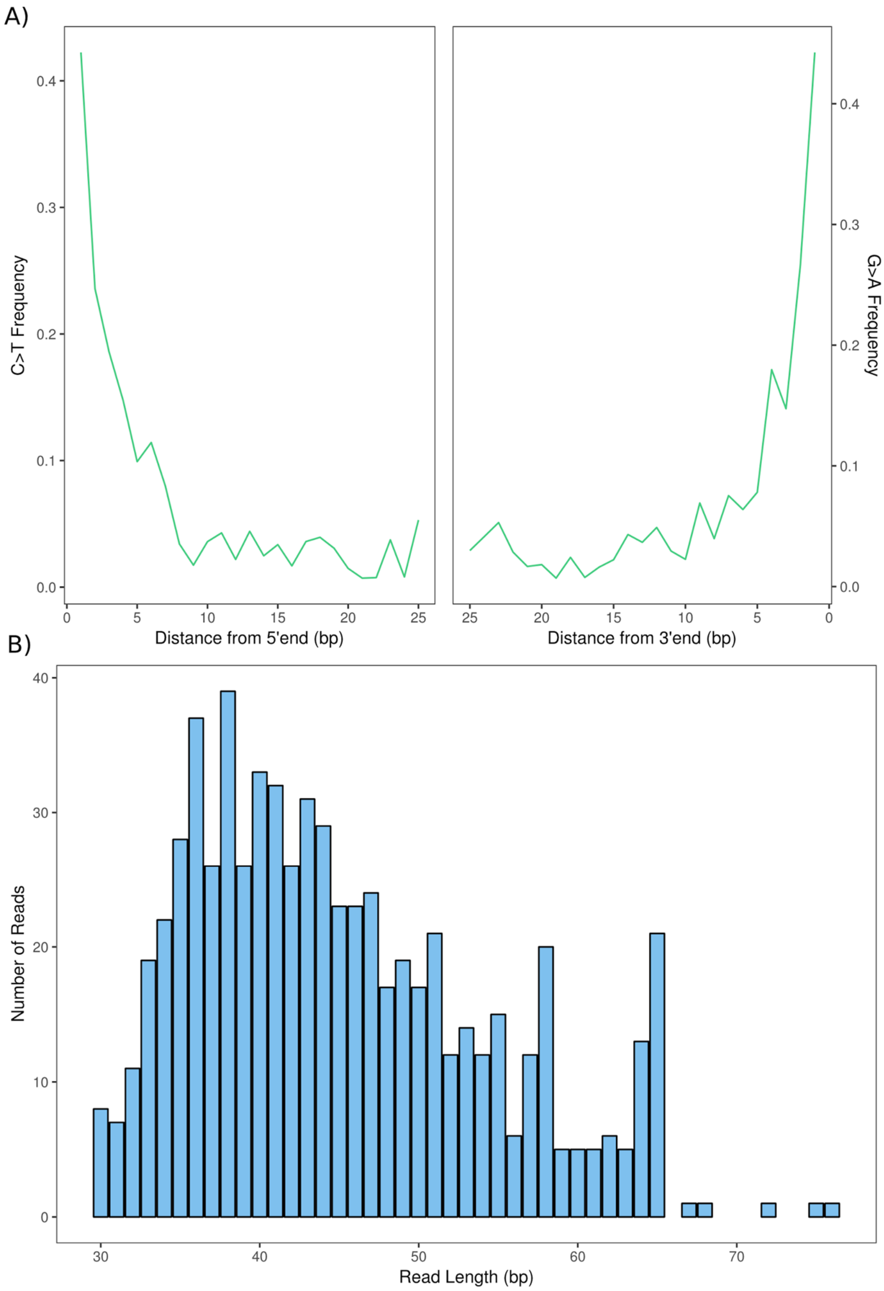


Fig 4.6. *Damage pattern (A) and read length distribution (B) of VEM212 - a 200,000 year old British Aurochs from Norton Bottoms Quarry in Lincolnshire, UK. The sample possess high substitution rates in the first base pair at both the 5'prime and 3'prime ends, expected of a sample of such an age. The fragmentation pattern of the DNA demonstrates the degraded nature of this sample.*

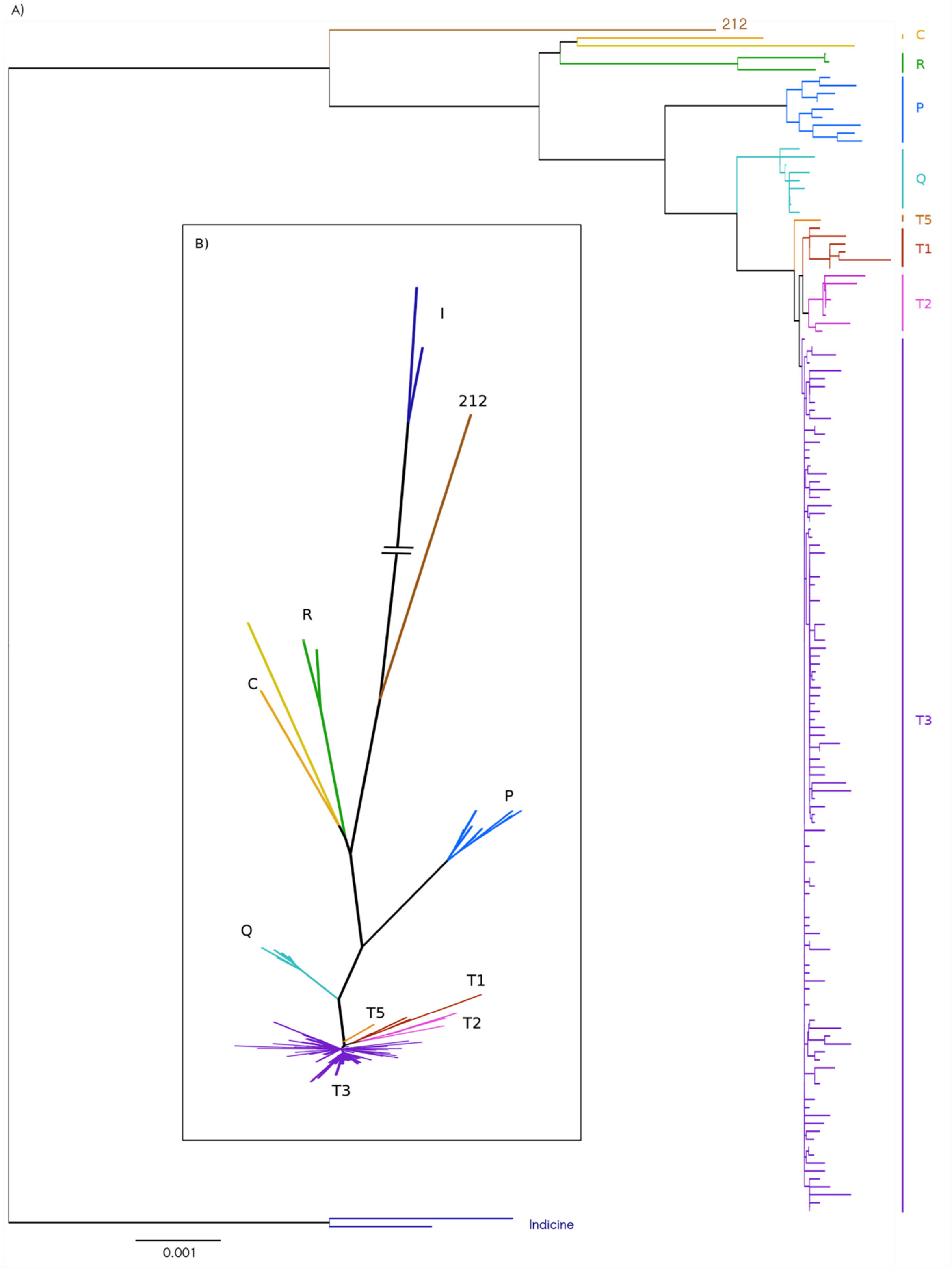


Fig 4.7. *Neighbour-joining tree of 126 ancient samples and X modern samples to demonstrate the basal placement of VEM212 - a 200,000 year old British auroch - in the Bos phylogeny. A) Tree rooted by Bos indicus, B) An unrooted tree. Published haplogroups are denoted on the right-hand side.*

Table 4.2. *Coverage and haplogroup results for samples subject to mtDNA capture and further analysed in this chapter. * Morpholigically Bos primigenius.*

Name	Cultural Period	Radiocarbon Date calBC (2 Sigma)	Site	Country	Coverage	Species	Haplogroup
VEM104	Bronze Age		Deir el Bahri	Egypt	14.82	*b. taurus*	T2
VEM090	Bronze Age		Thoth Hill	Egypt	134.08	*b. taurus*	T1
VEM132	Neolithic		Durrington Walls	England	6.83	*b. taurus*	T3
VEM212	Paleolithic		Norton Bottoms	England	1.86	*b. primigenius*	New
VEM189	Unknown		Cambridge Fens	England	7.66	*b. taurus* *	T3
VEM086	Neolithic		Ecsegfalva	Hungry	22.32	*b. primigenius*	P
VEM085	Neolithic		Ecsegfalva	Hungry	33.46	*b. primigenius*	P
VEM084	Chalcolithic		Maral Tappeh	Iran	17.78	*b. taurus*	Q
Dub2	Medieval		Dublin	Ireland	81.39	*b. taurus*	T3
VEM098	Neolithic		Kilshane	Ireland	9.74	*b. taurus*	T3
VEM097	Neolithic		Kilshane	Ireland	12.65	*b. taurus*	T3
VEM099	Neolithic		Kilshane	Ireland	29.02	*b. taurus*	T3
Turk6	Neolithic		Mentese	Turkey	11.47	*b. taurus*	T3
Turk12	Neolithic		Mentese	Turkey	17.42	*b. taurus*	T2
Turk11	Neolithic	6076-5983 BC	Mentese	Turkey	44.59	*b. taurus*	T3

Table 4.3. *Coverage and haplogroup results for mtDNA of samples subject to shotgun sequencing and analysed in this chapter.*

Name	Cultural Period	Radiocarbon Date calBC (2 Sigma)	Site	Country	Coverage	Species	Haplogroup
Gen1	Medieval		Gent	Belgium	21.24	*b. taurus*	T3
Dzh1	Early Neolithic		Dzulhunitsa	Bulgaria	2043.32	*b. taurus*	T3
Dzh2	Early Neolithic		Dzulhunitsa	Bulgaria	976.4	*b. taurus*	T3
Bis1	Neolithic	3529-3362BC	Bishops Cannings	England	201.64	*b. taurus*	T3
Bis2	Neolithic	3638-3512BC	Bishops Cannings	England	231.57	*b. taurus*	T3
Yor3	Medieval		Coppergate	England	201.29	*b. taurus*	T3
Yor4	Medieval		Coppergate	England	242.09	*b. taurus*	T3
Yor7	Medieval		Coppergate	England	15.84	*b. taurus*	T3
Da2	Iron		Danebury	England	20.79	*b. taurus*	T3
Da4	Iron		Danebury	England	29.49	*b. taurus*	T3
Da6	Iron	203-43BC	Danebury	England	182.42	*b. taurus*	T3
Da1	Iron		Danebury	England	172.28	*b. taurus*	T3
Da3	Iron	404-354BC	Danebury	England	188.33	*b. taurus*	T3
Da5	Iron	511-357BC	Danebury	England	131.67	*b. taurus*	T3
Dur3	Late Neolithic		Durrington Walls	England	33.10	*b. taurus*	T3
Dur1	Late Neolithic	2410-2199BC	Durrington Walls	England	158.40	*b. taurus*	T3
Dur2	Late Neolithic	2695-2457BC	Durrington Walls	England	119.63	*b. taurus*	T3
Yor9	Medieval		York	England	26.55	*b. taurus*	T3
Yor2	Medieval		York	England	212.69	*b. taurus*	T3
Yor8	Medieval		York	England	19.10	*b. taurus*	T3
Yor10	Medieval		York	England	35.00	*b. taurus*	T3
Yor12	Medieval		York	England	23.50	*b. taurus*	T3
Yor6	Medieval		York	England	169.10	*b. taurus*	T3
Yor5	Medieval		York	England	122.50	*b. taurus*	T3
Pot2	Late Bronze Age		Potterne	England	17.32	*b. taurus*	T3
Pot5	Late Bronze Age		Potterne	England	43.49	*b. taurus*	T3
Pot6	Late Bronze Age		Potterne	England	94.91	*b. taurus*	T3
Pot3	Late Bronze Age		Potterne	England	46.12	*b. taurus*	T3
Pot1	Late Bronze Age		Potterne	England	217.29	*b. taurus*	T3
Pot4	Early Iron	769-472BC	Potterne	England	251.18	*b. taurus*	T3
Yor1	Roman	2 BC- 135 AD	York	England	603.00	*b. taurus*	T3
Yor11	Roman		York	England	22.60	*b. taurus*	T3
Bed3	Mesolithic	9805-9385 BC	Bedburg-Königshoven	Germany	3532.02	*b. primigenius*	P
Bed4	Mesolithic	10036-9647 BC	Bedburg-Königshoven	Germany	657.91	*b. primigenius*	P

Dyr1	Medieval		Dyrotz 25	Germany	103.38	*b. taurus*	T3
Fal7	Medieval		Falkenwalde	Germany	195.92	*b. primigenius*	P
Hxh1	Neolithic		Herxheim	Germany	980.22	*b. taurus*	T3
Hxh2	Neolithic		Herxheim	Germany	2209.33	*b. primigenius*	P
Kir18	Neolithic	3798-3641 BC	Kirschbaumhöhle	Germany	1474.92	*b. taurus*	T3
Tri1	Mesolithic	8261-7960 BC	Trier-Brüderkrankenhaus	Germany	822.2	*b. primigenius*	P
Kir4	Neolithic	2816-2671 BC	Kirschbaumhöhle	Germany	812.77	*b. taurus*	new
Bal1	Medieval		Ballybane	Ireland	57.72	*b. taurus*	T3
Bal2	Medieval		Ballybane	Ireland	54.52	*b. taurus*	T3
Dub1	Medieval		Dublin	Ireland	153.36	*b. taurus*	T3
New2	Late Iron	321-425 AD	Newgrange	Ireland	194.30	*b. taurus*	T3
New3	Neolithic		Newgrange	Ireland	345.90	*b. taurus*	T3
New4	Neolithic		Newgrange	Ireland	244.46	*b. taurus*	T3
New1	Neolithic	2795-2580 BC	Newgrange	Ireland	777.90	*b. taurus*	T3
New5	Neolithic	3024-2894 BC	Newgrange	Ireland	322.50	*b. taurus*	T3
Par1	Neolithic		Parknabinia	Ireland	31.96	*b. taurus*	T3
Par2	Neolithic		Parknabinia	Ireland	31.18	*b. taurus*	T3
Rou4	Bronze Age		Roughan Hill	Ireland	7.56	*b. taurus*	T3
Rou3	Bronze Age		Roughan Hill	Ireland	27.74	*b. taurus*	T3
Rou1	Bronze Age		Roughan Hill	Ireland	251.50	*b. taurus*	T3
Rou2	Bronze Age		Roughan Hill	Ireland	165.79	*b. taurus*	T3
HF1	Bronze Age		Haughey's Fort	N. Ireland	73.50	*b. taurus*	T3
Win1	Roman	417-561 AD	Winsum	Netherlands	855.71	*b. taurus*	T3
Dro1	Roman		Dronrijp-zuid	Netherlands	317.94	*b. taurus*	T3
Bri1	Roman		Britsum	Netherlands	207.49	*b. primigenius**	T3
Otb1	Roman		Oosterbeintum	Netherlands	166.77	*b. taurus*	T3
Els1	Roman		Elst	Netherlands	179.29	*b. taurus*	T3
Fir1	Medieval		Firdgum	Netherlands	185.48	*b. taurus*	T3
Sch1	Neolithic		Schipluiden	Netherlands	158.61	*b. taurus*	T3
Ste1	Iron Age/ Roman		Stevenshofjesp	Netherlands	257.5	*b. taurus*	T3
Hou1	Medieval		Houten-Loerik	Netherlands	202.11	*b. taurus*	T3
Viz1	Bronze Age		Vianen-Zijdervel	Netherlands	574.78	*b. taurus*	T3
Viz2	Bronze Age		Vianen-Zijdervel	Netherlands	251.41	*b. taurus*	T3
Hou2	Medieval		Houten-Hoogdij	Netherlands	194.52	*b. taurus*	T3
Bun1	Iron Age/ Roman		Bunnik	Netherlands	242.46	*b. taurus*	T3
Bun2	Iron Age		Bunnik	Netherlands	252.56	*b. taurus*	T3
Kalisz2	no info		Kalisz	Poland	416.71	*b. primigenius*	P
Kalisz3	no info		Kalisz	Poland	11.44	*b. primigenius*	Q
Kie1	Neolithic		Kierzkowo	Poland	5.52	*b. taurus*	T3

Kie2	Neolithic		Kierzkowo	Poland	13.18	*b. taurus*	T3
Lud1	Neolithic		Ludwinowo	Poland	10.3	*b. taurus*	T3
Lud2	Neolithic		Ludwinowo	Poland	29.53	*b. taurus*	T3
Lud3	Neolithic		Ludwinowo	Poland	31.46	*b. taurus*	T3
Lud4	Neolithic		Ludwinowo	Poland	16.06	*b. taurus*	T3
Bor1	Iron Age		Bornais	Scotland	33.90	*b. taurus*	T3
Cla1	Bronze	1054-826 BC	Cladh Hallan	Scotland	679.56	*b. taurus*	T3
Cla2	Bronze		Cladh Hallan	Scotland	187.79	*b. taurus*	T3
Cla6	Iron Age		Cladh Hallan	Scotland	28.55	*b. taurus*	T3
Cla7	Bronze Age		Cladh Hallan	Scotland	28.10	*b. taurus*	T3
Cla3	Bronze Age		Cladh Hallan	Scotland	15.54	*b. taurus*	T3
Cla4	Bronze Age		Cladh Hallan	Scotland	27.07	*b. taurus*	T3
Cla5	Bronze Age		Cladh Hallan	Scotland	220.03	*b. taurus*	T3
Cla8	Bronze Age	1060-899BC	Cladh Hallan	Scotland	211.55	*b. taurus*	T3
Cla9	Bronze Age	1118-924BC	Cladh Hallan	Scotland	241.36	*b. taurus*	T3
Fis1	Iron Age		Fiskavaig	Scotland	227.63	*b. taurus*	T3
Fis2	Iron Age		Fiskavaig	Scotland	243.29	*b. taurus*	T3
Kil1	Medieval -Norse		Kilpheder	Scotland	26.66	*b. taurus*	T3
Ness4	Late Neolithic		Ness of Brodgar	Scotland	24.98	*b. taurus*	T3
Ness3	Late Neolithic		Ness of Brodgar	Scotland	30.51	*b. taurus*	T3
Ness2	Late Neolithic		Ness of Brodgar	Scotland	170.49	*b. taurus*	T3
Ness5	Late Neolithic	2588-2464BC	Ness of Brodgar	Scotland	223.96	*b. taurus*	T3
Sir1	Medieval - Norse	1039-1215AD	Snusgar	Scotland	269.19	*b. taurus*	T3
Bel1	Neolithic		Belovode- Veliko	Serbia	12.22	*b. taurus*	Q
Bel2	Neolithic		Belovode- Veliko	Serbia	26.25	*b. taurus*	Q
Bub1	Neolithic		Bubanj	Serbia	53.09	*b. taurus*	T3
Plo1	Neolithic		Plocnik	Serbia	10.41	*b. taurus*	Q
Plo2	Neolithic		Plocnik	Serbia	36.47	*b. taurus*	Q
Plo3	Neolithic		Plocnik	Serbia	152.81	*b. taurus*	T3
Plo4	Neolithic		Plocnik	Serbia	153.66	*b. taurus*	T2
Plo5	Neolithic		Plocnik	Serbia	13.17	*b. taurus*	T3
Plo6	Neolithic		Plocnik	Serbia	7.96	*b. taurus*	T3
Plo7	Neolithic		Plocnik	Serbia	19.2	*b. taurus*	T3
Plo8	Neolithic		Plocnik	Serbia	23.82	*b. taurus*	T2
Stu1	Neolithic		Stubline	Serbia	20.14	*b. taurus*	T3
Men1	Neolithic		Mentese	Turkey	333.85	*b. taurus*	T3
Sub1	Neolithic	6221-6024 BC	Suberde	Turkey	987.5	*b. taurus*	T2
Men2	Neolithic	6048-5893 BC	Mentese	Turkey	88.04	*b. taurus*	Q
Ch22	Neolithic		Catal Hoyuk	Turkey	46.89	*b. taurus**	T3

Table 4.4. *Corresponding site names for the site numbers featured in Figure 4.4*

Site Number	Site Name	Country
1	Bedburg-Königshoven	Germany
2	Trier-Brüderkrankenhaus	Germany
3	Herxheim	Germany
4	Kirschbaumhöhle	Germany
5	Kalisz	Poland
6	Ecsegfalva	Hungary
7	Plocnik	Serbia
8	Belovode- Veliko	Serbia
9	Mentese	Turkey
10	Suberde	Turkey
11	Catal Hoyuk	Turkey
12	Dzulhunitsa	Bulgaria
13	Stubline	Serbia
14	Kierzkowo	Poland
15	Ludwinowo	Poland
16	Schipluiden	Netherlands
17	Bishops Cannings	England
18	Durrington Walls	England
19	Ness of Brodgar	Scotland
20	Newgrange	Ireland
21	Kilshane	Ireland
22	Parknabinia	Ireland
23	Roughan Hill	Ireland
24	Ballybane	Ireland
25	Haughey's Fort	N. Ireland
26	Dublin	Ireland
27	Snusgar	Scotland
28	Fiskavaig	Scotland
29	Cladh Hallan	Scotland
30	Bornais	Scotland
31	Kilpheder	Scotland
32	York	England
33	Potterne	England
34	Danebury	England
35	Gent	Belgium
36	Dyrotz	Germany
37	Falkenwalde	Germany
38	Vianen-Zijdervel	Netherlands
39	Houten-Loerik	Netherlands
40	Stevenshofjesp	Netherlands
41	Elst	Netherlands
42	Bunnik	Netherlands
43	Firdgum	Netherlands
44	Oosterbeintum	Netherlands
45	Britsum	Netherlands
46	Winsum	Netherlands
47	Maral Tappeh	Iran
48	Thoth Hill	Egypt
49	Deir el Bahri	Egypt
50	Cambridge Fens	England

4.5. Discussion

4.5.1. Target Capture aDNA Preservation

The majority of the samples on which mtDNA capture was attempted had very low endogenous DNA percentages when subjected to shotgun sequencing analysis. However some were successfully captured and were able to be used for further analysis, suggestive of different preservation of mtDNA and nuDNA. Not only is mtDNA in high copy number within cells (Miller et al. 2003), experiments have indicated that nuDNA degrades faster than mtDNA (Allentoft et al. 2012). Therefore it is not surprising that capture of mtDNA from samples with little nuDNA concentration can be possible.

The sequence reads from the two Egyptian Bronze Age skin samples show less damage when compared to those from bone samples from similar time periods and climate. There are several possible explanations. Both samples are from archaeologically significant sites near Luxor Egypt. However neither are directly dated, so it could be argued that the samples are younger than believed and therefore less DNA deamination has taken place. One could argue the sequencing could be contamination. However controls and precautions standard to aDNA were taken, no other T1 and only one other T2 was captured at the same time. Also coverages of these samples, especially VEM90, would suggest that they are not the result of contamination. The postmortem environment of the samples, and possibly the treatment post-mortem (the samples were partially mummified), may have helped reduce the damage to DNA. aDNA extracted from parchment has also shown reduced deamination patterns than contemporaneous bone samples, possibly due to parchment-making process which takes animal hide and turns it into parchment via numerous stages including soaking in lime and scraping off hair (Teasdale et al. 2017). Additionally both parchment and these two samples are soft tissue, as opposed to bone; perhaps DNA degradation within such proceeds in a different manner to within bone.

4.5.2. Holocene Cattle Phylogeny

4.5.2.1. Haplogroup Geography and Temporal Variation

Several past analysis have discussed the distribution of haplogroups utilising modern and ancient mtDNA. However to date no ancient cattle mitogenome dataset of this size has been published. The maximum likelihood phylogeny of the modern and ancient samples (Fig 4.3) mirrors previous whole mitogenomes analysis of modern cattle, in addition to partial mtDNA phylogenies of modern and aDNA samples (Troy et al. 2001; Edwards et al. 2007; Ho et al. 2008; Achilli et al. 2009; Stock et al. 2009; Bonfiglio et al. 2012; Olivieri et al. 2015). As the majority of the published haplogroups denoted on the trees have ancient samples assigned, allowing a geographic and temporal look at the cattle haplogroups.

The variety of haplogroups sequenced in the Neolithic Near Eastern animals concurs with past work on partial mtDNA sequences that have identified the Near East as the centre of domestication, due to the haplotype diversity present in modern and ancient samples from this region (Troy et al. 2001; Bollongino et al. 2006; Scheu et al. 2015). During the Neolithic this diversity in haplogroups is not only present in the Anatolia, but also in Southeast Europe, supporting a previous ancient DNA paper using partial D-loop sequences (Scheu et al. 2015), and indicates that diversity of cattle haplogroups was not lost on the initial migration route from Anatolia to Southeast Europe. Although the Polish sample Kalisz3, identified morphologically as auroch (Daniel Bradley *per. comm.*), possesses a Q mtDNA haplogroup lack of temporal information regarding this sample makes it hard to discern the ramifications on our current understanding of the Q haplogroup. Previous studies have shown Q to be a divergent minor taurine haplogroup, originally identified in modern Italian cattle (Achilli et al. 2008), but has since been identified in ancient cattle from the Southeast Europe as well as at low frequency in Neolithic Central Europe, while in modern cattle it is present in the near East and Egypt (Olivieri et al. 2015; Scheu et al. 2015). The presence of Q in a morphologically ID auroch from Poland could suggest a wide geographic range for the wild haplogroup, though some may argue the sample a hybrid or to be a very large domestic. A radiocarbon date is essential to learn if this animal predates domestic cattle in Europe or not.

As previous papers utilising modern mitogenomes and partial ancient mtDNA sequences have identified (Troy et al. 2001; Bollongino et al. 2006; Scheu et al. 2015), a loss of haplogroup diversity occurs as cattle migrated to northern parts of Europe where T3 is the almost exclusive haplogroup. Post-Neolithic T3 remains the only haplogroup identified in the north west European ancient samples presented here, suggesting that while haplogroup diversity in Southeastern Europe survived into the Bronze Age (Scheu et al. 2015) there were no large scale subsequent migrations of cattle into Northwestern Europe brought with them more haplogroup diversity. There is little chance of attributing any temporal or geographic structure to the T3 haplogroup. The comb-like structure of the T3 haplogroup is indicative of a bottleneck and a subsequent rapid expansion event, such as domestication from a small genetic pool and subsequent migration (Bollongino et al. 2012). Geographically isolated island populations show a diversity in haplotypes present, suggesting that the herds from sites such as the Neolithic site of the Ness of Brodgar

(Orkney) and the Bronze Age site of Cladh Hallan (Outer Hebrides) were founded on multiple mitochondrial lineages.

The lack of P haplogroups in the ancient domestic cattle sequenced here, is in keeping with the majority of other mtDNA studies (Troy et al. 2001; Edwards et al. 2004; Edwards et al. 2007; Scheu et al. 2008). The P haplogroup has been sequenced in a modern Korean animal (Accession number: DQ124389), a Neolithic domestic individual from Switzerland and a Neolithic domestic individual from Orkney (Schibler et al. 2014; Fraser et al. 2017). However in general the P haplogroup is not observed in domestic animals. This lack of presence was argued as proof of lack of Northern European auroch introgression into domestic stock (Edwards et al. 2007; Scheu et al. 2008). However more recently nuclear analysis has shown auroch introgression into modern European breeds (Park et al. 2015). As this introgression is not traceable on the maternal line, it was likely male mediated.

The one European Neolithic sample that defied the T3 haplogroup assignment, was Kir4, a German Late Neolithic sample associated with the Corded Ware culture. The divergent mitochondrion clusters with the other ancient divergent mitochondria from the Chinese Mesolithic sample (Zhang et al. 2013). This juvenile sample is from a cave where the majority of animal samples are domestic (Seregély 2014). Kir4 is a representation of a new haplogroup, perhaps from an auroch population not previously sequenced. The highly divergent Neolithic E haplogroup from Germany indicates that there were other auroch populations in Northern Europe during the Neolithic (Edwards et al. 2007), perhaps other haplogroups were present but at a lower frequency than P and therefore have not survived the archaeological record. Alternatively, Kir4 may be an alien, a domestic migrant from the east that travelled with, for example, human steppe migrations that are associated with the Corded Ware culture (Haak et al. 2015; Allentoft et al. 2015). Perhaps a relict of a female introgression into a herd prior to a migration to the west of Europe.

While only three samples represent post-Neolithic Near East and Egypt, none of these samples are T3. An Iranian Q, an Egyptian T2 and an Egyptian T1 indicate a similar trend in ancient cattle as described in modern cattle, that this region has a greater haplogroup diversity (Troy et al. 2001; Olivieri et al. 2015). The placement of the Bronze Egyptian as an outgroup to Neolithic Anatolian, South Eastern Neolithic and modern T2 perhaps indicates a geographic structure to T2, however more ancient and modern whole mitogenomes samples are needed to test this. These two Egyptian samples, mark the earliest domestic cattle ancient mtDNA from the African continent and the first whole ancient mitogenomes, indicating that both T2 and the main African haplogroup of T1 were present in Egypt during the Bronze Age. While Iron Age samples and modern African cattle are almost exclusively T1 (Loftus et al. 1994; Bradley et al. 1996; Troy et al. 2001; Edwards et al. 2004; Bonfiglio et al. 2012), diversity of haplogroups does exist in modern day Egypt (Lenstra et al. 2014). The T1 haplogroup in modern day cattle has far reaching geographic distribution; from Africa to Europe to South America, due to human mediated movement (Bonfiglio et al. 2012; Lenstra et al. 2014). The assignment of T1b is in keeping with results from modern African mitogenomes. In Bonfiglio and colleagues (2012) the T1b sub haplogroup was the most common haplogroup in Africa (63.8%)

The new European auroch samples presented here, overwhelmingly belong to the P haplogroup, clustering together with the previously published British Mesolithic auroch CPC (Edwards et al. 2010), and supporting previous analysis indicating a predominant P haplogroup identity for Northern European auroch (Bailey et al. 1996; Troy et al. 2001; Scheu et al. 2008; Stock et al. 2009; Zeyland et al. 2013). None of the morphologically denoted European auroch presented here clustered with the C specimen from Mesolithic China (Zhang et al. 2013), suggesting this haplogroup did not populate Europe. However, the auroch haplotype E (Edwards et al. 2007; Stock et al. 2009), which was not included due to the partial nature of the sequence, and the presence of other haplogroups, such as Kir4 and the Q indicates some haplogroup diversity in Northern European aurochs.

The structure of P haplogroup itself shows a deep divergence from sister clades, as well as divergence within the haplogroup. A lack of temporal structure tentatively suggests that at least between the Mesolithic and the Neolithic there were no major population constrictions and subsequent loss of diversity within the European population that could have taken place with the arrival of farming and likely habitat loss. However this hypothesis needs to be tested further, for example a Bayesian skyline plot would help to determine demographic history. Hints of possible geographical structure exist with majority of the Eastern European samples clustering together and the majority of the Western European samples clustering together. Perhaps with a greater number of complete mitogenomes from European auroch would help clarify if an east vs west Europe structure did exist. Tentative morphological evidence suggests a west-east cline of body size of auroch populations with larger animals present in Germany and Poland compared with Britain paralleling differences in size of European wild boar (Wright & Viner-Daniels 2015).

It is interesting that even the Medieval sample denoted as auroch contains the P haplogroup, especially as during and post-Bronze Age it is likely that populations of auroch contracted and became restricted due to landscape alterations by humans (Van Vuure 2005). Given the increase in the number of domestic cattle that would have accompanied farming, it is interesting that there has been no domestic cattle introgression in the maternal line of the Fal7 auroch. The analysis of modern cattle and an auroch nuclear genomes, has shown auroch introgression into domestic cattle occurred prior to the aurochs extinction in 1600s (Park et al. 2015; Upadhyay et al. 2016). Granted female introgression would be more difficult to achieve than male introgression as a domestic cow would need to become feral and join a wild herd. Introgression by domestic male animals can not be estimated using mtDNA.

Sample VEM189, a petrous sampled from a large skull in the Booth Natural History Museum, was believed to be an auroch due to the size of the skull. The sample itself was collected in the 1800s from the Cambridge fens area and limited information regarding it survives (John Cooper *per. comm.*). Bri1 is a Roman Netherlands specimen that is larger than other archaeological bull remains from the period, hence its classification of auroch (Marta Verdugo/Canan Çakırlar *per. comm.*). Both animals exhibit a T3 mitochondria, contrary to expectation as in previous papers on Northern European auroch samples, as all samples bar one have been identified as the P haplotype (Bailey et al. 1996; Troy et al. 2001; Edwards et al. 2007; Scheu et al. 2008; Edwards et al. 2010; Scheu et al. 2015). T3 aurochs have been identified in Italy, and argued as evidence for an independent Italian domestication (Beja-Pereira et al. 2006), but it is more likely that there were different populations of auroch across Europe, as suggested by the morphological evidence (Wright & Viner-Daniels 2015). Perhaps VEM189 and Bri1 were descended from a different auroch population than other Northern European animals, however there are several other highly plausible explanations. These samples may have been descended from an admixture event between a male auroch and a female domestic T3 animal, explaining both the animals size and mtDNA haplogroup. It is also possible that these samples are large domestic animals that have been mistakenly denoted auroch. As aurochs were extirpated in Britain during the Bronze Age (Clutton-Brock 1989), a radiocarbon date would help to distinguish if the VEM189 sample post dates this local extinction or even predates domestic cattle in Britain. Without temporal information for this sample no solid conclusions can be drawn.

Morphologically Ch22, from the Neolithic Catal Hoyuk site in Anatolia (Turkey), cannot be determined as domestic or wild (Amelie Scheu *per. comm.*). Unfortunately, mtDNA can not help to determine this either as the T3 haplotype is associated with domestic cattle, both modern and ancient, in Europe and the Near East. As cattle were domesticated from

Near Eastern auroch (Troy et al. 2001), diversity in the wild these would have been incorporated into the domestic stock.

The only haplogroup not to be represented by ancient data in this dataset is the divergent R haplogroup. This haplogroup has only been found in modern Italian domestic cattle and has been suggested as possible evidence for Italian domestication (Bonfiglio et al. 2010). However to date, no ancient auroch population with the R haplotype has been published, therefore while a European origin for the R haplotype fits the current published data, human mediated migration of cattle cannot be ruled out. For example, mtDNA variation within modern Portuguese cattle has been suggested could be due to movement of cattle during the Moorish occupation of the Southern Iberian peninsula (Cymbron et al. 1999). With the Roman Empire across Europe, the Near East and North Africa, the transportation of animals from a variety of geographical locations would have been plausible. As more ancient samples are sampled more new haplogroups representative of lost diversity, such as Kir4 and C (Zhang et al. 2013), will emerge. Perhaps the origins of R will be uncovered through aDNA.

4.5.2.2. Timing the Divergence of Bos Haplogroups

The timing of the divergence of the cattle mtDNA haplogroups via a Bayesian framework has been attempted before using partial sequences (Edwards et al. 2007; Finlay et al. 2007; Ho et al. 2008; Lari et al. 2011) and complete mitogenomes (Massilani et al. 2016), however it has never been attempted with this many ancient genomes; 26 genomes when only radiocarbon dated genomes are included and 72 samples when some non radiocarbon dated samples were included. Differences in the divergence dates of the haplogroups across the two analysis are likely due to the incorporation of non radiocarbon dated samples as dates given may not have been correct.

The 95% HPD dates (105,425-149,080 yrBP) for the divergence *B. indicus* and *B. taurus* encompasses dates previously calculated using BEAST for partitioned modern cattle mitogenomes (Massilani et al. 2016) of 130,000-185,000 yrBP and 84,000 – 219,000 yrBP for partial mtDNA sequences (Ho et al. 2008). While dating using autosomal variation suggested a separation of 280,000 yrBP (sdr 22Kyr) (Murray et al. 2010). However the date is much younger than the 335,000 yrB reported by Achilli and colleagues 2009 which was calculated using only modern DNA and a maximum likelihood approach and without the partitioning of the different coding regions of the mtDNA. As different regions and codon positions of the mitochondria mutate at different rates it is sensible to partition the data to ensure a more accurate result (Kainer & Lanfear 2015). As this date of the split predates domestication, it supports the long held view that the modern *Bos taurus* and *Bos indicus* were domesticated from diverse populations of auroch (Loftus et al. 1994) as supported and studies using autosomal variation (Murray et al. 2010).

To date, the C/R/Kir4 vs P/Q/T has no published split time. The split of P vs Q/T (95% HPD 27,475-39,255 yrBP) is within the range of 26,000-82,000 yrBP of the P vs T split given by Ho and colleagues 2008. This split of the traditional Northern European auroch haplogroup (P) and the domestic haplogroups of Q and T predates both domestication in the Near East and the Last Glacial Maximum in Europe, circa 20,000 yrBP (Hewitt 2000). The timing of this split is similar to the timing of a rapid decline of Beringian bison believed to have started to occur approximately 37,000 yrBP and correlated with to environmental changes linked to the onset of the LGM (Shapiro et al. 2004). It is possible that these changes in the environment caused this split in the wild population; perhaps via different refugia, where the ancestral Q/T population in the Near East and the P population elsewhere.

The 95% HPD for the TMRCA of the P haplogroup, 14,650-19,914 yrBP, is similar to past BEAST analysis using a 360 bp portion of the control region of mtDNA. Edwards and colleagues (2007) timed this to 10,050 – 32,320 yrBP (95% HPD). The split time in this analysis would correspond with a population range expansion that likely occurred with the retreating of the ice sheets post LGM.

The Q vs T split also predates domestication (95% HPD 15,478-23,685 yrBP). This split date again predates domestication, suggesting that some of the genetic variation incorporated into the Near Eastern domestic herds predated domestication. The timing of this split overlaps the LGM, during this period of time the Southwest Asia was covered in a steppe vegetation (grass/shrubland). However local variation in the pollen record does occur, with areas closer to the coast more amenable to tree growth than area such as the mountains of western Iran where signatures of arid steppe vegetation have been recorded (Elenga et al. 2000). Perhaps the split between Q and the macro haplogroup T was caused by environmental conditions during this period of time.

The TMRCA of the macro haplogroup T (95% HPD 9,182-12,565 yrBP) is similar to previous dates from Ho and colleagues 2008 (95% HPD 9,000-15,000 yrBP) and span the domestication of *Bos taurus,* which occurred approximately 11,000 yrBP in the Upper Euphrates valley (Helmer et al. 2005). Both the TMRCA between the ancient Q and ancient T2 with their associated haplogroup modern samples are post domestication, suggesting much of the genetic variation in the modern samples does not predate domestication. Therefore a large proportion of the starburst pattern of T can likely be appropriated to a bottleneck and rapid expansion event; comparable to a domestication event.

4.5.3. 200,000 year old *Bos primigenius*

The successful sequencing of a 200,000 year old sample is a rarity. At present aDNA has been successfully sequenced from only a handful of samples older and of similar age than VEM212. The oldest sample to be successfully sequence is a 560-780 kyrBP Equid persevered in the permafrost of Alaska (Orlando et al. 2013), followed by; hominins from the Sima de los Huesos cave from approximately 430 kyrBP (Meyer et al. 2016), a 300kyrBP cave bear (Dabney et al. 2013), forest community from the Greenland Ice core dated to 116-130 kyrBP (Willerslev et al. 2007) and a 110 -130 kyrBP bear sample from Svalbard (Miller et al. 2012). Unlike these published samples, VEM212 is not from an environment favorable to DNA preservation, such as permafrost or cave systems (Hofreiter et al. 2001). However both the endogenous level of DNA (0.2%) and the number of aligned reads are low. Nonetheless the damage patterns and read distribution are typical of ancient DNA. Two different approaches were taken to sequence mtDNA reads, shotgun and target capture, both producing aligned reads confirming the reads were not due to contamination during one of the sequencing runs.

When aligned to a reference genome of a Mesolithic British auroch, VEM212 sits basal to the taurine clades post the indicine split. While this sample is believed to predate the split time of taurine v indicine calculated by BEAST (4.3.2.2) the high amounts of missingness and damage in this sample means its placement on the *Bos* tree is not fully secure. However the high sequence divergence and the presence of previously described mutations away from the *B. taurus* reference mtDNA that predate the REPQT split (Achilli et al. 2009), advocates for the samples age and authentic aDNA. Future analysis will try to sequence more reads from this sample and attempt to align the sample using MIA – mapping-iterative-assembler – an ancient DNA aware reference assisted assembly program that has been used to assemble Neanderthal and early modern human mitochondria (Briggs et al. 2009; Krause et al. 2010).

4.5.4 Limitations and Future Analysis

Mitochondrial DNA as a matrilineal single locus has limitations, as it does not account for all the variation within a population. The lack of variation within domestic cattle, especially within the T3 haplogroup, means few conclusions about past cattle migration and demographic events can be drawn. However even with these limitations future mtDNA analysis should incorporate more radiocarbon dated samples into a Bayesian framework to help to narrow down split times. It would also be sensible to run this analysis minus the D-loop to ensure this highly mutable region does not confound the analysis. Furthermore, continued sequencing of the 200,000 year old auroch will help to inform our understanding of the Eurasian Pleistocene auroch population.

With the advancement in sequencing technology, the publication of the bovine nuclear genome, the sequencing and publication of the first auroch genome, as well as genome wide SNP typing platforms, there are many opportunities for the advancement of our understanding of cattle domestication (Bovine Genome Sequencing and Analysis Consortium et al. 2009; Park et al. 2015); (Bovine HapMap Consortium et al. 2009). Studies utilising modern cattle genomes have already demonstrated the superiority of autosomal data to that of mtDNA for understanding cattle genetics. For example phylogenetic structure of modern cattle breeds (Decker et al. 2014), as well admixture of domestic cattle with wild European auroch (Park et al. 2015; Upadhyay et al. 2016), have been inferred using autosomal data. Ancient nuclear genomes will have an important role to play to understand the process of domestication and migration of cattle prior to the formation of breeds.

4.5. Conclusion

This mtDNA analysis has demonstrated that the study of ancient mtDNA is warranted as new haplogroups are still emerging from ancient samples, providing insight to the diversity present in the wild population. The sequencing of mtDNA has allowed an insight into the genetic history of samples which have too poor endogenous DNA to sequence a nuclear genome from. Ancient mitogenomes from across Europe, the Near East, Africa confirm previous publications of modern and ancient D-loop sequences, showing the high haplogroup diversity from the Near East dwindling as domestic cattle moved into Europe. The African cattle advocate for ancient haplogroup diversity in Egypt as opposed to this being a modern phenomenon, and confirm the presence of T1 in Africa dates to at least the Egyptian Bronze Age. Bayesian analysis has shown how timings of some of the major haplogroups could have been linked to environmental conditions, with the dating of some of the branches occurring for the first time. This analysis suggests the formation of the macro haplogroup of T likely coincided with the beginnings of domestication and modern diversity has occurred since the Neolithic. Additionally, while the split from Q was prior to domestication, much of the modern diversity analysed here has also occurred since the Neolithic.

While there are limitations to mtDNA analysis, low endogenous DNA survival will likely be a challenge for the aDNA community for some time to come. In these situations the target capture and sequencing of mtDNA will still have a role to play in understanding past population events, especially as demonstrated here, more mtDNA haplogroups and past variation may still be waiting to be discovered.

4.6. References

Achilli, A. et al., 2008. Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Current biology: CB*, 18(4), pp.R157–8.

Achilli, A. et al., 2009. The multifaceted origin of taurine cattle reflected by the mitochondrial genome. *PloS one*, 4(6), p.e5753.

Allentoft, M.E. et al., 2015. Population genomics of Bronze Age Eurasia. *Nature*, 522(7555), pp.167–172.

Allentoft, M.E. 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), pp.4724–4733.

Anderson, S. et al., 1982. Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. *Journal of molecular biology*, 156(4), pp.683–717.

Anisimova, M. et al., 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic biology*, 60(5), pp.685–699.

Arbuckle, B.S., 2014. Pace and process in the emergence of animal husbandry in Neolithic Southwest Asia. *Bioarchaeol. Near East*, 8, pp.53–81.

Bailey, J.F. et al., 1996. Ancient DNA suggests a recent expansion of European cattle from a diverse wild progenitor species. *Proceedings. Biological sciences / The Royal Society*, 263(1376), pp.1467–1473.

Barnes, I. et al., 2002. Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science*, 295(5563), pp.2267–2270.

Beja-Pereira, A. 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), pp.8113–8118.

Bollongino, R. et al., 2006. Early history of European domestic cattle as revealed by ancient DNA. *Biology letters*, 2(1), pp.155–159.

Bollongino, R. et al., 2012. Modern taurine cattle descended from small number of near-eastern founders. *Molecular biology and evolution*, 29(9), pp.2101–2104.

Bonfiglio, S. et al., 2012. Origin and spread of Bos taurus: new clues from mitochondrial genomes belonging to haplogroup T1. *PloS one*, 7(6), p.e38601.

Bonfiglio, S. et al., 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), p.e15760.

Bovine Genome Sequencing and Analysis Consortium et al., 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, 324(5926), pp.522–528.

Bovine HapMap Consortium et al., 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science*, 324(5926), pp.528–532.

Bradley, D.G. et al., 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), pp.5131–5135.

Bridgland, D.R. et al., 2014. *Quaternary of the Trent*, Oxbow Books.

Briggs, A.W. et al., 2009. Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science*, 325(5938), pp.318–321.

Burbano, H.A. et al., 2010. Targeted investigation of the Neandertal genome by array-based sequence capture. *Science*, 328(5979), pp.723–725.

Carpenter, M.L. et al., 2013. Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *American journal of human genetics*, 93(5), pp.852–864.

Clutton-Brock, J., 1989. Five thousand years of livestock in Britain. *Biological journal of the Linnean Society. Linnean Society of London*, 38(1), pp.31–37.

Cooper, A. et al., 2001. Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature*, 409(6821), pp.704–707.

Cymbron, T. et al., 1999. Mitochondrial sequence variation suggests an African influence in Portuguese cattle. *Proceedings. Biological sciences / The Royal Society*, 266(1419), pp.597–603.

Dabney, J. et al., 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America*, 110(39), pp.15758–15763.

Dalén, L. et al., 2007. Ancient DNA reveals lack of postglacial habitat tracking in the arctic fox. *Proceedings of the National Academy of Sciences of the United States of America*, 104(16), pp.6726–6729.

Dalén, L. et al., 2005. Population history and genetic structure of a circumpolar species: the arctic fox. *Biological journal of the Linnean Society. Linnean Society of London*, 84(1), pp.79–89.

Decker, J.E. et al., 2014. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PloS genetics*, 10(3), p.e1004254.

Drummond, A.J. & Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, 7, p.214.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research, 32(5), pp.1792–1797.

Edwards, C.J. et al., 2010. A complete mitochondrial genome sequence from a mesolithic wild aurochs (Bos primigenius). *PloS one*, 5(2), p.e9255.

Edwards, C.J. et al., 2004. Ancient DNA analysis of 101 cattle remains: limits and prospects. *Journal of archaeological science*, 31(6), pp.695–710.

Edwards, C.J. et al., 2007. Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1616), pp.1377–1385.

Elenga, H. et al., 2000. Pollen-based biome reconstruction for southern Europe and Africa 18,000 yr bp. *Journal of biogeography*, 27(3), pp.621–634.

Finlay, E.K. et al., 2007. Bayesian inference of population expansions in domestic bovines. *Biology letters*, 3(4), pp.449–452.

Frantz, L.A.F. et al., 2016. Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science*, 352(6290), pp.1228–1231.

Fraser, S. et al., 2017. Matrilines in Neolithic cattle from Orkney, Scotland reveals complex husbandry patterns of ancestry. *Journal of Archaeological Science: Reports*, 14, pp.46–54.

Gamba, C. et al., 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nature communications*, 5, p.5257.

Gilbert, M. et al., 2007. Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *science.sciencemag.org*. Available at: http://science.sciencemag.org/content/317/5846/1927.short.

Gilbert, M.T.P. et al., 2008. Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science*, 320(5884), pp.1787–1789.

Gouy, M., Guindon, S. & Gascuel, O., 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution*, 27(2), pp.221–224.

Green, R.E. et al., 2008. A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell*, 134(3), pp.416–426.

Guindon, S. et al., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59(3), pp.307–321.

Haak, W. et al., 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. Available at: http://dx.doi.org/10.1038/nature14317.

Hansen, A. et al., 2001. Statistical evidence for miscoding lesions in ancient DNA templates. *Molecular biology and evolution*, 18(2), pp.262–265.

Helmer, D. et al., 2005. Identifying early domestic cattle from Pre-Pottery Neolithic sites on the Middle Euphrates using sexual dimorphism. In J. D. Vigne, J. Peter, & D. Helmer, eds. *The First Steps of Animal Domestication*. Oxbow Books, Oxford, pp. 86–95.

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789), pp.907–913.

Hofreiter, M., Serre, D., et al., 2001. Ancient DNA. *Nature reviews. Genetics*, 2(5), pp.353–359.

Hofreiter, M., Jaenicke, V., et al., 2001. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic acids research*, 29(23), pp.4793–4799.

Ho, S.Y.W. et al., 2008. Correlating Bayesian date estimates with climatic events and domestication using a bovine case study. *Biology letters*, 4(4), pp.370–374.

Ivankovic, A. et al., 2014. Mitochondrial DNA-based genetic evaluation of autochthonous cattle breeds in Croatia. *Czech journal of animal science = Zivocisna vyroba / Ustav zemedelskych a potravinarskych informaci*, 59, pp.519–528.

Kainer, D. & Lanfear, R., 2015. The effects of partitioning on phylogenetic inference. *Molecular biology and evolution*, 32(6), pp.1611–1627.

Krause, J. et al., 2010. A complete mtDNA genome of an early modern human from Kostenki, Russia. *Current biology: CB*, 20(3), pp.231–236.

Lanfear, R. et al., 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular biology and evolution*, 29(6), pp.1695–1701.

Lari, M. et al., 2011. The complete mitochondrial genome of an 11,450-year-old aurochsen (Bos primigenius) from Central Italy. *BMC evolutionary biology*, 11, p.32.

Larson, G. et al., 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science*, 307(5715), pp.1618–1621.

Lefort, V., Longueville, J.-E. & Gascuel, O., 2017. SMS: Smart Model Selection in PhyML. *Molecular biology and evolution*, p.msx149.

Lenstra, J.A. et al., 2014. Meta-Analysis of Mitochondrial DNA Reveals Several Population Bottlenecks during Worldwide Migrations of Cattle. *Diversity*, 6(1), pp.178–187.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , 25(16), pp.2078–2079.

Loftus, R.T., MacHugh, D.E., Bradley, D.G., et al., 1994. Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences of the United States of America*, 91(7), pp.2757–2761.

Loftus, R.T., MacHugh, D.E., Ngere, L.O., et al., 1994. Mitochondrial genetic variation in European, African and Indian cattle populations. *Animal genetics*, 25(4), pp.265–271.

Magee, D.A. et al., 2002. A partial african ancestry for the creole cattle populations of the Caribbean. *The Journal of heredity*, 93(6), pp.429–432.

Mannen, H. et al., 1998. Mitochondrial DNA variation and evolution of Japanese black cattle (Bos taurus). *Genetics*, 150(3), pp.1169–1175.

Massilani, D. et al., 2016. Past climate changes, population dynamics and the origin of Bison in Europe. *BMC biology*, 14(1), p.93.

McKenna, A. et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), pp.1297–1303.

Meyer, M. et al., 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science*, 338(6104), pp.222–226.

Meyer, M. et al., 2016. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature*. Available at: http://dx.doi.org/10.1038/nature17405.

Miller, F.J. et al., 2003. Precise determination of mitochondrial DNA copy number in human skeletal and cardiac muscle by a PCR-based assay: lack of change of copy number with age. *Nucleic acids research*, 31(11), pp.e61–e61.

Miller, W. 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), pp.E2382–90.

Murray, C. et al., 2010. Cattle demographic history modelled from autosomal sequence variation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1552), pp.2531–2539.

Okonechnikov, K., Conesa, A. & García-Alcalde, F., 2015. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* . Available at: http://dx.doi.org/10.1093/bioinformatics/btv566.

Olivieri, A. 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), p.e0141170.

Orlando, L. et al., 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, 499(7456), pp.74–78.

Paijmans, J.L.A., Gilbert, M.T.P. & Hofreiter, M., 2013. Mitogenomic analyses from ancient DNA. *Molecular phylogenetics and evolution*, 69(2), pp.404–416.

Park, S. et al., 2015. Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and evolution of cattle. *Genome Biology*, 16(1), p.234. Available at: http://genomebiology.com/2015/16/1/234.

Poinar, H.N. et al., 2006. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science*, 311(5759), pp.392–394.

Rambaut, A., 2007. FigTree, a graphical viewer of phylogenetic trees. *See http://tree. Bio. Ed. Ac. Uk/software/figtree*.

Rambaut A, Suchard MA, Xie D, & Drummond AJ, 2014. *Tracer*, Available at: http://tree.bio.ed.ac.uk/software/tracer/.

Rasmussen, M. et al., 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*, 463(7282), pp.757–762.

Reich, D. et al., 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, 468(7327), pp.1053–1060.

Rogaev, E.I. et al., 2006. Complete Mitochondrial Genome and Phylogeny of Pleistocene Mammoth Mammuthus primigenius. *PloS biology*, 4(3), p.e73.

Rohland, N. & Reich, D., 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome research*, 22(5), pp.939–946.

Scheu, A. et al., 2008. Ancient DNA provides no evidence for independent domestication of cattle in Mesolithic Rosenhof, Northern Germany. *Journal of archaeological science*, 35(5), pp.1257–1264.

Scheu, A. et al., 2015. The genetic prehistory of domesticated cattle from their origin to the spread across Europe. *BMC genetics*, 16, p.54.

Schibler, J. et al., 2014. Incorporation of aurochs into a cattle herd in Neolithic Europe: single event or breeding? *ncbi.nlm.nih.gov*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4107343/.

Schubert, M. et al., 2014. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proceedings of the National Academy of Sciences of the United States of America*. Available at: http://dx.doi.org/10.1073/pnas.1416991111.

Seregély, T., 2014. Human and animal remains from three eras. New documentation methods of a vertical cave of the Northern Franconian low mountain range and its inventory. *The European Archaeologist*, (41), pp.17–19.

Shapiro, B. et al., 2004. Rise and fall of the Beringian steppe bison. *Science*, 306(5701), pp.1561–1565.

Stock, F. et al., 2009. Cytochrome b sequences of ancient cattle and wild ox support phylogenetic complexity in the ancient and modern bovine populations. *Animal genetics*, 40(5), pp.694–700.

Teasdale, M.D. et al., 2017. The York Gospels: a 1000-year biological palimpsest. *Royal Society open science*, 4(10), p.170988.

Troy, C.S. et al., 2001. Genetic evidence for Near-Eastern origins of European cattle. *Nature*, 410(6832), pp.1088–1091.

Upadhyay, M.R. et al., 2016. Genetic origin, admixture and population history of aurochs (Bos primigenius) and primitive European cattle. *Heredity*. Available at: http://dx.doi.org/10.1038/hdy.2016.79.

Van Vuure, C., 2005. *Retracing the aurochs: history, morphology and ecology of an extinct wild ox*, Pensoft Pub.

Vilstrup, J.T. et al., 2013. Mitochondrial phylogenomics of modern and ancient equids. *PloS one*, 8(2), p.e55950.

Willerslev, E. et al., 2009. Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution. *BMC evolutionary biology*, 9, p.95.

Willerslev, E. et al., 2007. Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science*, 317(5834), pp.111–114.

Wright, E. & Viner-Daniels, S., 2015. Geographical variation in the size and shape of the European aurochs (Bos primigenius). *Journal of archaeological science*, 54, pp.8–22.

Zeyland, J. et al., 2013. Complete mitochondrial genome of wild aurochs (Bos primigenius) reconstructed from ancient DNA. *Polish journal of veterinary sciences*, 16(2), pp.265–273.

Zhang, H. et al., 2013. Morphological and genetic evidence for early Holocene cattle management in northeastern China. *Nature communications*, 4, p.2755.

Chapter 5. Whole Genome Analysis of Ancient Near Eastern and European *Bos* Samples

5.1. Introduction

Archaeozoological and genetic studies indicate the domestication of taurine cattle occurred in the Upper and Middle Euphrates Valley of the Near East approximately 9,000 BC from the wild progenitor *Bos primigenius,* auroch*,* (Troy et al. 2001; Helmer et al. 2005; Bollongino et al. 2006; Edwards et al. 2007; Achilli et al. 2008; Arbuckle 2014). Likely domesticated from a small effective number of female aurochs, subsequent migration from South Anatolia through to Western Anatolia and the Aegean occurred after 7,000 BC. From there an introduction in Europe around 6,400 BC, eventually reaching Northern Europe 4,100 BC (Bollongino et al. 2012; Scheu et al. 2015). Over time, selection for a variety of phenotypic traits has resulted in a diverse array of breeds adapted to different environmental conditions and with different production attributes.

Until relatively recently the understanding of cattle population histories has been largely based upon modern and ancient mitochondrial and Y chromosome studies (Loftus et al. 1994; Bradley et al. 1996; Mannen et al. 1998; Troy et al. 2001; Bollongino et al. 2006; Edwards et al. 2007; Achilli et al. 2009; Speller et al. 2013; Scheu et al. 2015; Niemi et al. 2015). A number of developments such as: advancements in sequencing technology, the publication of the bovine nuclear genome (Bovine Genome Sequencing and Analysis Consortium et al. 2009), genome wide SNP typing platforms (Bovine HapMap Consortium et al. 2009) and the ongoing sequencing of 1000 Bulls genomes (Daetwyler et al. 2014; Chung et al. 2017), have enabled population genetic studies to focus on genome-wide variation as opposed to single maternal or paternal loci. Such studies include analysing past demographic events (MacLeod et al. 2013), patterns of admixture (Decker et al. 2014; Upadhyay et al. 2016; Park et al. 2015) and phylogenetic structure of different taurine and indicine populations (Decker et al. 2014; Chung et al. 2017).

Utilising modern breeds to project back into the past to understand domestication has limitations. Historic population movement and the formation of breeds may confound projections to the past when trying to understand domestication from modern samples. For example the historical movement of merino rams across Europe, the movement of British sheep to Australia and African sheep to the Caribbean all would complicate an analysis looking to understand sheep domestication on the basis of modern sheep breeds (Kijas et al. 2012). In a recent aDNA study of Roman British humans excavated in York, it was noted that the ancient samples were significantly divergent from the modern inhabitants of area (Martiniano et al. 2016). To understand cattle domestication, and subsequent migration, ancient samples from a variety of geographical locations and time periods are necessary.

The publishing of the first ancient bovid genome, a British Mesolithic auroch, enabled the study of introgression from European auroch into domestic stock (Park et al. 2015). No consensus on auroch introgression in European cattle had been reached using only maternal or paternal loci (Gotherstrom et al. 2005; Beja-Pereira et al. 2006; Bollongino et al. 2008; Edwards et al. 2007; Achilli et al. 2008), however using the auroch genome in conjunction with genome-wide SNP data of modern breeds indicated European auroch introgression did occur (Park et al. 2015; Upadhyay et al. 2016). The addition of further ancient genomes should enable us to explore the temporal and geographical detail of this introgression.

This study presents 113 ancient cattle genomes from Anatolia to Ireland, spanning from the Neolithic to Medieval period. 42 Neolithic genomes spread geographically across Europe enables the study of the migration of cattle and interaction with European aurochs from Anatolia right across Northern Europe to Britain and Ireland. In particular 77 genomes create a time series at the North Atlantic Edge, from the Neolithic through to the Medieval period, consisting of samples from Britain, Ireland and the Netherlands. Past human migrations have had a significant effect on the genetic diversity within Europe (Skoglund et al. 2012; Gamba et al. 2014; Haak et al. 2015; Jones et al. 2015; Cassidy et al. 2016; Martiniano et al. 2016). However until now, it has been unknown if these human migrations had a genetic effect on the livestock that these human populations relied upon. The aim of this study was twofold; i) to gain an insight into the genetic structure present in Neolithic Europe, ii) examine how this genetic structure varied through time, in particular on the Atlantic Edge.

5.2. Materials and Methods

5.2.1. Materials

The majority of cattle samples presented in this chapter are sampled from the petrous portion of the temporal bone. The samples were screened as detailed in the overall methods and results presented in Chapter 3. In total 64 bones from Britain and Ireland were chosen for further sequencing (Fig 5.1 & Table 5.1). This chapter is supplemented with another 49 samples from Eurasia, processed in the laboratory and for whole genome alignment (2.2.2) by Marta Verdugo and Amelie Schue (Table 5.2).

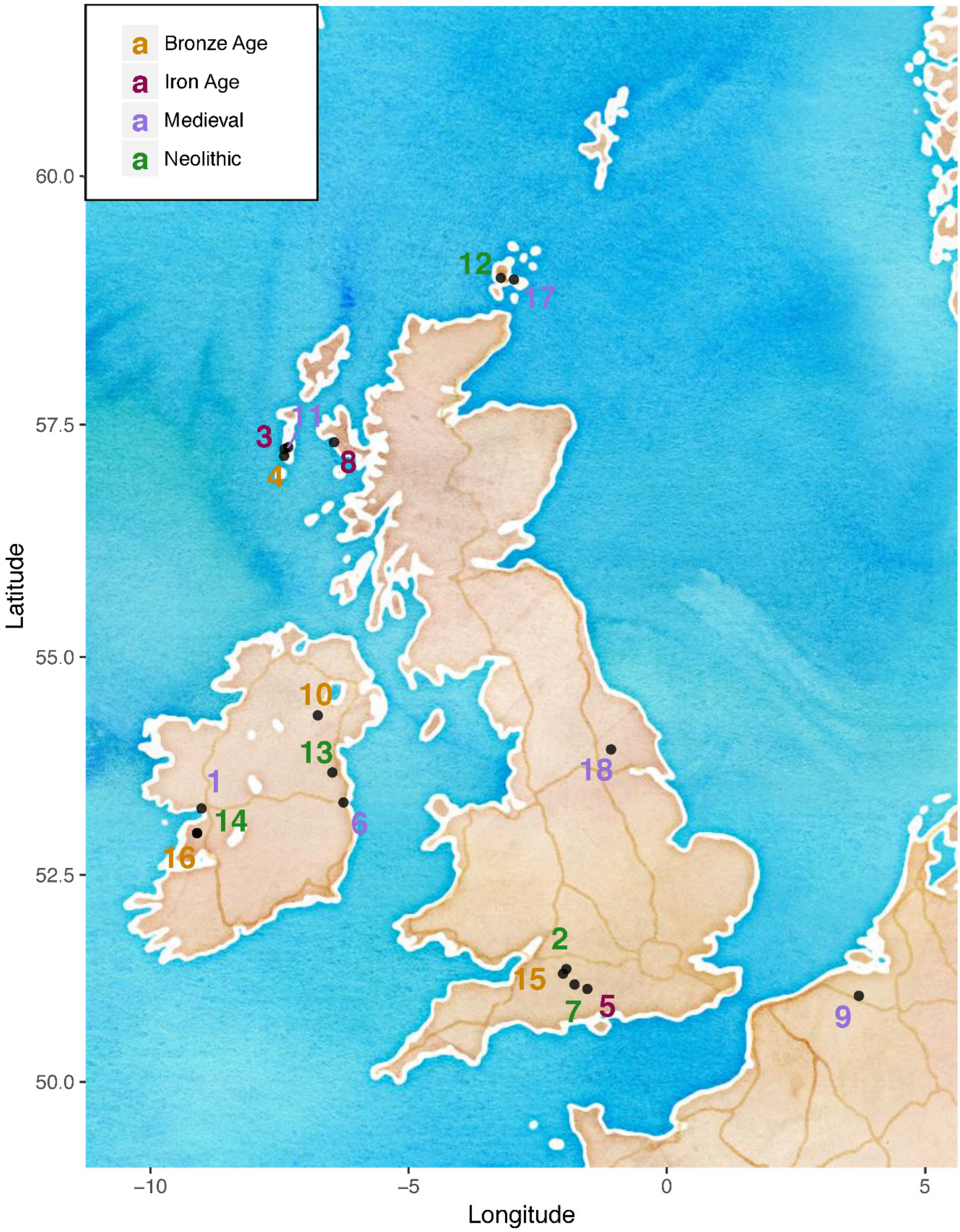


Fig 5.1. *Map of sites sampled for this chapter, predominantly from Britain and Ireland. Sites are coloured by the predominant cultural period of the site. For site number see Appendix Table 7.*

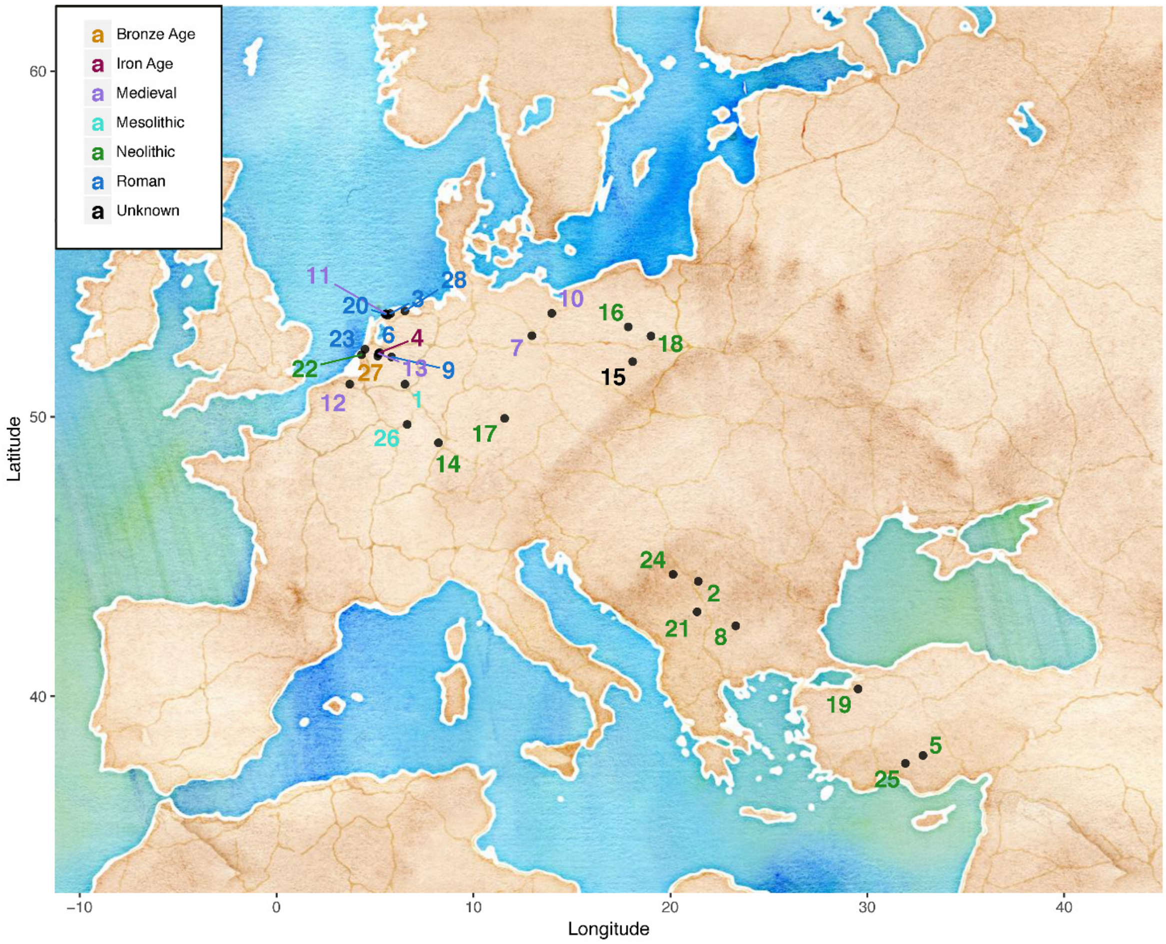


Fig 5.2. *Map of sites sampled by MV and AS. Sites are coloured by the predominant cultural period of the site. For site number see Appendx Table 8.*

Table 5.1. *Samples predominantly from Britain and Ireland processed by Victoria Mullin and analysed in this chapter.*

Name	Cultural Period	Radiocarbon Date calBC (2 Sigma)	Site	Country
Bis1	Neolithic	3529-3362BC	Bishops Cannings, Wiltshire	England
Bis2	Neolithic	3638-3512BC	Bishops Cannings, Wiltshire	England
Dur1	Late Neolithic	2410-2199BC	Durrington Walls, Wiltshire	England
Dur2	Late Neolithic	2695-2457BC	Durrington Walls, Wiltshire	England
Dur3	Late Neolithic		Durrington Walls, Wiltshire	England
Pot1	Late Bronze Age		Potterne, Wiltshire	England
Pot2	Late Bronze Age		Potterne, Wiltshire	England
Pot3	Late Bronze Age		Potterne, Wiltshire	England
Pot5	Late Bronze Age		Potterne, Wiltshire	England
Pot6	Late Bronze Age		Potterne, Wiltshire	England
Pot4	Early Iron	769-472BC	Potterne, Wiltshire	England
Da1	Iron Age		Danebury, Hampshire	England
Da2	Iron Age		Danebury, Hampshire	England
Da3	Iron Age	404-354BC	Danebury,Hampshire	England
Da4	Iron Age		Danebury, Hampshire	England
Da5	Iron Age	511-357BC	Danebury,Hampshire	England
Da6	Iron Age	203-43BC	Danebury, Hampshire	England
Yor1	Roman	2 BC- 135 AD	Tanner Row, York	England
Yor11	Roman		Tanner Row, York	England
Yor10	Medieval		Hungate, York	England
Yor12	Medieval		Hungate, York	England
Yor2	Medieval		Hungate, York	England
Yor3	Medieval		Coppergate, York	England
Yor4	Medieval		Coppergate, York	England
Yor5	Medieval		Hungate,York	England
Yor6	Medieval		Hungate, York	England
Yor7	Medieval		Coppergate, York	England
Yor8	Medieval		Hungate, York	England
Yor9	Medieval		Fishergate, York	England
New4	Neolithic		Newgrange, Co Meath	Ireland
New5	Neolithic	3024-2894 BC	Newgrange, Co Meath	Ireland
Par1	Neolithic		Parknabinia, Co Clare	Ireland
Par2	Neolithic		Parknabinia, Co Clare	Ireland
New1	Neolithic	2795-2580 BC	Newgrange, Co Meath	Ireland
New3	Neolithic		Newgrange, Co Meath	Ireland
Rou1	Bonze Age		Roughan Hill, Co Clare	Ireland
Rou2	Bonze Age		Roughan Hill, Co Clare	Ireland
Rou3	Bonze Age		Roughan Hill, Co Clare	Ireland
Rou4	Bonze Age		Roughan Hill, Co Claire	Ireland
New2	Late Iron Age	321-425 AD	Newgrange, Co Meath	Ireland
Bal1	Medieval		Ballybane, Co Galway	Ireland
Dub2	Medieval		Dublin	Ireland
Bal2	Medieval		Ballybane, Co Galway	Ireland
Dub1	Medieval		Dublin	Ireland
HF1	Bonze Age		Haughey's Fort, Co Armagh	N. Ireland
Ness2	Late Neolithic		Ness of Brodgar, Orkney	Scotland
Ness3	Late Neolithic		Ness of Brodgar, Orkney	Scotland
Ness4	Late Neolithic		Ness of Brodgar, Orkney	Scotland
Ness5	Late Neolithic	2588-2464BC	Ness of Brodgar, Orkney	Scotland
Cla1	Bronze Age	1054-826 BC	Cladh Hallan, South Uist	Scotland
Cla2	Bronze Age		Cladh Hallan, South Uist	Scotland
Cla3	Bronze Age		Cladh Hallan, South Uist	Scotland
Cla4	Bronze Age		Cladh Hallan, South Uist	Scotland
Cla5	Bronze Age		Cladh Hallan, South Uist	Scotland
Cla7	Bronze Age		Cladh Hallan, South Uist	Scotland
Cla8	Bronze Age	1060-899BC	Cladh Hallan, South Uist	Scotland
Cla9	Bronze Age	1118-924BC	Cladh Hallan, South Uist	Scotland
Bor1	Iron Age		Bornais, South Uist	Scotland
Cla6	Iron Age		Cladh Hallan, South Uist	Scotland
Fis1	Iron Age		Fiskavaig, Isle of Skye	Scotland
Fis2	Iron Age		Fiskavaig, Isle of Skye	Scotland
Kil1	Medieval (Norse)		Kilpheder, South Uist	Scotland
Sir1	Medieval (Norse)	1039-1215AD	Snusgar, Orkney	Scotland
Gen1	Medieval		Gent	Belgium

Table 5.2. *Samples from Eurasia analysed in this chapter processed by Marta Verdugo and Amelie Scheu (per. comm.).*

Name	Cultural Period	Radiocarbon Date calBC (2 Sigma)	Site	Country	Processed by
Ch22	Neolithic		Catal Hoyuk	Turkey	AS
Men1	Neolithic		Mentese	Turkey	MV
Men2	Neolithic	6048-5893 BC	Mentese	Turkey	MV
Sub1	Neolithic	6221-6024 BC	Suberde	Turkey	MV
Dzh1	Early Neolithic		Dzulhunitsa	Bulgaria	AS
Dzh2	Early Neolithic		Dzulhunitsa	Bulgaria	AS
Bel1	Neolithic		Belovode- Veliko	Serbia	MV
Bel2	Neolithic		Belovode- Veliko	Serbia	MV
Bub1	Neolithic		Bubanj	Serbia	MV
Plo1	Neolithic		Plocnik	Serbia	MV
Plo2	Neolithic		Plocnik	Serbia	MV
Plo3	Neolithic		Plocnik	Serbia	MV
Plo4	Neolithic		Plocnik	Serbia	MV
Plo5	Neolithic		Plocnik	Serbia	MV
Plo6	Neolithic		Plocnik	Serbia	MV
Plo7	Neolithic		Plocnik	Serbia	MV
Plo8	Neolithic		Plocnik	Serbia	MV
Stu1	Neolithic		Stubline	Serbia	MV
Lud1	Neolithic		Ludwinowo	Poland	MV
Lud2	Neolithic		Ludwinowo	Poland	MV
Lud3	Neolithic		Ludwinowo	Poland	MV
Lud4	Neolithic		Ludwinowo	Poland	MV
Kie1	Neolithic		Kierzkowo	Poland	MV
Kie2	Neolithic		Kierzkowo	Poland	MV
Bed3	Mesolithic	9805-9385 BC	Bedburg-Königshoven	Germany	AS
Bed4	Mesolithic	10036-9647 BC	Bedburg-Königshoven	Germany	AS
Tri1	Mesolithic	8261-7960 BC	Trier-Brüderkrankenhaus	Germany	AS
Kalisz2	no info		Kalisz	Poland	MV
Kalisz3	no info		Kalisz	Poland	MV
Hxh1	Neolithic		Herxheim	Germany	AS
Hxh2	Neolithic		Herxheim	Germany	AS
Kir18	Neolithic	3798-3641 BC	Kirschbaumhöhle	Germany	AS
Kir4	Neolithic	2816-2671 BC	Kirschbaumhöhle	Germany	MV/VEM
Fal7	Medieval		Falkenwalde	Germany	AS
Dyr1	Medieval (late slavic)		Dyrotz 25	Germany	AS
Sch1	Neolithic		Schipluiden	Netherlands	MV
Viz1	Bronze Age		Vianen-Zijdervel	Netherlands	MV
Viz2	Bronze Age		Vianen-Zijdervel	Netherlands	MV
Bun2	Iron Age		Bunnik	Netherlands	MV
Bun1	Iron Age/ Roman		Bunnik	Netherlands	MV
Ste1	Iron Age/ Roman		Stevenshofjesp	Netherlands	MV
Bri1	Roman		Britsum	Netherlands	MV
Dro1	Roman		Dronrijp-zuid	Netherlands	MV
Els1	Roman		Elst	Netherlands	MV
Otb1	Roman		Oosterbeintum	Netherlands	MV
Win1	Roman	417-561 AD	Winsum	Netherlands	MV
Fir1	Medieval		Firdgum	Netherlands	MV
Hou1	Medieval		Houten-Loerik	Netherlands	MV
Hou2	Medieval		Houten-Hoogdij	Netherlands	MV

5.2.2. Methods

5.2.2.1. Read Processing

Illumina HiSeq^{TM} reads were processed as described in section 2.2.2.3.

5.2.2.2. SNP calling and Modern Dataset Merge

To analyse the samples in the context of modern variation, the SNPs from the curated 770K Illumina Bovine HD SNP chip dataset (section 2.2.2.4.1) were called in all the samples (section 2.2.2.5.1). Ancient samples were then merged with the modern dataset, previously filtered for a minor allele frequency of 1%, and a 1% SNP missingness using PLINK (Purcell et al. 2007).

5.2.2.3. Population genetic Analysis

5.2.2.3.1. PCA

In order to visualise autosomal variation among the ancient cattle genomes in the context of the modern breeds a LASER PCA (Wang et al. 2015) was performed and PCs 1-2 plotted using ggplot2 in R (Wickham 2009; R Core Team 2015).

5.2.2.3.2. Ancestry Estimation

Prior to estimating ancestry fractions the dataset was filtered for linkage disequilibrium using PLINK option *--indep-pairwise 50 5 0.5*, leaving 253,502 SNPs. Ancestry coefficients of K=1-6 were computed using ADMIXTURE (Alexander et al. 2009) (section 2.2.3.2). Each value of K was run ten times. For each K the replicate with the lowest likelihood was chosen to be visualized in R (R Core Team 2015).

5.2.2.3.3. Testing for Admixture and Clade Integrity

For estimating possible admixture three different statistics from the AdmixTools (Patterson et al. 2012a), the *D*-statistic, f_3-statistic and f_4-ratio, were implemented (sections 2.2.3.3, 2.2.3.4.1 & 2.2.3.4.3). Results visualized using ggplot2 in R (Wickham 2009; R Core Team 2015), as well as compiled in table format.

5.2.2.3.3.1. D-statistics

The outgroup for the *D*-statistics was the Gaur, a bovine from the Indian Subcontinent, which has previously been genotyped on the Illumina Bovine HD SNP chip. Tests were performed with modern samples left diploid and for some tests where possible ancient samples were grouped into temporal and geographical populations (Appendix Table 10). Tests were performed using the general formula (Ref1,Ref2),X,Gaur) where X is the introgressor and Ref1 and Ref2 represent ancient and modern samples or populations.

The *D*-statistic was also employed to test for clade integrity between ancient samples. Results were compiled in table format. Tests were performed using the general formula (Ref1,Ref2),X,Gaur) where Ref1 and Ref2 represent the calde being tested and X represents the sample or population attempting to break the clade.

5.2.2.3.3.2. f_3-statistics

Modern samples were left diploid and where possible ancient samples were grouped into populations (Appendix Table 10). The inbreed:YES option was applied when the target population was a group of pseudo-diplodised population. Tests were performed using the general formula (C;A,B) where C is the target, while A and B represent the admixing populations.

5.2.2.3.3.3. f_4-ratio

To estimate the size of ancestry proportions from an admixture event the f_4-ratio statistic was calculated. The test was set to calculate introgression from the previously published British Mesolithic Auroch CPC (Park et al. 2001) into the test sample X. The test was set up as below:

$$f4(OGR,CPC;X,SUB1)/f4(OGR,CPC;BED3,SUB1)$$

Results were plotted on a map using QGIS opensource software (QGIS Development Team 2009) (John Lancsater *per. comm.*)

5.2.2.3.4. Shared Drift

To test for shared drift between samples the outgroup f_3-statistic was performed in a pairwise fashion using AdmixTools (Patterson et al. 2012a) (section 2.2.3.4.2). The indicine breed Gir was utilised as the outgroup for this test. Tests were performed in a pairwise fashion using the general formula (C;A,B), where C represents the outgroup of the test, while A and B represent the two test populations.

5.2.2.4. Runs of Homozygosity

To analyse the recent demography of the ancient samples an analyses of runs of homozygosity was performed on the high coverage samples. Samples with a minimum coverage of 8x were subject to diploid SNP calling (section 2.2.2.5.2) for the SNPs discovered by Run 6 of the 1K Bulls project (section 2.2.2.4.2). Samples were merged in PLINK (Purcell et al. 2007) with the 1K Bulls modern dataset, which was previously filtered for a minimum minor allele frequency of 1% and no missingness within SNPs, leaving 915,542 SNPs. Runs of homozygosity were then calculated in PLINK (Purcell et al. 2007) using the following options similar to previously published analysis (Purfield et al. 2012; Gamba et al. 2014); --homozyg-window-kb 5000, --homozyg-window-snp 50, --homozyg-window-het 1, --homozyg-window-missing 2, --homozyg-window-threshold 0.05, --homozyg-snp 50, --homozyg-kb 500, --homozyg-density 50, --homozyg-gap 100. The results were then binned into 7 different ROH length categories and the mean total length within these categories calculated using an in house script (E. Jones *per. comm.*). Results were then visualized using R (R Core Team 2015).

5.2.2.5. Allele Presence - DGAT1

To examine allele presence for the missense mutations in DGAT1 - Diglyceride acyltransferase (rs109234250, rs109326954), GATK pileup (McKenna et al. 2010) was used to call the positions from BAM files filtered for removal of duplicates and a mapping quality of 25. Using a custom script, calls were then filtered for a minimum base quality of 20 and the alleles present noted.

5.3. Results

5.3.1. Overview

The average genome coverage of the cattle analysed in this study ranges from 0.01x to 22.84x (Table 5.3 and 5.4).

Table 5.3. *Sample information post Hiseq sequencing for samples predominantly from Britain and Ireland sampled and processed by Victoria Mullin.*

Name	Cultural Period	Radiocarbon Date calBC (2 Sigma)	Site	Country	Coverage	USER	Sex
Bis2	Neolithic	3638-3512BC	Bishops Cannings	England	2.78	Yes	M
Bis1	Neolithic	3529-3362BC	Bishops Cannings	England	3.07	Yes	M
Dur3	Late Neolithic		Durrington Walls	England	0.38	Yes	F
Dur1	Late Neolithic	2410-2199BC	Durrington Walls	England	2.48	Yes	F
Dur2	Late Neolithic	2695-2457BC	Durrington Walls	England	2.79	Yes	F
Pot2	Late Bronze Age		Potterne	England	0.15	Yes	F
Pot3	Late Bronze Age		Potterne	England	0.32	Yes	M
Pot5	Late Bronze Age		Potterne	England	0.34	Yes	F
Pot1	Late Bronze Age		Potterne	England	2.60	Yes	F
Pot6	Late Bronze Age		Potterne	England	0.35	Yes	M
Pot4	Early Iron Age	769-472BC	Potterne	England	2.68	Yes	F
Da2	Iron Age		Danebury	England	0.21	Yes	M
Da4	Iron Age		Danebury	England	0.31	Yes	F
Da3	Iron Age	404-354BC	Danebury	England	2.56	Yes	F
Da5	Iron Age	511-357BC	Danebury	England	2.87	Yes	F
Da6	Iron Age	203-43BC	Danebury	England	3.51	Yes	F
Da1	Iron Age		Danebury	England	3.51	Yes	F
Yor11	Roman		York	England	0.17	No	F
Yor1	Roman	2 BC- 135 AD	York	England	8.00	Yes	F
Yor8	Medieval		York	England	0.13	No	M
Yor7	Medieval		Coppergate	England	0.17	No	M
Yor12	Medieval		York	England	0.18	No	M
Yor10	Medieval		York	England	0.25	No	F
Yor9	Medieval		York	England	0.26	No	F
Yor5	Medieval		York	England	1.17	No	M
Yor6	Medieval		York	England	2.19	No	F
Yor3	Medieval		Coppergate	England	2.54	Yes	M
Yor4	Medieval		Coppergate	England	3.19	Yes	F
Yor2	Medieval		York	England	3.97	Yes	M
New4	Neolithic		Newgrange	Ireland	4.31	Yes	F
New5	Neolithic	3024-2894 BC	Newgrange	Ireland	4.78	Yes	M
New3	Neolithic		Newgrange	Ireland	5.29	Yes	M
New1	Neolithic	2795-2580 BC	Newgrange	Ireland	16.62	Yes	M
Par1	Neolithic		Parknabinia	Ireland	0.26	Yes	F
Par2	Neolithic		Parknabinia	Ireland	0.38	Yes	M

Rou4	Bronze		Roughan Hill	Ireland	0.08	Yes	F
Rou3	Bronze Age		Roughan Hill	Ireland	0.27	Yes	F
Rou1	Bronze Age		Roughan Hill	Ireland	2.98	Yes	F
Rou2	Bronze Age		Roughan Hill	Ireland	3.03	Yes	F
New2	Late Iron Age	321-425 AD	Newgrange	Ireland	3.99	Yes	M
Dub2	Medieval		Dublin	Ireland	0.19	No	M
Dub1	Medieval		Dublin	Ireland	2.13	Yes	F
Bal1	Medieval		Ballybane	Ireland	0.50	Yes	F
Bal2	Medieval		Ballybane	Ireland	0.90	Yes	F
HF1	Bonze Age		Haughey's Fort	N. Ireland	0.17	No	F
Ness4	Late Neolithic		Ness of Brodgar	Scotland	0.35	Yes	F
Ness3	Late Neolithic		Ness of Brodgar	Scotland	0.41	Yes	F
Ness2	Late Neolithic		Ness of Brodgar	Scotland	2.83	Yes	F
Ness5	Late Neolithic	2588-2464BC	Ness of Brodgar	Scotland	3.15	Yes	M
Cla3	Bronze Age		Cladh Hallan	Scotland	0.15	Yes	F
Cla7	Bronze Age		Cladh Hallan	Scotland	0.30	Yes	F
Cla4	Bronze Age		Cladh Hallan	Scotland	0.33	Yes	F
Cla2	Bronze Age		Cladh Hallan	Scotland	1.89	Yes	M
Cla9	Bronze Age	1118-924BC	Cladh Hallan	Scotland	2.98	Yes	F
Cla8	Bronze Age	1060-899BC	Cladh Hallan	Scotland	3.15	Yes	F
Cla5	Bronze Age		Cladh Hallan	Scotland	3.39	Yes	F
Cla1	Bronze Age	1054-826 BC	Cladh Hallan	Scotland	15.51	Yes	F
Cla6	Iron Age		Cladh Hallan	Scotland	0.36	Yes	F
Bor1	Iron Age		Bornais	Scotland	0.49	Yes	M
Fis2	Iron Age		Fiskavaig	Scotland	2.74	Yes	F
Fis1	Iron Age		Fiskavaig	Scotland	2.97	Yes	F
Kil1	Medieval (Norse)		Kilpheder	Scotland	0.28	Yes	F
Sir1	Medieval (Norse)	1039-1215AD	Snusgar, Orkney	Scotland	3.48	Yes	F
Gen1	Medieval		Gent	Belgium	0.26	Yes	F

Table 5.4. *Sample information post Hiseq sequencing for samples from Eurasia, sampled and processed by Marta Verdugo and Amelie Scheu. Names in bold are European B. primigenius samples, while * denotes uncertainty as to domestic or wild sample.*

Name	Cultural Period	Radiocarbon Date calBC/AD (2 Sigma)	Site	Country	Coverage	USER	Sex
Sub1	Neolithic	6221-6024 BC	Suberde	Turkey	13.50	Yes	M
Men2	Neolithic	6048-5893 BC	Mentese	Turkey	1.88	Yes	F
Men1	Neolithic		Mentese	Turkey	2.97	Yes	M
Ch22*	Neolithic		Catal Hoyuk	Turkey	0.31	Yes	M
Dzh1	Neolithic		Dzulhunitsa	Bulgaria	9.83	Yes	M
Dzh2	Neolithic		Dzulhunitsa	Bulgaria	4.17	Yes	F
Bel1	Neolithic		Belovode- Veliko	Serbia	0.09	Yes	F
Bel2	Neolithic		Belovode- Veliko	Serbia	0.13	Yes	M
Bub1	Neolithic		Bubanj	Serbia	0.43	No	F
Plo1	Neolithic		Plocnik	Serbia	0.13	Yes	F
Plo2	Neolithic		Plocnik	Serbia	0.35	Yes	F
Plo3	Neolithic		Plocnik	Serbia	2.66	Yes	M

Plo4	Neolithic		Plocnik	Serbia	2.74	Yes	M
Plo5	Neolithic		Plocnik	Serbia	0.08	Yes	M
Plo6	Neolithic		Plocnik	Serbia	0.12	Yes	M
Plo7	Neolithic		Plocnik	Serbia	0.11	Yes	F
Plo8	Neolithic		Plocnik	Serbia	0.20	Yes	F
Stu1	Neolithic		Stubline	Serbia	0.15	Yes	M
Lud1	Neolithic		Ludwinowo	Poland	0.15	Yes	F
Lud2	Neolithic		Ludwinowo	Poland	0.36	Yes	F
Lud3	Neolithic		Ludwinowo	Poland	0.29	Yes	M
Lud4	Neolithic		Ludwinowo	Poland	0.19	Yes	F
Kie1	Neolithic		Kierzkowo	Poland	0.04	Yes	M
Kie2	Neolithic		Kierzkowo	Poland	0.21	Yes	M
Kalisz2	no info		Kalisz	Poland	2.70	Yes	M
Kalisz3	no info		Kalisz	Poland	0.01	Yes	M
Bed3	Mesolithic	9805-9385 BC	Bedburg-Königshoven	Germany	18.50	Yes	F
Bed4	Mesolithic	10036-9647 BC	Bedburg-Königshoven	Germany	2.13	Yes	F
Tri1	Mesolithic	8261-7960 BC	Trier-Brüderkrankenhaus	Germany	2.27	Yes	M
Hxh1	Neolithic		Herxheim	Germany	8.54	Yes	F
Hxh2	Neolithic		Herxheim	Germany	22.84	Yes	F
Kir18	Neolithic	3798-3641 BC	Kirschbaumhöhle	Germany	0.77	Yes	F
Kir4	Neolithic	2816-2671 BC	Kirschbaumhöhle	Germany	15.14	Yes	F
Fal7	Medieval		Falkenwalde	Germany	2.07	Yes	M
Dyr1	Medieval (late slavic)		Dyrotz 25	Germany	2.35	Yes	F
Sch1	Neolithic		Schipluiden	Netherlands	1.73	Yes	F
Viz1	Bronze Age		Vianen-Zijdervel	Netherlands	0.72	Yes	M
Viz2	Bronze Age		Vianen-Zijdervel	Netherlands	2.13	Yes	M
Bun1	Iron Age		Bunnik	Netherlands	1.79	Yes	F
Bun2	Iron Age		Bunnik	Netherlands	2.41	Yes	M
Ste1	Iron Age/Roman		Stevenshofjesp	Netherlands	0.96	Yes	F
Bri1	Roman		Britsum	Netherlands	3.26	Yes	M
Dro1	Roman		Dronrijp-zuid	Netherlands	2.88	Yes	M
Els1	Roman		Elst	Netherlands	2.56	Yes	M
Otb1	Roman		Oosterbeintum	Netherlands	2.62	Yes	M
Win1	Roman	417-561 AD	Winsum	Netherlands	19.03	Yes	F
Fir1	Medieval		Firdgum	Netherlands	3.21	Yes	F
Hou1	Medieval		Houten-Loerik	Netherlands	2.60	Yes	M
Hou2	Medieval		Houten-Hoogdij	Netherlands	2.67	Yes	F

5.3.2. Geographic Genetic Variation

5.3.2.1. Principal Components Analysis

Principal components analysis of modern cattle diversity creates a recognisable triangle pattern (Fig 5.3 & Appendix Fig 1) splitting modern domestic *Bos* among European taurine animals, African taurine animals and Asian indicine animals with hybrids intermediate to theise vertices. PC1 splits European taurine from Asian indicine, whilst PC2 splits African taurine from the other two clusters. The modern European cluster splits into Southern European breeds, such as Charolais and Limousin, and the Northern European breeds including Holstein, Kerry and Highland.

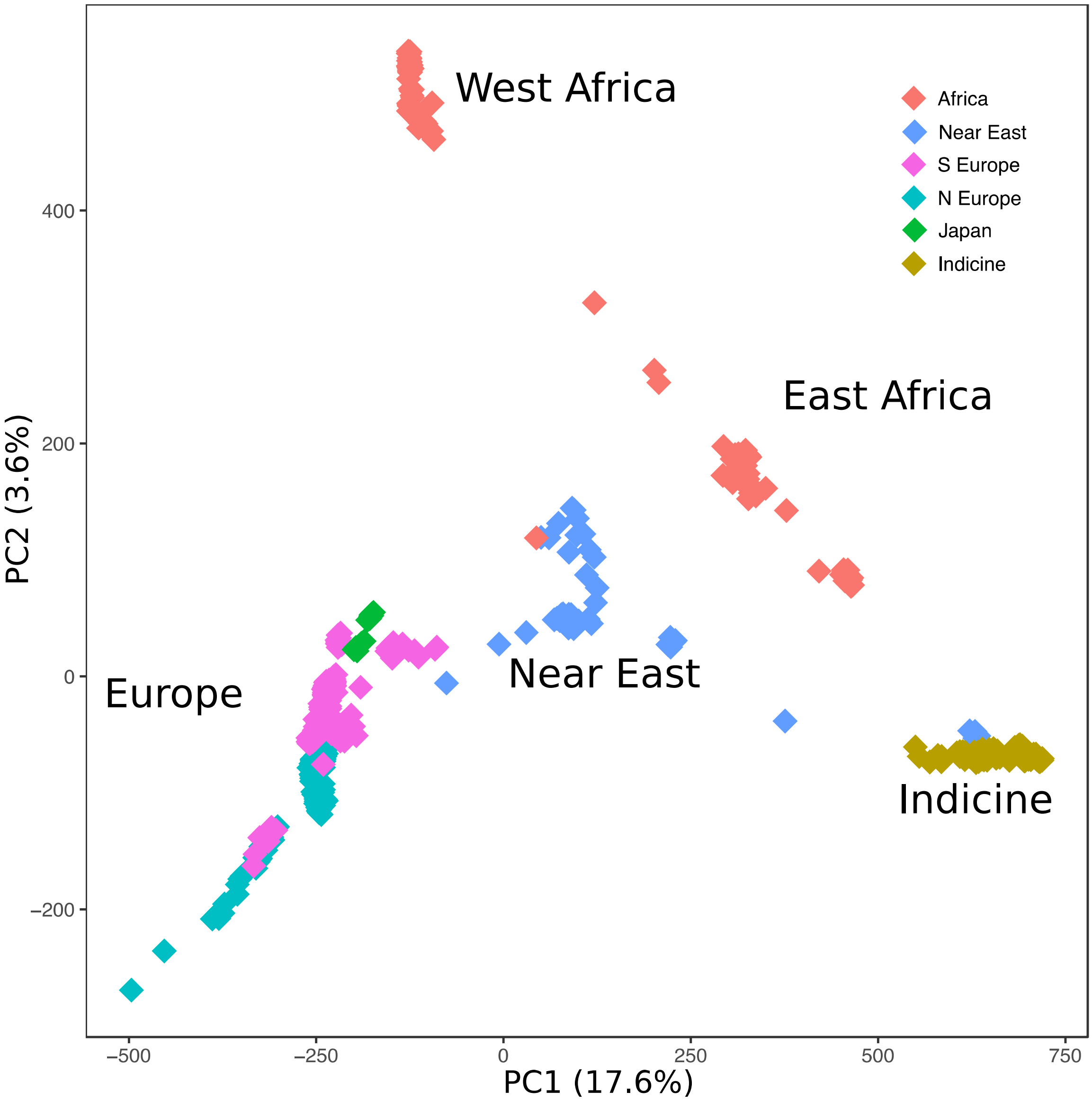


Fig 5.3. *LASER PCA of world wide modern cattle breeds using varaints from the 770K Illumina HD Bovine chip. PC1 splits taurine cattle and indicine cattle, whilst PC2 splits African taurine from European taurine breeds. Europe itself splits into Southern Europe and Northern European breeds. For exact breed placing seen Appendix Fig 1.*

When ancient samples are projected onto this modern background (Fig 5.4) the geographic placement of animals is concordant with the modern samples. To the side of the ancient and modern European cattle sits a cluster of the European auroch from geographical locations including Germany, Poland and Britain with temporal range from the Mesolithic to the Medieval Slavic period. No observable temporal or geographical structure is present within this population, with the exception of the Medieval Slavic Fal7 which is pulled towards the European taurine cluster.

Broadly speaking there are three clusters of Neolithic samples, 1) the Near East, 2) Eastern European/Germany and 3) the Atlantic Edge cluster including the Netherlands, Britain and Ireland. The early Anatolian domestic animals are positioned furthest from the modern European clusters. In general, the Neolithic animals clusters are placed in stepwise progression towards modern day European breeds, imitating the migration of the Neolithic through Europe. This migration culminates with the grouping of Neolithic British and Irish samples with modern day British and Irish Breeds such as Jersey, Kerry and Highland.

Exceptions to this Neolithic pattern include the Late Neolithic German Kir4 which is palced near the early Neolithic Bulgarian sample Dzh2 as opposed to clustering with the earlier German Neolithic sample of Hxh1. The Late Neolithic Polish sample Kie2 sits away from the ancient samples between the European taurine and European auroch clusters. It should also be noted that although the archaeological period of all three clusters is the same, the ages of the samples defining the clusters are quite different. The Anatolian samples are the oldest of the samples; Sub1 and Men2 (Table 5.4). Whilst these are the oldest directly dated domesticated samples analysed, they are approximately some 3-4K years post initial domestication. The low coverage genome Ch22 while not directly dated is from Çatalhöyük a large Neolithic settlement in Anatolia dating from 7,400-6,000 BC. Here, due to morphological characteristics of the *Bos* skulls found in the earlier phases it is believed wild aurochsen remains are present (Jacqui Mullville *per comm.*). This sample has a mtDNA haplogroup of T3 and unfortunately an ID of domestic or wild can not be achieved.

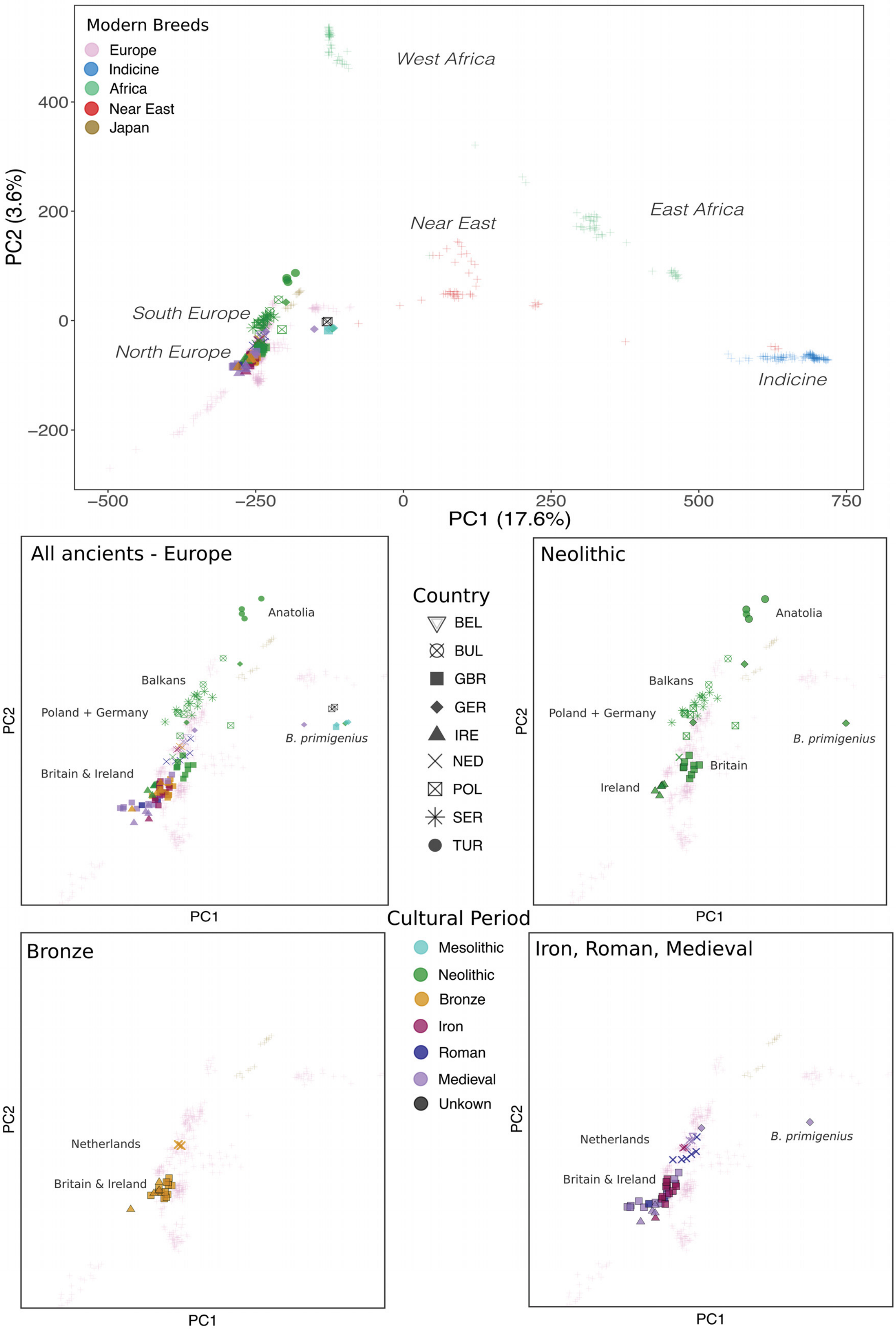


Fig 5.4. *LASER PCA of world wide modern cattle populations with the ancient samples projected on via a procrustes analysis, using variants from the 770K Illumina HD Bovine chip. Four zooms of Europe from the main PCA are then shown, with a breakdown into cultural periods. For a modern only PCA see Appendix Figure 1.*

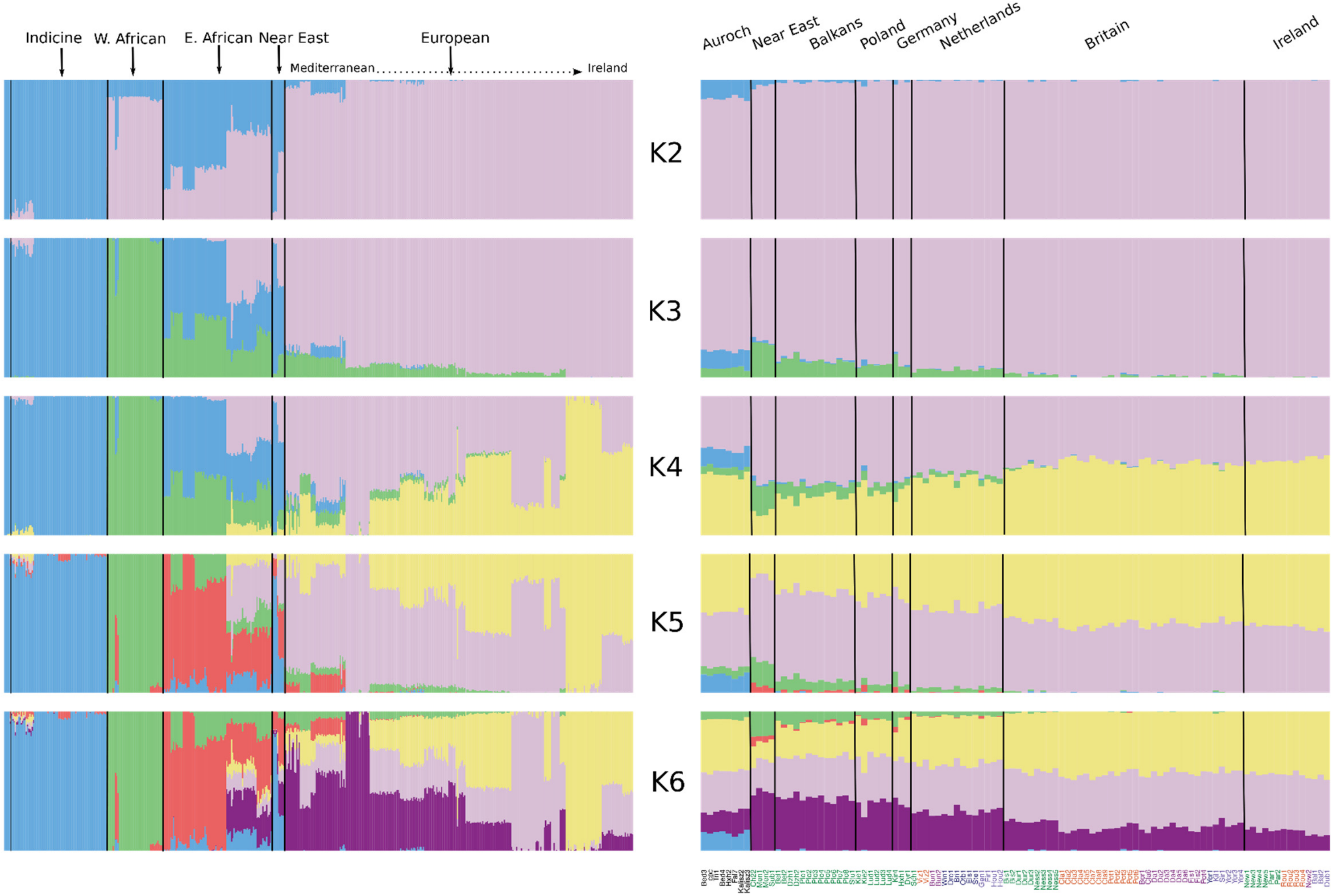


Fig 5.5. *Unsupervised ADMIXTURE analysis of modern and ancient cattle populations using variance from the 770K Illumina HD. Ancestry coefficient K2-6 are shown, with modern samples on the left and ancient samples on the right. The unmistakable split between indicine and taurine breeds is present at K=2, while European taurine, African taurine and Indicine split at K=3. Ancient samples are labelled and ancient taurine samples coloured on cultural period; Neolithic, Bronze Age, Iron Age, Roman, Medieval.*

The next large cluster of samples are from sites with an age range of approximately 6,050 BC to 3,000BC. The later samples in this group are the Polish samples, Kei1 and Kie2, which are positioned somewhat outside the cluster, with Kie2 towards the auroch cluster and Kie1 towards the Atlantic Edge samples. The British and Netherlands samples dating from approximately 3,500 BC until 2,500 BC cluster together, while sitting apart from this cluster are the Irish samples dating from approximately 3,000 BC until 2,500 BC.

Through the time transects of Britain and Ireland it is clear that the samples stay clustered to their geographic origin, with no discernible separation into subclusters of Bronze Age, Iron Age, Roman or Medieval samples. One apparent difference within the British time series is the clustering of the Neolithic animals away from the the Atlantic Edge cluster to which the Irish Neolithic and all post-Neolithic British samples belong. Within the Netherlands time transect there is a slight shift from the group containing the Neolithic sample Sch1 and a number of Roman samples upwards towards the Bronze Age, Iron Age and Medieval cluster.

5.3.2.2. Ancestral Estimation-ADMIXTURE

The modern geographic structure is also demonstrated in the admixture analysis (Fig 5.5), with K=2 splitting the components of taurine and indicine and K=3 distinguishing between European taurine, African taurine and indicine. As K values increase the breed structure emerges, for example the new component of K=4 maps to the Angus breed, whilst K=5 creates a split within the African breeds and K=6 distinguishes the Brown Swiss breed.

Within ancient samples, for K=2 all samples contain the pink European component, with the European aurochs samples alone displaying a substantial indicine component. With K=3 the African component is largest in the Anatolian Neolithic animals, diminishing through time and geographical distance. Intriguingly two non-Anatolian samples contain larger African components than their contemporaries and geographical neighbours, Neolithic Dzh2 (Bulgaria) and Neolithic Kir4 (Germany). With increased K values modern breeds influence the components of the ancients, at K=4 it is noticeable that the yellow componenet of Angus, a British Breed, is greater in the British and Irish samples than the continental European animals. Whilst at K=6 the purple Central European (Brown Swiss) component peaks within the Southwest Asian and Southeast and Central European ancient samples. K=5 introduces a red component which noticeably decreases with distance from the domestication centre, while the European aurochs do not contain this component. Again the sample Kie2 stands out as somewhat different when compared to sample Kei1 from the same site, as well as other Polish Neolithic samples.

5.3.2.3. Geographical and Temporal Continuity

5.3.2.3.1. Neolithic Europe

5.3.2.3.1.1. Outgroup f_3-statistic

The patterns of shared drift (Fig 5.6 & Appendix Fig 2) within the Neolithic echoes the stepwise progression of the PCA (Fig5.4). Neolithic samples with the greatest shared drift are those closest geographically and temporally. Interestingly, the contemporaneous Late British Neolithic samples from The Ness of Brodgar show high shared drift with one another. Additionally, the Irish and British Neolithic samples share higher values of shared drift with the samples from the Polish site Ludwinowo than they do with the majority of Neolithic samples from Serbia, Bulgaria and Anatolia. The two Polish Late Neolithic samples (Kei & Kei2) do not show a great affinity to the earlier Polish Neolithic sites, however sample Kie1 is only 0.04x average genome coverge. Once again Kir4 for is an exception to the pattern of shared drift with no obvious leaders in shared drift.

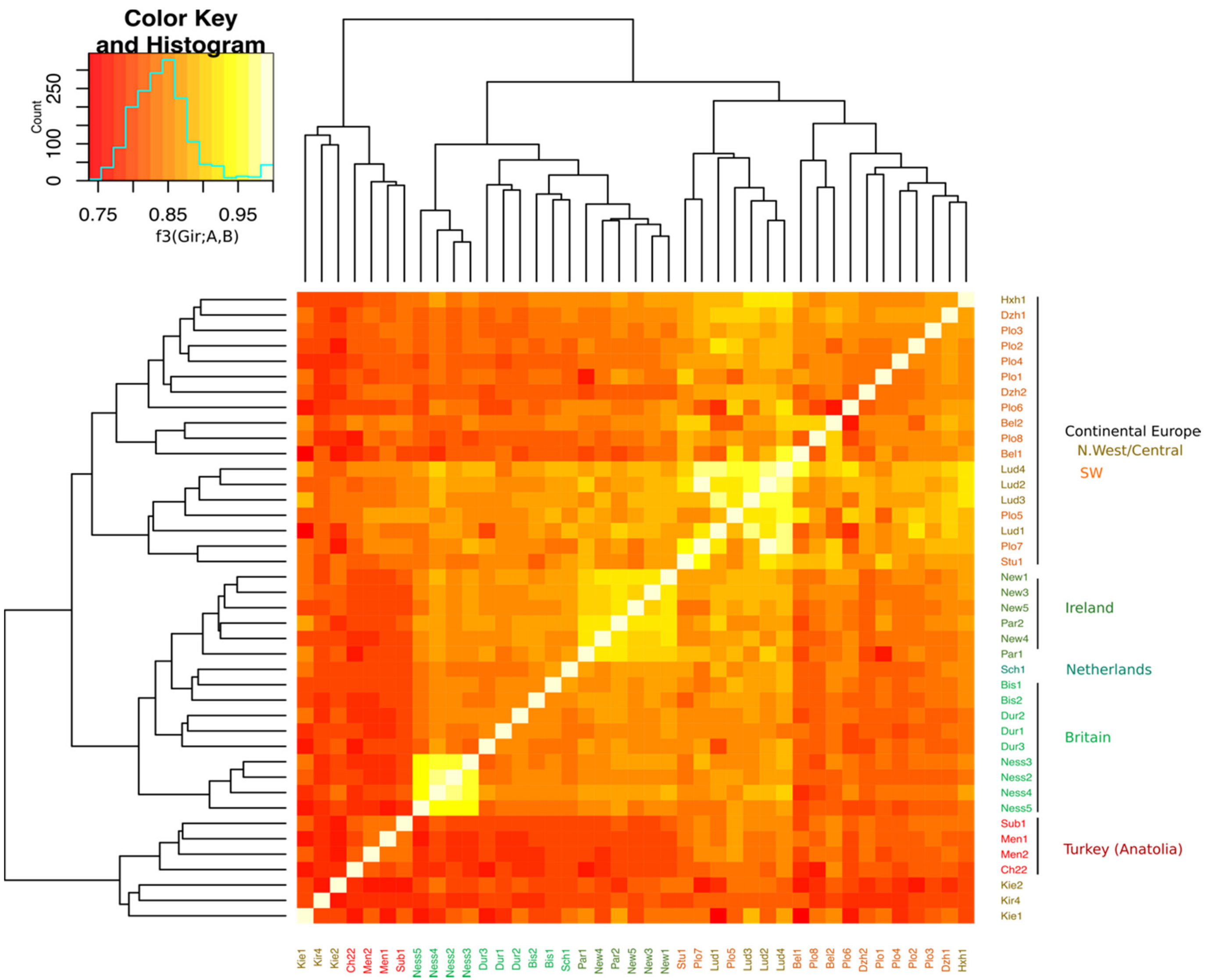


Fig 5.6. *Pairwise test of outgroup f_3-statistic of Neolithic samples demonstrates the greatest shared drift is between those closest geographically and temporally, mirrowing the PCA.The indicine breed Gir was used as the outgroup to measure shared drift between samples. Sample names are coloured by geographical affinity.*

5.3.2.3.1.2. *Measurement of Neolithic Population Clade Integrity - The D-statistic*

The *D*-statistic was used to test clade integrity of the Neolithic clades as implied from PCA and outgroup f_3 (Fig 5.4 & 5.6), see below for the three research questions addressed. For these statistics populations of ancient samples were formed (Appendix Table 10). In general, the results of the *D*-statistic tests confirm the groupings of outgroup f_3 and the PCA of populations from the three main Neolithic clusters. Some anomalies were present in the tests for clade integrity within the Atlantic Edge; the (Irish,British) clade was disrupted by Neolithic German sample (Hxh1) to varying degrees dependent on which British and Irish populations were used.

1) *Does gene flow from Neolithic Anatolia disrupt clade integrity of Neolithic Southeastern European populations (Bulgaria and Serbia) with other Neolithic European populations?*

The clades grouping Neolithic Bulgaria and Early Neolithic Poland plus Neolithic Bulgaria and Neolithic Germany hold. However, for all other tests the Neolithic Bulgarian animals share derived alleles unevenly with the test population **(**Table 5.5), suggesting Neolithic Bulgarian samples have a closer relationship to the Anatolia samples than the test populations. If Neolithic Bulgaria is substituted for Neolithic Serbia, clades with Neolithic Germany and Early Neolithic Poland hold their integrity, while all other clades do not, echoing the placement of samples on the PCA.

Table 5.5. *The Testing of Neolithic caldes in respect to Neolithic Turkey (Anatolia) population using the D-statistic. Here clades formed of the two Southeastern European populations, Bulgaria and Serbia respectively, with other Neolithic populations are tested. Negative D and Z-score indicates more shared derived alleles between Pop2 and Pop3, positive indicates more shared derived alleles between Pop2 and Pop4.GBR= Britain, IRE= Ireland, NED=Netherlands, SRB= Serbia, DEU = Germany, POL= Poland. See Appendix Table 10. for definition of ancient populations. Z-scores are coloured to indicate significance level: green= P ≤ 0.001, orange = 0.001 < P ≤ 0.01, red = 0.01 < P ≤ 0.05.*

Pop2	Pop3	Pop4	*D*	stdr error	*Z-score*	BABA	ABBA	Sites
D(Pop4,Neolithic Bulgaria)Neolithic Turkey, Gaur))								
Neolithic Turkey	Neolithic Bulgaria	Neolithic GBR 3	-0.0839	0.00751	-11.171	6394	7565	95973
		Neolithic GBR 2	-0.0721	0.00666	-10.82	6492	7501	95593
		Neolithic GBR 1	-0.0609	0.00672	-9.057	6457	7295	95922
		Neolithic IRE 1	-0.0516	0.00623	-8.281	6588	7304	96266
		Neolithic NED	-0.069	0.00897	-7.69	5336	6126	80705
		Neolithic IRE2	-0.0414	0.00935	-4.429	3324	3611	48295
		Neolithic SRB	-0.0159	0.00612	-2.599	6472	6681	96072
		Neolithic DEU	-0.0123	0.00863	-1.425	6479	6641	96107
		Neolithic POL	-0.0049	0.00861	-0.569	4234	4276	63089

D(Pop4,Neolithic Serbia)Neolithic Turkey, Gaur))								
Neolithic Turkey	Neolithic Serbia	Neolithic GBR 3	-0.0686	0.00695	-9.876	6515	7475	95804
		Neolithic GBR 2	-0.0568	0.00588	-9.655	6594	7389	95428
		Neolithic GBR 1	-0.0455	0.00627	-7.261	6629	7260	95753
		Neolithic IRE 1	-0.0364	0.00545	-6.684	6674	7179	96091
		Neolithic NED	-0.0507	0.00831	-6.101	5500	6088	80585
		Neolithic IRE2	-0.0261	0.00867	-3.011	3407	3590	48237
		Neolithic DEU	0.0037	0.00851	0.435	6647	6599	95936
		Neolithic POL	0.0088	0.00775	1.135	4370	4293	62993

2) *Does gene flow from Neolithic Germany disrupt the clade integrity of the Neolithic Netherlands and the Neolithic populations from Britain and Ireland?*

The clade integrity of Neolithic Netherlands and Neolithic Britain holds (Table 5.6), which again follows this grouping on the PCA (Fig5. 4). However, the positive *D*-scores (*Z-score* > 2) for the Neolithic Irish samples are suggestive of the Neolithic Irish sharing more derived alleles with Neolithic Germany than the Neolithic Netherlands sample.

Table 5.6. *Testing of the Neolithic of the Netherlands and Ireland clades, in respect to Neolithic German sample Hxh1, using the D-statistic. Here clades of (Neolithic Netherlands, Neolithic Britain) and (Neolithic Netherlands, Neolithic Ireland) are tested. GBR= Britain, IRE= Ireland. Negative D and Z-score indicates more shared derived alleles between Pop2 and Pop3, positive indicates more shared derived alleles between Pop2 and Pop4. See Appendix Table 10 for definition of ancient populations. Z-scores are coloured to indicate significance level: green= P ≤ 0.001, orange = 0.001 < P ≤ 0.01, red = 0.01 < P ≤ 0.05.*

Pop2	**Pop3**	**Pop4**	***D***	**stdr error**	***Z-score***	**BABA**	**ABBA**	**Sites**
D(Pop4,Neolithic Netherlands) Neolithic Germany, Gaur))								
Neolithic Germany	Neolithic Netherlands	Neolithic GBR 2	-0.0113	0.01037	-1.09	5959	6095	80115
		Neolithic GBR 3	-0.0097	0.01083	-0.896	5925	6041	80406
		Neolithic GBR 1	0.0034	0.01104	0.308	5867	5828	80366
		Neolithic IRE 1	0.0219	0.00961	2.279	6057	5797	80613
		Neolithic IRE 2	0.0385	0.01413	2.725	3103	2873	40840

3) *Does gene flow from Neolithic Germany disrupt clade integrity of the Neolithic Irish and British populations?*

The clade of the two Irish populations (W. Ireland & E. Ireland) holds (Table 5.7). The clade between the Irish Neolithic and Late Neolithic British populations of the Ness of Brodgar (Neolithic GBR 3) and Durrington Walls (Neolithic GBR 2), respectively, do not hold as the Neolithic Irish populations share more derived alleles with the German Neolithic sample. The clade between the Irish Neolithic populations and the Early Neolithic population of Bishops Cannings (Neolithic GBR 1) is also disrupted, but with lower *D* values and significance than the other two British populations (*Z*-score >2). This pattern of the Early British Neolithic behaving differently to the Late British Neolithic is detected again when the clades of the different British Neolithic are tested. While the Late Neolithic populations remain as a clade, that between these Later British Neolithic samples with the Early British Neolithic is disrupted by Hxh1 (*Z*-score >2).

Table 5.7. *Testing of Neolithic caldes of Britain and Ireland, in respect to Neolithic German sample Hxh1, using the D-statistic. Here a number of clades formed of British and Irish Neolithic populations are tested for clade integrity. GBR= Britain, IRE= Ireland. Negative D and Z-score indicates more shared derived alleles between Pop2 and Pop3, positive indicates more shared derived alleles between Pop2 and Pop4. See Appendix 10 for definition of ancient populations. Z-scores are coloured to indicate significance level: green= P ≤ 0.001, orange = 0.001 < P ≤ 0.01, red = 0.01 < P ≤ 0.05.*

Pop2	Pop3	Pop4	*D*	stdr error	Z-score	BABA	ABBA	Sites
D(Pop4,Neolithic Ireland 1) Neolithic Germany, Gaur))								
Neolithic Germany	Neolithic Ireland 1	Neolithic GBR 2	-0.0365	0.00712	-5.123	6890	7412	95478
		Neolithic GBR 3	-0.0382	0.00795	-4.804	6864	7409	95853
		Neolithic GBR 1	-0.0176	0.00805	-2.185	6943	7192	95800
		Neolithic IRE 2	0.0008	0.01000	0.08	3346	3340	48233
D(Pop4,Neolithic Ireland 2 Neolithic Germany, Gaur))								
Neolithic Germany	Neolithic Ireland 2	Neolithic GBR 2	-0.0397	0.01111	-3.573	3413	3695	47946
		Neolithic GBR 3	-0.0361	0.01167	-3.093	3431	3688	48110
		Neolithic GBR 1	-0.0348	0.01164	-2.989	3389	3633	48089
D(Pop4,Neolithic Britain 1) Neolithic Germany, Gaur))								
Neolithic Germany	Neolithic Britain 3	Neolithic GBR 2	0.0021	0.00875	0.24	7169	7139	95209
		Neolithic GBR 1	0.0203	0.00888	2.286	7268	6978	95526
D(Neolithic Britain 3 ,Neolithic Britain 2) Neolithic Germany, Gaur))								
Neolithic Germany	Neolithic Britain 2	Neolithic GBR 1	0.0187	0.00829	2.255	7232	6967	95151

5.3.2.3.2. Atlantic Edge Time Series

5.3.2.3.2.1. Outgroup f_3-statistic

To measure the shared drift between some samples from the Atlantic Edge, the outgroup f_3-statistic was performed in a pairwise fashion (Gir;test,test) (Fig 5.7 & Appendix Fig 2). This test demonstrates that the British Bronze Age samples share more drift, therefore indicating a closer relationship, with Irish Neolithic than the British Neolithic, while the Irish Bronze Age shares most drift with the Irish Neolithic. Overall, post-Bronze Age the patterns of shared drift become more difficult to differentiate. However, it is evident that British and Irish samples share most drift with samples from the two islands, as opposed to the Iron Age British and the Iron Netherlands samples demonstrating high levels of drift. Interestingly a number of the Netherlands samples (Hou2, Hou1, Viz2, Bri1, Otb1) show lowest values of shared drift with the other samples in the time series. Additionally, shared drift between the ancient British/Irish samples with the modern British/Irish breeds is greater than the shared drift between the ancient samples and the continental breeds (Appendix Fig 2). Through the time series it is evident that samples show greater degree of geographical, as opposed to temporal affinity, in keeping with the PCA results.

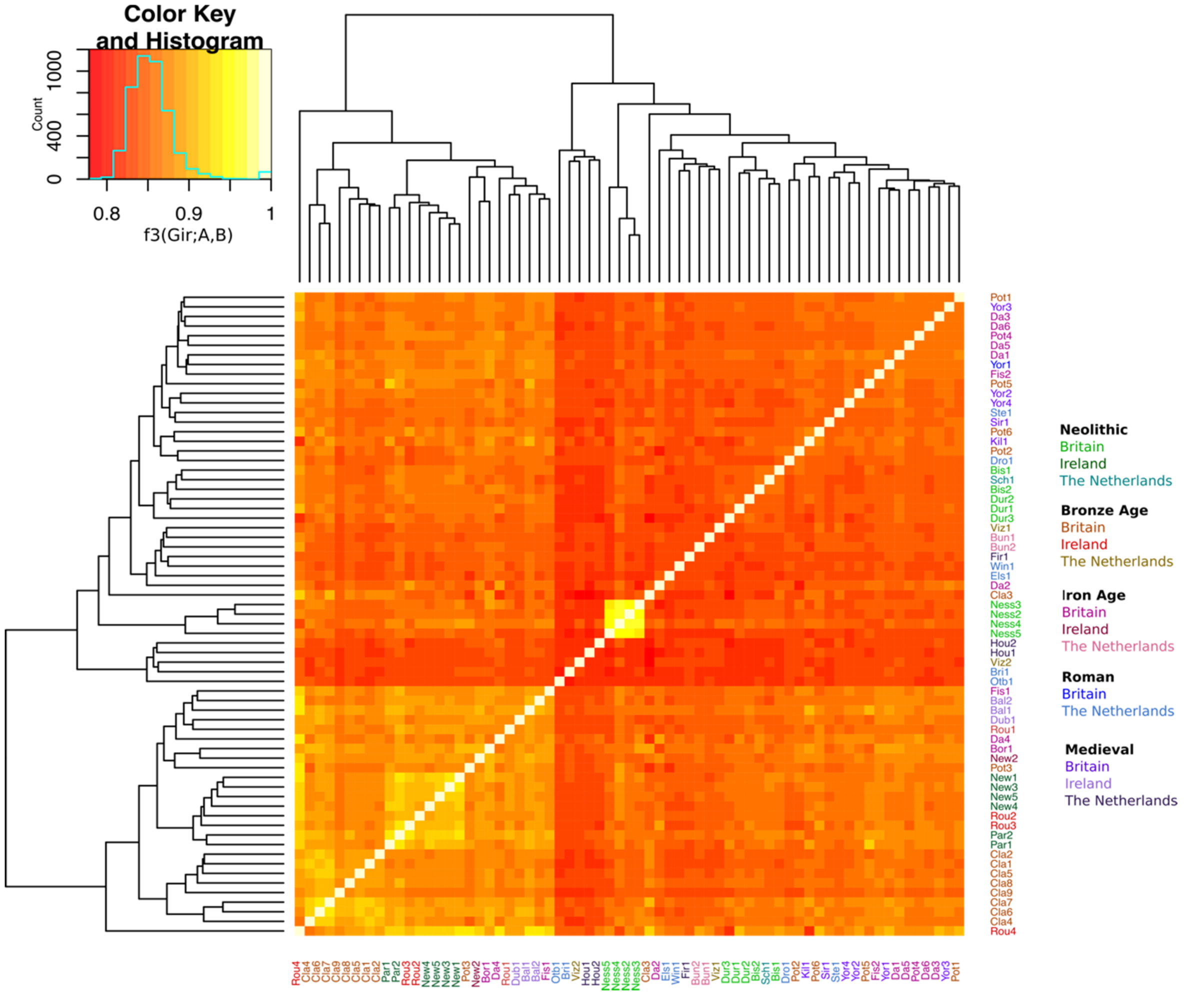


Fig.5.7. *Pairwise test of outgroup f_3-statistic of the Atlantic Edge time series demonstrates the greatest shared drift withinthe Late Neolithic British site of the Ness of Brodgar and between the Irish Neolithic and Bronze populations. Additionally the British Bronze Age samples share greater drift with the Irish Neolithic than the British. The indicine breed Gir was used as the outgroup to measure shared drift between samples. Sample names are coloured by geographical and temporal affinity.*

5.3.2.3.2.2. Measurement of Clade Integrity of the Atlantic Edge - The D-statistic

To test clade integrity four research questions were proposed. The results of these *D*-statistic tests indicate clade integrity between the more recent populations from Britain, such as the Medieval Britain and modern British populations. However, disruption of clade integrity does occur between the Late British Neolithic populations and the British Bronze populations.

1) *Do the Neolithic and Bronze Age Irish populations disrupt the clade integrity of the British Neolithic and Bronze Age?*

The clade integrity between the Late British Neolithic populations from Durrington Walls and The Ness of Brodgar (Neolithic GBR 2 & 3 respectively) and British Bronze Age is disrupted by the Irish Neolithic and Irish Bronze Age (*Z*-score > 2) (Table 5.8). However clade integrity between the Early Neolithic British (Neolithic GBR 1) and the British Bronze Age holds.

Table 5.8. *Testing of the British Neolithic and Bronze Age caldes, in respect to Neolithic and Bronze Age Ireland, using the D-statistic. Here a number of clades formed of British and Irish Neolithic populations are tested for clade integrity. GBR= Britain, IRE= Ireland. Negative D and Z-score indicates more shared derived alleles between Pop2 and Pop3, positive indicates more shared derived alleles between Pop2 and Pop4. See Appendix Table 10. for definition of ancient populations. Z-scores are coloured to indicate significance level: green= P ≤ 0.001, orange = 0.001 < P ≤ 0.01, red = 0.01 < P ≤ 0.05.*

Pop2	Pop3	Pop4	*D-stat*	stdr error	*Z-score*	BABA	ABBA	Sites
D(Bronze GBR, Neolithic GBR),Neolithic Ireland 1, Gaur))								
Neolithic Ireland 1	Neolithic GBR 1	Bronze GBR 2	0.0048	0.0066	0.731	7056	6989	92754
	Neolithic GBR 1	Bronze GBR 1	0.0075	0.0057	1.311	7323	7214	95950
	Neolithic GBR 2	Bronze GBR 2	0.015	0.0061	2.465	7106	6897	92466
	Neolithic GBR 2	Bronze GBR 1	0.018	0.0052	3.495	7362	7101	95622
	Neolithic GBR 3	Bronze GBR 2	0.0274	0.0071	3.853	7237	6850	92811
	Neolithic GBR 3	Bronze GBR 1	0.0304	0.0060	5.048	7506	7064	96002
D(Pop4,Pop3),Neolithic Ireland 1, Gaur))								
Bronze Ireland	Neolithic GBR 1	Bronze GBR 2	0.0081	0.0070	1.154	7058	6944	92551
	Neolithic GBR 1	Bronze GBR 1	0.0117	0.0060	1.969	7315	7146	95716
	Neolithic GBR 2	Bronze GBR 2	0.0142	0.0067	2.131	7054	6857	92259
	Neolithic GBR 2	Bronze GBR 1	0.0178	0.0055	3.259	7324	7067	95384
	Neolithic GBR 3	Bronze GBR 2	0.0263	0.0075	3.525	7198	6828	92604
	Neolithic GBR 3	Bronze GBR 1	0.0311	0.0065	4.82	7481	7029	95767

2) *Do the Neolithic and Bronze Age populations from the Netherlands disrupt the clade integrity between the Neolithic and Bronze Age populations of Britain and Ireland, respectively?*

Gene flow from the Neolithic Netherlands sample (Sch1) does not disrupt the clades of Late British Neolithic (Neolithic GBR 2 & 3) and British Bronze Age (Table 5.9). However, some of the clades between the Late Neolithic British population and the Bronze Age populations are disrupted by the Bronze population from the Netherlands (*Z*-score > 2). Although this pattern is not repeated between the Early British Neolithic and the Bronze Age, suggesting clade integrity and no gene flow.

Table 5.9. *Testing of the British and Irish Neolithic and Bronze Age caldes, in respect to Neolithic and Neolithic and Bronze Age Netherlands, using the D-statistic. Here a number of clades formed of British and Irish Neolithic populations are tested for clade integrity. GBR= Britain, IRE= Ireland. Negative D and Z-score indicates more shared derived alleles between Pop2 and Pop3, positive indicates more shared derived alleles between Pop2 and Pop4. See Appendix Table 10 for definition of ancient populations. Z-scores are coloured to indicate significance level: green= P ≤ 0.001, orange = 0.001 < P ≤ 0.01, red = 0.01 < P ≤ 0.05.*

Pop2	Pop3	Pop4	*D-stat*	stdr error	*Z-score*	BABA	ABBA	Sites
D(Bronze GBR, Neolithic GBR),Neolithic Ireland 1, Gaur))								
Neolithic Netherlands	Neolithic GBR 1	Bronze GBR 1	-0.0225	0.0075	-3.019	5803	6070	80464
	Neolithic GBR 1	Bronze GBR 2	-0.0201	0.0085	-2.367	5650	5881	77933
	Neolithic GBR 2	Bronze GBR 1	0.0054	0.0069	0.788	5966	5901	80206
	Neolithic GBR 2	Bronze GBR 2	0.0064	0.0081	0.79	5772	5699	77706
	Neolithic GBR 3	Bronze GBR 2	0.0112	0.0096	1.17	5843	5714	77973
	Neolithic GBR 3	Bronze GBR 1	0.0101	0.0082	1.232	6026	5906	80502
D(Bronze GBR, Neolithic GBR),Neolithic Ireland 1, Gaur))								
Bronze Netherlands	Neolithic GBR 1	Bronze GBR 2	-0.002	0.0073	-0.274	6516	6543	88471
	Bronze IRE	Neolithic IRE 2	-0.0016	0.0108	-0.148	3226	3236	46053
	Bronze IRE	Neolithic IRE 1	-0.0007	0.0055	-0.127	6428	6437	91530
	Neolithic GBR 1	Bronze GBR 1	0.0028	0.0064	0.437	6752	6715	91435
	Neolithic GBR 2	Bronze GBR 2	0.0074	0.0070	1.059	6521	6426	88207
	Neolithic GBR 2	Bronze GBR 1	0.012	0.0059	2.042	6794	6633	91139
	Neolithic GBR 3	Bronze GBR 2	0.0178	0.0074	2.396	6674	6441	88522
	Neolithic GBR 3	Bronze GBR 1	0.0224	0.0066	3.405	6894	6592	91486

3) *Does gene flow from Iron Age and Medieval Netherland populations disrupt the clade integrity of British and Irish clades post-Bronze Age?*

The Netherlands Iron Age population does not disrupt the clades of British Iron Age and British Bronze Age with high significance (Z-score ≥ 3) (Table 5.10). However the British Iron Age populations tend to have more shared derived Iron Age Netherlands alleles than the British Bronze Age. Additionally, Iron Age Britain holds as a clade with Medieval Britain.

Table 5.10. *Testing of the post-Bronze Age British and irish caldes, in respect to Iron Age and medieval Netherlands, using the D-statistic. Here a number of clades formed of British and Irish Neolithic populations are tested for clade integrity. GBR= Britain, IRE= Ireland. Negative D and Z-score indicates more shared derived alleles between Pop2 and Pop3, positive indicates more shared derived alleles between Pop2 and Pop4. See Appendix Table 10 for definition of ancient populations. Z-scores are coloured to indicate significance level: green= P ≤ 0.001, orange = 0.001 < P ≤ 0.01, red = 0.01 < P ≤ 0.05.*

Pop2	Pop3	Pop4	*D-stat*	stdr error	*Z-score*	BABA	ABBA	Sites
D(Pop4, Iron GBR),Iron Netherlands, Gaur))								
Iron Netherlands	Iron GBR 2	Bronze GBR 1	-0.0133	0.00542	-2.453	6785	6968	94284
	Iron GBR 1	Bronze GBR 1	-0.0081	0.00439	-1.846	6928	7042	94840
	Iron GBR 2	Bronze GBR 2	-0.0074	0.00671	-1.103	6656	6755	91218
	Iron GBR 2	Iron GBR 1	-0.0051	0.00548	-0.93	6891	6962	94281
	Iron GBR 1	Bronze GBR 2	-0.0023	0.00550	-0.418	6729	6761	91702
	Iron GBR 2	Iron GBR 3	-0.0017	0.00876	-0.194	4104	4118	56995
	Iron GBR 1	Iron GBR 2	0.0051	0.00551	0.93	6962	6891	94281
	Iron GBR 1	Iron GBR 3	0.0075	0.00784	0.956	4231	4168	57271
D(Medieval GBR, Iron GBR),Medieval Netherlands, Gaur))								
Medieval Netherlands	Iron GBR 3	Medieval GBR	-0.0047	0.00798	-0.589	4268	4308	58045
	Iron GBR 2	Medieval GBR	-0.0027	0.00542	-0.498	7008	7047	95655
	Iron GBR 1	Medieval GBR	0.0009	0.00419	0.215	7127	7114	96237
D(Medieval GBR, Medieval IRE),Medieval Netherlands, Gaur))								
Medieval Netherlands	Medieval GBR	Medieval IRE	0.0091	0.00508	1.79	7088	6960	96122

4) *Does gene flow from the Medieval Netherlands disrupt clade integrity with Medieval Britain and modern breeds?*

The modern breeds British and Irish of Highland, Galloway, Kerry and Jersey all hold as a clade with the Medieval British populations (Table 5.11). The only clades to be significantly disrupted between a British modern breed and the Medieval population is that of the Angus and Guernesey breeds (minimum *Z*-score > 2).

Table 5.11. *The testing of clades formed of British Medieval and modern British and irish breeds, in respect to the Medieval Netherlands, using the D-statistic. Here a number of clades formed of British and Irish Neolithic populations are tested for clade integrity. GBR= Britain. Negative D and Z-score indicates more shared derived alleles between Pop2 and Pop3, positive indicates more shared derived alleles between Pop2 and Pop4. See Appendix Table 10. for definition of ancient populations. Z-scores are coloured to indicate significance level: green= P ≤ 0.001, orange = 0.001 < P ≤ 0.01, red = 0.01 < P ≤ 0.05.*

Pop2	Pop3	Pop4	*D-stat*	stdr error	*Z-score*	BABA	ABBA	Sites
D(Pop4, Medieval GBR), Medieval Netherlands, Gaur))								
Medieval Netherlands	Medieval GBR	ANG	-0.035	0.00472	-7.42	7086	7600	96243
		GNS	-0.0122	0.00499	-2.446	7158	7334	96243
		JER	-0.002	0.00426	-0.47	7296	7326	96243
		KER	-0.0019	0.00445	-0.427	7210	7238	96243
		HIG	-0.0019	0.00525	-0.362	7122	7148	96243
		GAL	0.0018	0.00462	0.39	7166	7140	96243

5.3.2.3.2.2. *Population admixture in the Atlantic Edge - f_3-statistic*

The only Bronze Age population within which admixture was significantly detected is the Irish Bronze Age samples from the west of Ireland (Table 5.12). This population is a mixture of Neolithic Irish populations, and another population such as British Bronze, British Neolithic or Bronze Netherlands. The other test with a significant (p-value < 0.05) is the admixture between Iron Age Netherlands and the Bronze Age Cladh Hallan samples forming the Iron Age samples from the Outer Hebrides (Appendix Table 9).

Table 5.12. *Testing admixture scenarios using the f_3-statistic (Target; Source1, Source2) that gave rise to the Irish Bronze Age population. For full results of population admixture in the Atlantic Edge see Appendix Table 9. Z-scores are coloured to indicate significance level: green= P ≤ 0.001, orange = 0.001 < P ≤ 0.01, red = 0.01 < P ≤ 0.05.*

Source 1	Source2	Target	F3	Stdr. err	Z	SNP no.
Bronze GBR	Neolithic IRE	Bronze IRE	-0.009811	0.00238	-4.121	55066
Bronze GBR	Neolithic IRE	Bronze IRE	-0.009811	0.00238	-4.121	55066
Bronze NED	Neolithic IRE	Bronze IRE	-0.010965	0.002809	-3.904	44097
Neolithic GBR 3	Neolithic IRE	Bronze IRE	-0.010301	0.002705	-3.809	47216
Neolithic GBR 2	Neolithic IRE	Bronze IRE	-0.009672	0.002669	-3.623	48685
Bronze GBR 2	Neolithic IRE	Bronze IRE	-0.009601	0.002796	-3.434	46461
Bronze GBR 2	Neolithic IRE	Bronze IRE	-0.009601	0.002796	-3.434	46461
Neolithic GBR 1	Neolithic IRE	Bronze IRE	-0.008425	0.002566	-3.283	47905
Bronze NED	Neolithic IRE 2	Bronze IRE	-0.014074	0.004331	-3.25	19443
Neolithic NED	Neolithic IRE	Bronze IRE	-0.005778	0.003148	-1.835	36372
Bronze GBR 2	Neolithic IRE 2	Bronze IRE	-0.007257	0.004365	-1.662	20191
Bronze GBR 2	Neolithic IRE 2	Bronze IRE	-0.007257	0.004365	-1.662	20191

5.3.3. *Bos primigenius* Introgression into Domestic Cattle

The overall trend of European Auroch introgression is an increase in auroch introgression with distance from the domestication centre (Fig 5.8 & Appendix Fig 3). The highest levels of auroch introgression occur in the samples from the Atlantic Edge, in particular in Late Neolithic and Bronze Age Britain. Additionally, while the highest levels of introgression are recorded in the Neolithic, there are differing trends of auroch introgression Post-Neolithic within the Atlantic Edge transect.

5.3.3.1. D-statistics to Test European Auroch Introgression into Domestic Taurine Cattle.

The overall trend in the *D*-statistic test for European auroch introgression, *D*(Gaur,Bed3;Sub1,Test ancient), (Fig 5.8) is an increase of shared derived alleles between the test and the auroch sample as we move geographically further away from the Near East, suggestive of auroch introgression increasing with distance from domestication centre. A result echoed by the test *D*(Gaur,CpC;Sub1,Test ancient) (Appendix Fig 3) utilising the presviously published CpC (Park et al 2015).

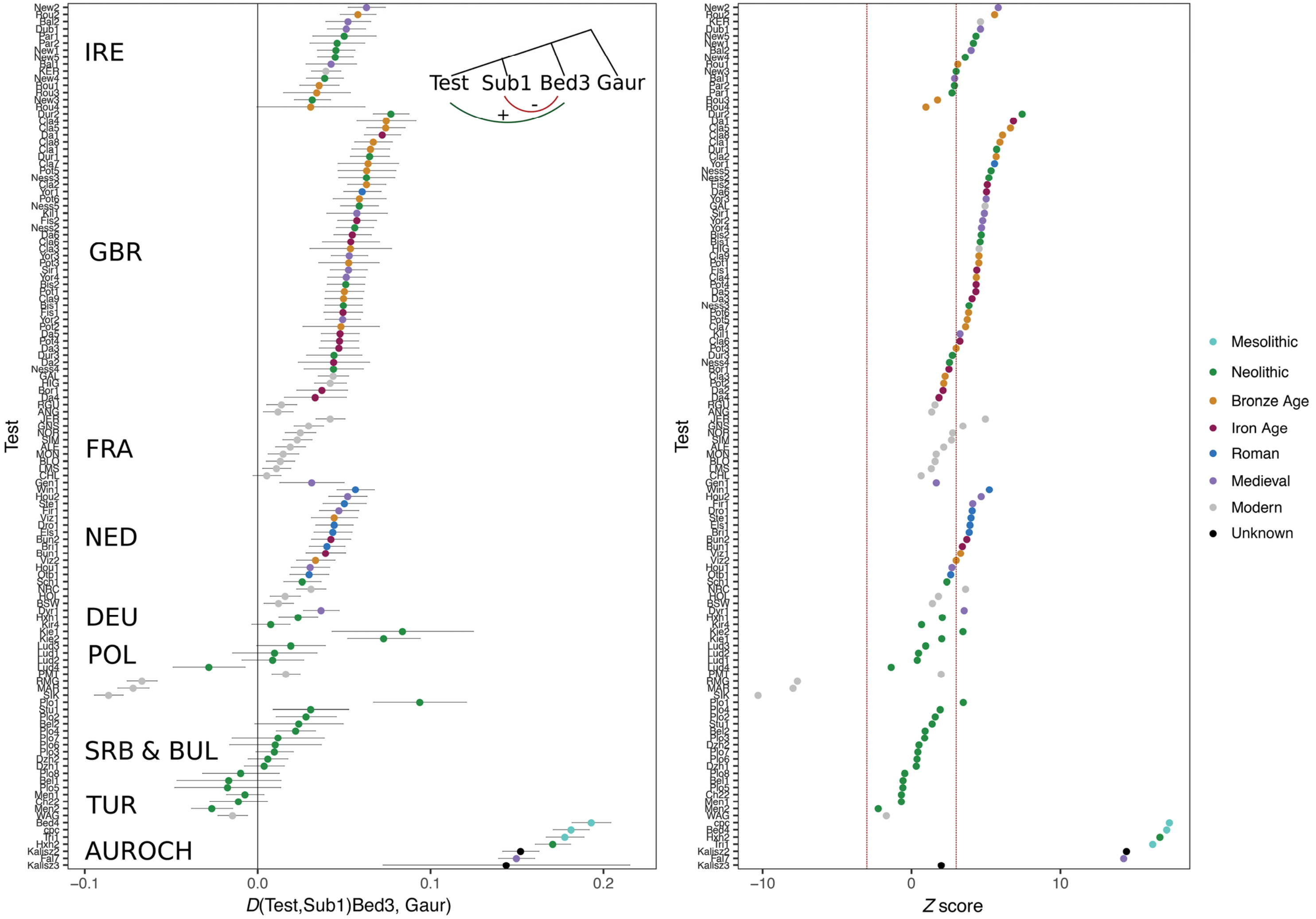


Fig 5.8. *D-statistic test to measure introgression from a Mesolithic European B. primigenius (Bed3) into ancient samples, in comparison to early the Neolithic Anatolian sample Sub1. Red line denotes Z score of 3 or -3. Points are coloured for cultural period. There is a trend of increasing D further from Anatolia.*

5.3.3.1.1. European Auroch Introgression in Neolithic Populations

Below are the results of four groups of specific *D*-statistic tests to calculate the difference in auroch derived alleles in different Neolithic populations. The trend of an increase in auroch introgression with geographical distance from Anatolia remains.

1) *Do Neolithic populations have a greater number of shared auroch derived alleles than the Neolithic Anatolian populations?*

Where possible, samples were grouped into geographical and temporal populations (Appendix Table 10). The *D*-statistic results of (Gaur,auroch;Neolithic Anatolia, test), show a similar trend shown with individual ancient samples (Table 5.13). However, with the population groupings the trends of significance are clearer, with all groups bar the Bulgarian Neolithic, with a Z-score ≤3.

Table 5.13. *D-statistic test to measure introgression from the European auroch population into Neolithic populations, in comparison to early the Neolithic Anatolian (Turkey=TUR) population. GBR= Britain, IRE= Ireland, NED=Netherlands, SRB= Serbia, DEU = Germany, POL= Poland. Negative D and Z-score indicates more shared derived alleles between Pop2 and Pop3, positive indicates more shared derived alleles between Pop2 and Pop4. See Appendix Table 10. for definition of ancient populations. Z-scores are coloured to indicate significance level: green= P ≤ 0.001, orange = 0.001 < P ≤ 0.01, red = 0.01 < P ≤ 0.05.*

Pop3	Pop4	*D*-stat	stdr error	*Z*-score	BABA	ABBA	Sites
D(Pop4,Neolithic TUR)auroch,Gaur))							
Neolithic TUR	Neolithic BUL	0.0115	0.006317	1.814	5930	5795	96252
	Neolithic DEU KIR4	0.0181	0.007397	2.451	6315	6090	96262
	Neolithic POL 1	0.0216	0.007743	2.792	4012	3842	63090
	Neolithic DEU	0.0322	0.007853	4.1	6233	5844	96116
	Neolithic NED	0.0381	0.007566	5.031	5448	5049	80709
	Neolithic SRB	0.0316	0.005385	5.869	6168	5790	96081
	Neolithic IRE 2	0.0544	0.009207	5.913	3339	2995	48293
	Neolithic POL 2	0.0822	0.013342	6.163	1819	1542	25129
	Neolithic GBR 1	0.0576	0.00626	9.203	6645	5921	95929
	Neolithic IRE 1	0.0492	0.005208	9.441	6632	6010	96276
	Neolithic GBR 3	0.067	0.006553	10.23	6755	5906	95986
	Neolithic GBR 2	0.0825	0.00609	13.551	6872	5825	95606

2) *Do Neolithic samples have a greater number of shared auroch derived alleles than the Neolithic Serbian samples?*

The Bulgarian and Anatolian Neolithic populations have significantly less shared derived alleles. Converesley, Late Polish Neolithic, British and Irish Neolithic all have significantly more sharing than the Serbian Neolithic (Z ≤3) (Table 5.14.). While the other Central European populations have no significant difference.

Table 5.14. *D-statistic test to measure introgression from the European auroch population into Neolithic populations, in comparison to early the Neolithic Serbian population. GBR= Britain, IRE= Ireland, NED=Netherlands, SRB= Serbia, DEU = Germany, POL= Poland. Negative D and Z-score indicates more shared derived alleles between Pop2 and Pop3, positive indicates more shared derived alleles between Pop2 and Pop4. See Appendix Table 10. for definition of ancient populations. Z-scores are coloured to indicate significance level: green= P ≤ 0.001, orange = 0.001 < P ≤ 0.01, red = 0.01 < P ≤ 0.05.*

Pop3	Pop4	*D*-stat	stdr error	*Z*-score	BABA	ABBA	Sites
D(Pop4,Neolithic SRB)auroch,Gaur))							
Neolithic SRB	Neolithic TUR	-0.0316	0.0054	-5.869	5790	6168	96081
	Neolithic BUL	-0.0208	0.0062	-3.354	5649	5889	96065
	Neolithic DEU KIR4	-0.0124	0.0076	-1.63	6165	6320	96074
	Neolithic POL 1	-0.0075	0.0078	-0.962	3760	3816	62991
	Neolithic DEU	0.0013	0.0081	0.164	5832	5817	95935
	Neolithic NED	0.0081	0.0075	1.084	5137	5054	80585
	Neolithic IRE 2	0.0258	0.0094	2.732	3160	3002	48235
	Neolithic IRE 1	0.0197	0.0056	3.491	6244	6002	96086
	Neolithic POL 2	0.0579	0.0136	4.262	1750	1559	25092
	Neolithic GBR 1	0.0282	0.0063	4.468	6341	5993	95749
	Neolithic GBR 3	0.0382	0.0068	5.65	6438	5965	95804
	Neolithic GBR 2	0.0544	0.0062	8.785	6552	5876	95426

3) *Do Neolithic samples have a greater number of shared auroch derived alleles than the Neolithic Irish population from Newgrange?*

The earlier continental Neolithic populations of Anatolia, Serbia, Bulgaria and Poland all have significantly fewer (*Z* ≤3) shared auroch derived alleles than the Irish Neolithic population from Newgrange (IRE 1) (Table 5.15). The Late Neolithic British populations from Durrington Walls (GBR 2) and Ness Brodgar (GBR 3) contain significantly (*Z* ≤3) more shared derived alleles than the Irish Newgrange population. There is no significant difference between the two Irish Neolithic populations, or the Neolithic British population from Bishops Cannings (GBR 1).

Table 5.15. *D-statistic test to measure introgression from the European auroch population into Neolithic populations, in comparison to early the Neolithic Irish population. GBR= Britain, IRE= Ireland, NED=Netherlands, SRB= Serbia, DEU = Germany, POL= Poland. Negative D and Z-score indicates more shared derived alleles between Pop2 and Pop3, positive indicates more shared derived alleles between Pop2 and Pop4. See Appendix Table 10. for definition of ancient populations. Z-scores are coloured to indicate significance level: green= P ≤ 0.001, orange = 0.001 < P ≤ 0.01, red = 0.01 < P ≤ 0.05.*

Pop3	Pop4	*D*-stat	stdr error	*Z*-score	BABA	ABBA	Sites
D(Pop4,Neolithic IRE 1)auroch,Gaur))							
Neolithic IRE 1 (Newgrange)	Neolithic TUR	-0.0492	0.0052	-9.441	6010	6632	96276
	Neolithic BUL	-0.0396	0.0061	-6.449	5891	6376	96260
	Neolithic KIR4	-0.0313	0.0072	-4.376	6143	6541	96273
	Neolithic SRB	-0.0197	0.0056	-3.491	6002	6244	96086
	Neolithic POL	-0.0261	0.0077	-3.388	3825	4030	63091
	Neolithic DEU	-0.0187	0.0079	-2.353	5877	6101	96123
	Neolithic NED	-0.0123	0.0075	-1.649	4945	5069	80711
	Neolithic IRE 2	0.0004	0.008826	0.049	2865	2862	48296
	Neolithic POL 2	0.0187	0.012817	1.455	1699	1637	25129
	Neolithic GBR 1	0.009	0.005752	1.57	6087	5978	95936
	Neolithic GBR 3	0.0188	0.006114	3.074	6214	5985	95994
	Neolithic GBR 2	0.035	0.00562	6.233	6305	5878	95612

4) *Do Neolithic samples have a greater number of shared auroch derived alleles than the Late Neolithic British population from Durrington Walls?*

The general trend is for a significant decrease (*Z* ≤3) in the number of shared auroch derived alleles in the test population when compared to the Durrington Walls samples

(GBR 2) (Table 5.16). Only the Late Neolithic Polish samples share a comparable number of auroch derived alleles.

Table 5.16. *D-statistic test to measure introgression from the European auroch population into Neolithic populations, in comparison to early the Neolithic Anatolian (Turkey) population. GBR= Britain, IRE= Ireland, NED=Netherlands, SRB= Serbia, DEU = Germany, POL= Poland. Negative D and Z-score indicates more shared derived alleles between Pop2 and Pop3, positive indicates more shared derived alleles between Pop2 and Pop4. See Appendix Table 10. for definition of ancient populations. Z-scores are coloured to indicate significance level: green= P ≤ 0.001, orange = 0.001 < P ≤ 0.01, red = 0.01 < P ≤ 0.05.*

Pop3	Pop4	*D*-stat	stdr error	Z-score	BABA	ABBA	Sites
D(Pop4,Neolithic GBR 2)auroch,Gaur))							
Neolithic GBR 2 (Durrington Walls)	Neolithic TUR	-0.0825	0.0061	-13.551	5825	6872	95606
	Neolithic BUL	-0.0736	0.0067	-11.01	5754	6669	95589
	Neolithic SRB	-0.0544	0.0062	-8.785	5876	6552	95426
	Neolithic KIR4	-0.0645	0.0074	-8.664	5987	6813	95600
	Neolithic POL	-0.0614	0.0079	-7.817	3774	4267	62768
	Neolithic DEU	-0.0538	0.0084	-6.44	5794	6452	95470
	Neolithic IRE1	-0.035	0.0056	-6.233	5878	6305	95612
	Neolithic NED	-0.0494	0.0080	-6.21	4882	5390	80203
	Neolithic GBR 1	-0.0268	0.0066	-4.058	5911	6237	95273
	Neolithic IRE 2	-0.0382	0.0095	-4.036	2937	3170	48001
	Neolithic GBR 3	-0.017	0.0068	-2.492	6015	6223	95334
	Neolithic POL 2	0.002	0.0129	0.155	1656	1650	25022

5.3.3.1.2. European Auroch Introgression Through a Time Transect of The Atlantic Edge

The post-Neolithic individual samples from the Atlantic Edge, have a greater number of shared derived European auroch alleles when compared with the Anatolian Neolithic sample Sub1 (Fig 5.8). Multiple *D*-statistic tests, comparing different populations through the transect show differing temporal and geographical patterns of auroch admixture within the Atlantic Edge. These tests are used to answer two specific research questions.

1) *Is there is an increase in auroch introgression during the Bronze Age?*

Within the Netherlands and Irish time series there is an increase (*Z*-score>2) in shared derived alleles in the Bronze Age populations when compared to the Neolithic populations (Table 5.17.). The British Bronze Age populations show no significant difference in the number of shared derived auroch alleles with the Late Neolithic samples (Neolithic GBR 3) from Ness of Brodgar, Orkney. Although an increase in shared derived auroch alleles is demonstrated in the British Bronze Age when comprared to Early Neolithic Britain. Moreover, there is a decrease when they are compared with the Late Neolithic site of Durrington Walls, although the significance of this low.

Table 5.17. *D-statistic test to measure introgression from the European auroch population into the Bronze Age populations of the Atlantic Edge, in comparison to the Neolithic populations. GBR= Britain, IRE= Ireland, NED=Netherlands Negative D and Z-score indicates more shared derived alleles between Pop2 and Pop3, positive indicates more shared derived alleles between Pop2 and Pop4. SeeAppendix Table 10. for definition of ancient populations. Z-scores are coloured to indicate significance level: green= P ≤ 0.001, orange = 0.001 < P ≤ 0.01, red = 0.01 < P ≤ 0.05.*

Pop3	Pop4	*D*-stat	stdr error	*Z-score*	BABA	ABBA	Sites
D(Bronze, Neolithic)auroch,Gaur)							
Neolithic GBR 2	Bronze GBR 1	-0.0123	0.00551	-2.231	6066	6218	95612
Neolithic GBR 2	Bronze GBR 2	-0.0117	0.00641	-1.826	5849	5988	92458
Neolithic GBR 3	Bronze GBR 2	0.0032	0.00702	0.463	6007	5966	92803
Neolithic GBR 3	Bronze GBR 1	0.0041	0.00589	0.694	6197	6146	95993
Neolithic IRE 2	Bronze IRE	0.0073	0.01012	0.72	2966	2923	48199
Neolithic GBR 1	Bronze GBR 2	0.013	0.00639	2.2032	6019	5865	92745
Neolithic GBR 1	Bronze GBR 1	0.014	0.00574	2.447	6221	6049	95935
Neolithic NED (Sch1)	Bronze NED	0.0205	0.00815	2.514	5075	4871	77076

2) *Is there is an increase in auroch introgression post-Bronze Age?*

In post-Bronze Age Britain there are no significant increases in shared aurochs derived alleles. A general trend of a decrease in derived alleles when compared with Neolithic British populations is observed (Table 5.18). Conversely within the the Irish and Netherlands times series there is an increase in shared derived alleles with time, an indication of admixture.

Table 5.18. *D-statistic test to measure introgression from the European auroch population into the post-Bronze Age populations of the Atlantic Edge. GBR= Britain, IRE= Ireland, NED=Netherlands Negative D and Z-score indicates more shared derived alleles between Pop2 and Pop3, positive indicates more shared derived alleles between Pop2 and Pop4. See Appendix Table 10. for definition of ancient populations. Z-scores are coloured to indicate significance level: green= P ≤ 0.001, orange = 0.001 < P ≤ 0.01, red = 0.01 < P ≤ 0.05.*

Pop3	Pop4	*D*-stat	stdr error	Z-score	BABA	ABBA	Sites
D(Pop4, Pop3)auroch,Gaur)							
Neolithic GBR 2	Medieval GBR	-0.0228	0.00562	-4.058	6056	6339	95592
Neolithic GBR 2	Iron GBR 1	-0.0193	0.00512	-3.766	6031	6268	95607
Neolithic GBR 2	Iron GBR 2	-0.0179	0.00605	-2.961	5998	6216	95048
Bronze GBR 1	Medieval GBR	-0.0109	0.00476	-2.288	6105	6239	96263
Bronze GBR 2	Medieval GBR	-0.0102	0.00582	-1.754	5895	6017	93037
Neolithic GBR 2	Iron GBR 3	-0.0148	0.00845	-1.751	3645	3754	57726
Bronze GBR 1	Iron GBR 1	-0.0068	0.00401	-1.695	6095	6178	96280
Bronze GBR 2	Iron GBR 1	-0.0067	0.00506	-1.325	5897	5976	93049
Bronze GBR 1	Iron GBR 2	-0.0064	0.00541	-1.183	5992	6068	95695
Iron GBR 3	Medieval GBR	-0.0088	0.00827	-1.064	3663	3729	58051
Bronze GBR 2	Iron GBR 2	-0.0059	0.00611	-0.965	5825	5894	92546
Iron GBR 1	Medieval GBR	-0.0041	0.00457	-0.897	6149	6200	96259
Iron GBR 2	Medieval GBR	-0.0051	0.00592	-0.862	6065	6127	95677
Iron NED	Roman NED	-0.004	0.00597	-0.67	6113	6162	94837
Iron NED	Medieval NED	-0.0034	0.00620	-0.548	6112	6154	94809
Bronze GBR 1	Iron GBR 3	-0.0007	0.00761	-0.092	3594	3599	58059
Bronze NED	Medieval NED	0.0004	0.00648	0.062	5957	5953	91714
Bronze NED	Roman NED	0.0005	0.00590	0.088	5962	5956	91738
Roman NED	Medieval NED	0.0006	0.00488	0.123	6254	6247	96248
Neolithic GBR 1	Medieval GBR	0.0034	0.00602	0.568	6246	6203	95919
Bronze GBR 2	Iron GBR 3	0.0071	0.00869	0.815	3580	3530	56335
Neolithic GBR 1	Iron GBR 2	0.0081	0.00637	1.271	6123	6025	95360
Neolithic GBR 1	Iron GBR 1	0.0071	0.00532	1.331	6235	6147	95932
Neolithic GBR 1	Iron GBR 3	0.0152	0.00924	1.65	3748	3636	57892
Bronze IRE	Medieval IRE	0.0096	0.00540	1.776	5991	5878	95922
Neolithic NED (Sch1)	Iron NED	0.0201	0.00829	2.428	5204	4998	79586
Neolithic NED (Sch1)	Medieval NED	0.0187	0.00758	2.461	5332	5137	80686
Neolithic NED (Sch1)	Roman NED	0.0188	0.00708	2.654	5325	5128	80712
Neolithic IRE 1	Medieval IRE	0.0191	0.00485	3.937	6085	5857	96159

5.3.3.2. Testing for European Auroch Admixture Using the f_3-statistic

5.3.3.2.1. Neolithic Europe

The *f_3-statistic* results detect auroch admixture occuring in only a few of the admixture scenarios tested (Table 5.19). The only observed significant values for an admixed population were for the British Neolithic populations of Bishops Cannings (Neolithic GBR 1) and Neolithic Durrington Walls (Neolithic GBR 2). Interestingly when calculating the possibility of auroch admixture in the Durrington Walls samples, admixture is still significantly detected when the second source population is Bishops Cannings.

Table 5.19. *Testing auroch admixture scenarios using the f_3-statistic (Target; Source1, Source2). Only significant results are shown. For full results the f_3-statistic for auroch introgression see Appendix Table 11. Z-scores are coloured to indicate significance level: green= $P \leq 0.001$, orange = $0.001 < P \leq 0.01$, red = $0.01 < P \leq 0.05$.*

Source 1	Source2	Target	F3	Std. err	Z score	SNP no.
auroch	Neolithic BUL	Neolithic GBR 2	-0.019022	0.003328	-5.717	45373
auroch	Neolithic SRB	Neolithic GBR 2	-0.01028	0.003271	-3.143	48054
auroch	Neolithic BUL	Neolithic GBR 1	-0.009216	0.004107	-2.244	43237
auroch	Neolithic POL	Neolithic GBR 2	-0.026659	0.003614	-7.376	28655
auroch	Neolithic POL	Neolithic GBR 1	-0.017182	0.004315	-3.982	26636
auroch	Neolithic DEU (Hxh1)	Neolithic GBR 2	-0.024044	0.004024	-5.976	41099
auroch	Neolithic DEU (Hxh1)	Neolithic GBR 1	-0.014032	0.004671	-3.004	38099
auroch	Neolithic NED (Sch1)	Neolithic GBR 2	-0.012874	0.004125	-3.121	34713
auroch	Neolithic NED (Sch1)	Neolithic GBR 1	-0.009947	0.004728	-2.104	31980
auroch	Neolithic GBR 1	Neolithic GBR 2	-0.013905	0.003404	-4.086	45292

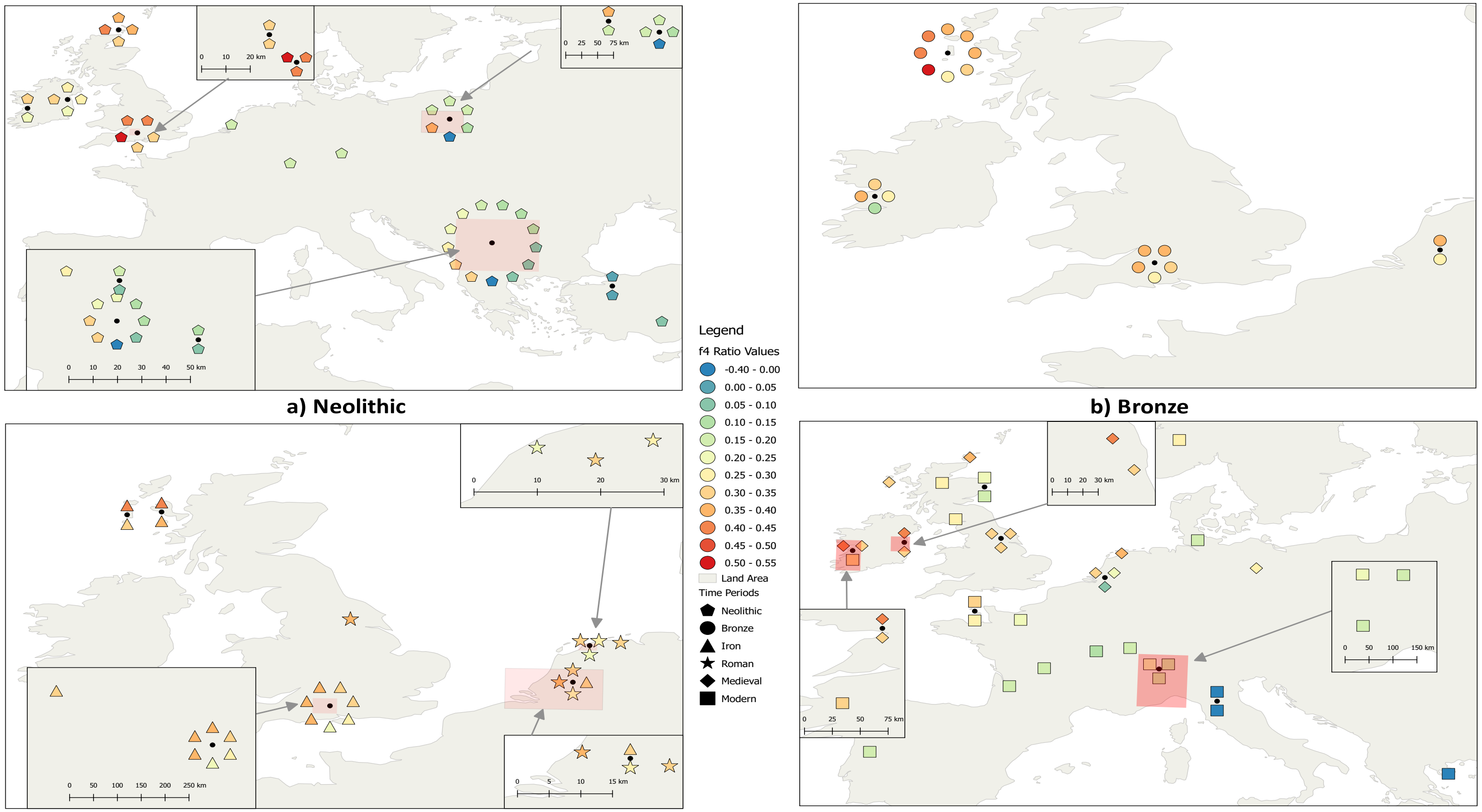
5.3.3.2.2. Atlantic Edge

Admixture is observable using the admixture f_3-statistic for some scenarios of Post-Neolithic introgression (Table 5.20), with results comparable to the *D* results noted above. Admixture was detected between auroch and Neolithic Irish cattle to give rise to Bronze and Medieval Irish cattle. Although negative results were achieved for admixture within the Netherlands time series, none of the results are significant. In contrast no admixture could be detected within the Post-Neolithic British time series (Appendix Table 12.)

Table 5.20. *Testing auroch admixture scenarios within the Atlantic Edge time series using the f_3-statistic (Target; Source1, Source2). Only significant results are shown, for full results the f_3-statistic for auroch introgression see Appendix Table 12. Z-scores are coloured to indicate significance level: green= $P \leq 0.001$, orange = $0.001 < P \leq 0.01$, red = $0.01 < P \leq 0.05$.*

Source 1	Source2	Target	F3	Std. err	Z score	SNP no.
auroch	Neolithic IRE 1	Bronze IRE	-0.014622	0.003014	-4.852	51630
auroch	Neolithic IRE 2	Bronze IRE	-0.0153	0.00443	-3.454	21528
auroch	Neolithic IRE 2	Medieval IRE	-0.010499	0.003619	-2.901	25196

5.3.3.2. Testing for Proportions of Auroch Admixture Using the f_4-ratio statistic



a) Neolithic

b) Bronze

c) Iron and Roman

d) Medieval and Modern

Fig 5.9. *Admixture proportion of auroch in domestic cattle genomes as tested by the f4-ratio, using the previously published Mesolithic auroch CPC (Park et al. 2001) and the German Mesolithic auroch Bed3. Test was set-up as: F4(OGR,CPC;X,SUB1)/f4(OGR,CPC;BED,SUB1) with OGR = Gaur and SUB1 = Neolithic sample from Anatolia. The test demonstrates an increase in auroch proportions within the genomes with distance from the Near East, with the highest value recorded in the Late Neolithic in Britain. Gradually overtime the auroch proportion in the genome decreases.*

5.3.3.2.1. Neolithic Europe

Calculations of admixture proportions of European auroch in the ancient domestic cattle, using the f_4-ratio statistic (Fig 5.9 & Appendix Table 13), show a pattern of increasing proportion of European auroch admixture the further the distance from Anatolia. The highest proportions are seen in the British samples, in particular the Late Neolithic Durrington Walls samples.

5.3.3.2.2. Atlantic Edge

In general through the post-Neolithic British timeline the European auroch proportions, as calculated by the f_4 ratio, reduced through time (Fig 5.9 & Appendix Table 13). The highest proportions are in the Late Neolithic period followed closely by some of the Late Bronze samples then a decrease through later periods. In contrast, both the Irish and Netherlands timelines show auroch proportions increase through time. With Neolithic samples generally showing lower proportions than the Bronze Age, Medieval, Iron Age and Roman samples.

5.3.3.3. Moderns breeds

Modern cattle populations also show an increase in shared derived auroch alleles and auroch proportions with distance from Anatolia (Fig 5.8 & 5.9 & Appendix Table 13). The Kerry, Galloway, Highland and Jersey breeds share more derived European auroch alleles when compared to the Neolithic Anatolian sample Sub1. In comparison to the early and late British Neolithic populations all modern breeds have statistically significant fewer derived auroch alleles, however when compared to Neolithic Irish population there is no significant difference (Appendix Tables 14,15&16).

5.3.4. Pattern of Taurine Introgression into Aurochs

The proportion of auroch in the Medieval Germany auroch Fal7, and Kalisz2 of unknown age, is reduced when compared to the Mesolithic and Neolithic aurochs (Fig 5.10 & Appendix Table 13).

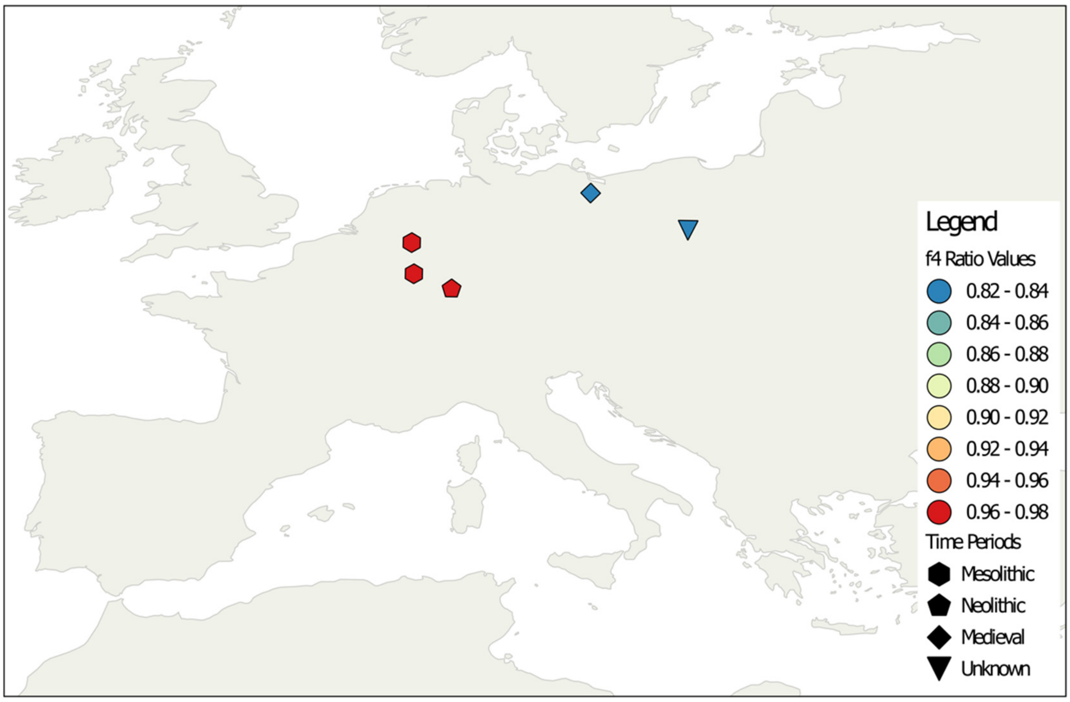


Fig 5.10*. Admixture proportion of auroch in auroch genomes as tested by the f4-ratio, using the previously published Mesolithic auroch CPC (Park et al. 2001) and the German Mesolithic auroch Bed3. Test was set-up as: F4(OGR,CPC;X,SUB1)/f4(OGR,CPC;BED,SUB1) with OGR = Gaur and SUB1 = Neolithic sample from Anatolia. The test demonstrates that two samples have lower proportions of auroch in their genomes.*

5.3.5. Demography - Runs of Homozygosity

The ROH analysis (Fig 5.11) indicates different demographic histories between the wild European auroch population and domestic cattle. Futhermore, the anaylsis demonstrates different demographic histories between the domestic cattle. Long runs of ROH are indicative of recent consanguinity/inbreeding, whilst short runs of ROH indicate restriction in breeding population further back in the ancestry of the individual.

The two German auroch genomes (Bed3 & Hxh2) display fewer short ROH than the domestic samples. Both have a relative excess of long ROH, with the Neolithic Hxh2 possessing a higher mean total of long ROH than the Mesolithic Bed3 sample. The majority of the Neolithic domesticates follow a similar pattern, however the Anatolian Sub1 sample has both shorter and fewer ROH than the other European Neolithic samples. The Early Neolithic samples from Bulgaria (Dzh1) and Germany (Hxh1) contain the highest mean total of short ROH runs of the Neolithic samples, while the Irish sample (New1) is intermediate between this Bulgarian/German group and the Anatolian sample. The Late Neolithic German sample Kir4 is an anomaly in comparison to other Neolithic samples, behaving similarly to the auroch samples in the short ROH range and but lacking long ROH, similar to the Neolithic genomes.

The British Bronze Age sample, Cla1, possess fewer stretches of short ROH than the Neolithic samples, however it possesses the highest proportion of of long ROH. The two Roman samples show somewhat different patterns of ROH to each other, Win1 from the Netherlands boasts a similar pattern to Kir4, with low numbers of short ROH when compared to other domesticates and low to no long ROH. The British Roman sample Yor1, shows a somewhat different pattern with a large number of short ROH, but again has little to no long ROH.

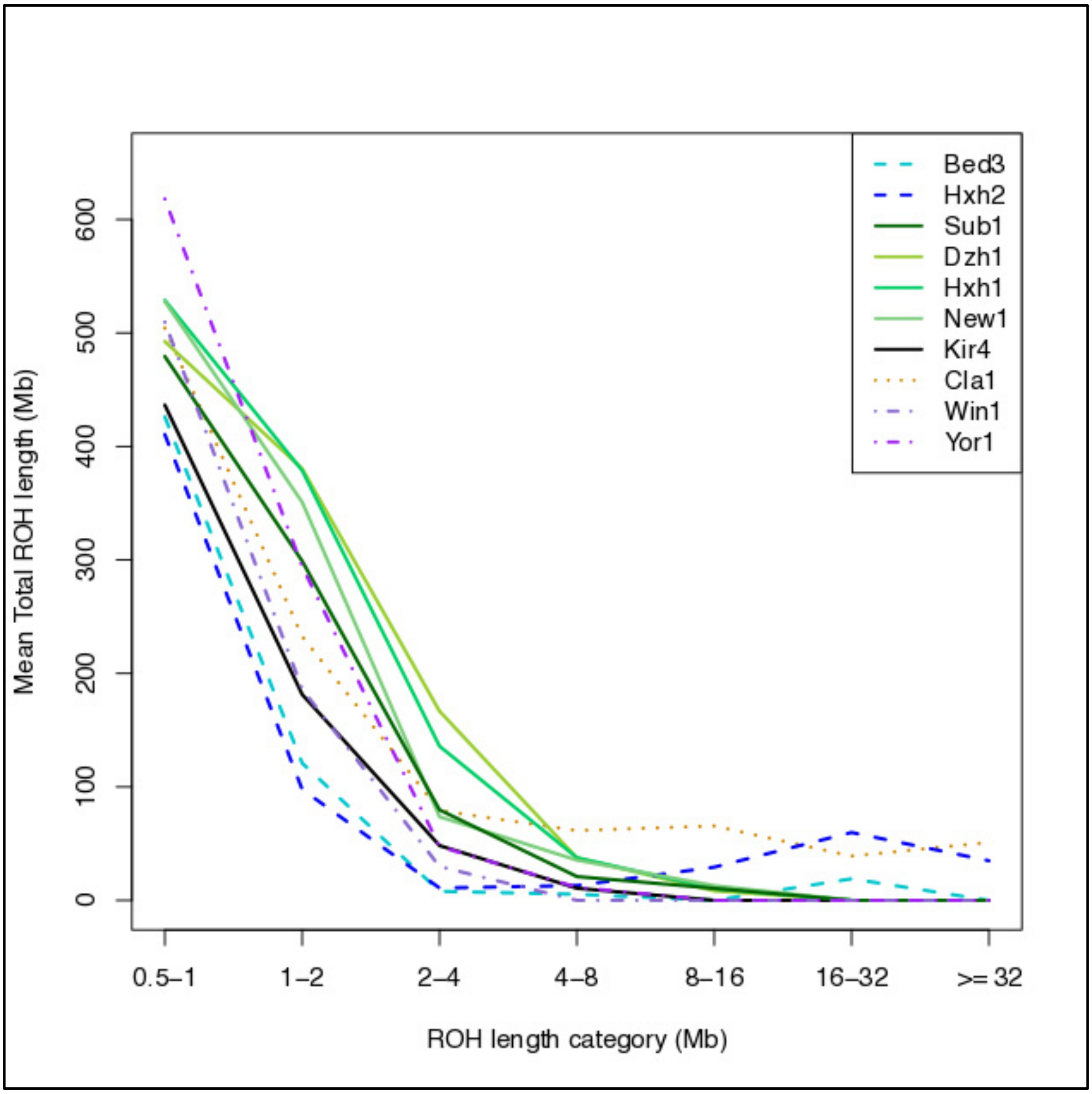


Fig 5.11. *Runs of Homozygosity of 10 ancient individual genomes demonstrating different demographic histroies. Blue broken line indicates B. primigenius samples, Green solid line indicates Neolithic genomes, Black solid line indicates outlier Neolithic genome, Orange dotted line indicates Late Bronze Age genome, Purple dotted/broken line indicates Roman genomes. Patterns of ROH differ for the different time periods and between the domesticated and wild Bos, suggesting different demographic histories.*

5.3.6. Evidence for Selection for Milk Production Traits

One of the most important genes in the modern dairying industry is Diglyceride acyltransferase (DGAT1) with variants strongly linked to milk production traits in multiple breeds (Grisart et al. 2002). Allele presence through time was explored (Table 5.21) for the SNPs rs109234250 and rs109326954. These two SNPs, side by side, code for the (K232A) missense mutation within the DGAT1 gene on Chr14. This mutation is associated with differing milk yields, with the derived (GC) alleles associated with higher milk yields (Grisart et al. 2002; Spelman et al. 2002; Kaupe et al. 2004).

The European auroch and the majority of the Neolithic samples have wild type DGAT1 alleles.The derived alleles first appear in the Serbian Neolithic in the sample Plo2 dated by phase to 5200-5000 BC. The next samples to have the alternative alleles present are samples Lud1 and Kir4, a Polish Early Neolithic and a German Late Neolithic sample respectively. Only one Bronze Age sample, Viz2 from the Netherlands, has the the alternative alleles present. The alternate alleles are not detected again until the Iron Age, when it is present in the Early Iron Age samples from the British site Danebury. In contrast to the earlier time periods, the majority of the Roman and Medieval samples have the derived alleles.

Table 5.21. *Allele presence of the SNPs rs109234250 and rs109326954 in the DGAT1 gene which are strongly linked to milk production traits in multiple cattle breeds. The wild type alleles are AA, while the derived are GC. An "X" denotes presence of alleles at the two sites which are in linkage disequilibrium. Therefore, if "X" is present in only AA column then only A alleles were present at the two sites. If "X" is denoted in both columns, then the sample is likely heterozygote at both sites.*

Time Period	Sample Name	AA	GC	Time Period	Sample Name	AA	GC
Bos primigenius				Bronze ~ 2400-800 BC	Viz2		X
Mesolithic	Bed3	X			Cla1	X	
	Bed4	X			Cla5	X	
	Tri1	X			Cla8	X	
	CPC	X			Cla9	X	
Neolithic	Hxh2	X			Pot5	X	
Medieval	Fal7	X			Rou1	X	
unknown	Kalisz2	X			Rou2	X	
Bos taurus				Iron/ Roman 800 BC - 450 AD	Pot4	X	
Neolithic ~ 6200-2500 BC	Men1	X			Cla6	X	
	Sub1	X			Da5	X	X
	Men2	X			Da1	X	X
	Dzh1	X			Da3		X
	Dzh2	X			Da6		X
	Plo2		X		Bor1	X	
	Plo4	X			Bun2	X	
	Lud1		X		Yor1		X
	Hxh1	X			New2	X	X
	Kie2	X			Ste1		X
	Bis1	X			Fis1	X	
	Bis2	X			Fis2		X
	Sch1	X			Els1		X
	New3	X			Win1		X
	New4	X			Dro1	X	X
	New1	X			Bri1		X
	New5	X			Otb1		X
	Kir4	X	X	Medieval 450 AD - 1400 AD	Fir1		X
	Par1	X			Hou1		X
	Par2	X			Dub1	X	
	Dur1	X			Yor2		X
	Dur2	X			Yor6	X	
	Ness2	X			Dyr1		X
	Ness4	X			Yor4	X	X
	Ness5	X			Yor12		X
					Sir1	X	X
					Hou2		X
					Yor3		X
					Bal1	X	
					Bal2	X	

5.4. Discussion

5.4.1. Geographic Structure

The domestication centre for the *Bos taurus* is supported by archaeological (Helmer et al. 2005) and genetic evidence (Troy et al. 2001; Edwards et al. 2007; Scheu et al. 2015). While some argue for independent domestication events in Europe (Beja-Pereira et al. 2006), the consensus is the primary origin lies in the Fertile Crescent in Southwest Asia, in concordance with with other ungulate domesticates, such as goats and sheep (Zeder 2011; Arbuckle 2014). A later domestication of *Bos primigenius* in the Indus valley, modern day Pakistan, resulted in *Bos indicus* (Loftus et al. 1994). The geographic structure of the genetic diversity of modern cattle populations from both types of cattle is exhibited in the PCA and ADMIXTURE and concurs with previous research (Matukumalli et al. 2009; Decker et al. 2014; Park et al. 2015; Sempéré et al. 2015; Chung et al. 2017). Demonstrating the differentiation between the taurine and the indicine sub species and supporting the hypothesis of separate domestication centres (Loftus et al. 1994).

5.4.1.1. Neolithic

With the addition of ancient samples to the PCA and ADMIXTURE analysis the separate domestication centre hypothesis (Loftus et al. 1994; Murray et al. 2010) is further supported. The position of the Neolithic Anatolian samples above the European samples on the PCA, indicates a close genetic relationship. In ADMIXTURE analysis, k=3, while the Anatolian samples have a small *B. indicus* component, likely due to ancestral variation, the major component is that shared with modern European *B. taurus*. If a shared domestication of *B. taurus* and *B. indicius* had occurred, then there would be an expectation for ancient Anatolian Neolithics to have a greater indicine component, and to be positioned to the middle of PCA, similar to modern Anatolian animals which all exhibit admixture.

A contrary view is that this result is not at odds with a single domestication hypothesis. Some argue domestication should refer only to the initial processes and any subsequent admixture between wild populations, however large, is a different process (Larson & Burger 2013). For example, a possible scenario could be a small population from Anatolia founded the Indus valley domestic population. The modern indicine variation being a consequence of many admixture events from a differentiated local auroch population followed by drift or selection (Larson & Burger 2013). Nonetheless it has been demonstrated that the taurine and indicine populations have different domestication histories (Murray et al. 2010) and these results shown here suggest a closer relationship of ancient Anatolian cattle with taurine as opposed to indicine cattle. The only definitive test would be to sequence nuclear data from early ancient cattle and aurochsen from Asia. The results shown here indicate that the indicine introgression in modern breeds from Anatolia, such as the South-Anatolian Red, is post-domestication and plausibly post Neolithic.

Movement of taurine cattle from the Near East into Europe is believed to have started approximately 6,400BC. Genetic evidence has shown movement between the Europe and Near East likely continued to occur until 5,000 BC (Scheu et al. 2015). The placement of the Neolithic samples on the PCA and ADMIXTURE components show a stepwise progression, echoing the migration of cattle from the Near East through to the Atlantic Edge islands of Britain and Ireland. The relevant migration path is the Danubian route (Zeder 2008) as the samples analysed are from either Eastern or Northern Europe. The exception may be the Irish Neolithic samples as the Irish Neolithic may have southern influences as suggested by human ancient genomics (Cassidy et al. 2016). Differences in modern sheep populations have been attributed to different routes of domestication (Kijas et al. 2012). However this hypothesis is difficult to test here due to the lack of ancient Iberian or Southern European samples.

The clustering of the Neolithic samples in the PCA, outgroup f_3-statistic and *D*-statistics can be equated to both geographical distribution and division between the early and later Neolithic. The Neolithic spread rapidly across mainland Europe, but halted before spreading into Britain and Ireland approximately 4,000 BC (Whittle et al. 2011). The Anatolian samples group closer on the PCA, and share more drift, to Southeast Europe, as opposed to Northern Europe. Somewhat in agreement with evidence from mtDNA of continued contact between Europe and Southwest Asia (Scheu et al. 2015).

Testing for clade integrity with *D*-statistics suggests the Bulgarian samples are closest to the Anatolian samples in concordance with time and geography, present day Bulgaria and Turkey share a border. Neolithic groups such as the Serbian, Polish and German are closer to each other than to Neolithic Anatolia, again all tracking the pattern on the PCA, geography and age. The high shared drift between the Atlantic Neolithic samples, and the cluster on the PCA away from the other European samples, shows a closer genetic relationships between these samples than others. It may also be indicative of a bottleneck in genetic diversity occurring due to genetic isolation from the continental populations.

Perhaps this difference may also suggest that Atlantic Edge population was influenced by the Mediterranean routes as the Neolithic humans were (Cassidy et al. 2016).

5.4.1.2. Atlantic Edge

Within the Atlantic edge cluster the differentiation between the Irish grouping and the British/Netherlands grouping on the PCA can be detected in both *D*-statistics and f_3 outgroup tests. Within the Irish cluster the highest shared drift is between other Irish Neolithic samples. However, within the British cluster this is not always so. The early Neolithic samples from Bishops Cannings show a higher degree of shared drift with the Irish and Netherlands Neolithic samples, rather than with the later Neolithic British samples such as the Durrington Walls or the Ness of Brodgar. This is additionally illustrated by *D*-statistic results where the German Neolithic sample Hxh1 breaks the clade of the Irish Neolithic samples and the later British Neolithic population, but not the earlier Bishops Cannings population. A significant positive or negative *D*-score is generally associated with an increase in shared derived alleles of the introgressing population into one of the two populations in a clade. However it is also possible that a significant value could be due to admixture from a ghost population (Patterson et al. 2012), perhaps auroch admixture (Park et al. 2015; Upadhyay et al. 2016) within the British population. What is apparent on the PCA is a difference in genetic affinity between the British and the Irish populations, perhaps due to different partial origins of the cattle, such as due to migration routes. Another possibility could be the occurrence of an additional or a more severe population bottleneck when cattle were brought to Ireland.

Samples from the Ness of Brodgar demonstrated the highest values of shared drift in the outgroup f_3-statistics. These were in fact the highest values in any of the tests, possibly suggesting that these samples were a herd and therefore related. All these samples were found in structure 10 of the Neolithic temple in the Ness of Brodgar complex. The skulls are believed to have been purposefully placed around the structure during an event that saw 400 cattle slaughtered and feasted upon and the temple decommissioned. To investigate the origins of these cattle, strontium isotope analysis was performed for six cattle and suggested the cattle grazed on Orkney (Ingrid Mainland *per. comm.*). The outgroup f_3-statistic supports this isotope analysis, suggesting these cattle were local as opposed to cattle from a number of different locations in Britain or Europe.

Accompanying the Bronze Age in Europe was a second large scale human migration from the Eurasian Steppe. Herders linked to cultures such as the Yamnaya have been shown to have substantially introgressed into the Neolithic populations of Northern Europe

(Allentoft et al. 2015; Haak et al. 2015; Cassidy et al. 2016; Olalde et al. 2017). The time series of cattle on the Atlantic Edge highlights a degree of continuity as opposed to population replacement in Britain and Ireland. The lack of a Bronze Age cluster on the PCA, as shown in the Irish ancient human timeseries (Cassidy et al. 2016), the high shared drift between the Irish Neolithic and the Irish/British Bronze samples, as opposed to high shared drift between all the Bronze samples, indicate no large population turnover of during the Bronze Age in this part of Europe. However, this does not preclude admixture.

While a large genetic influx from the continent is unlikely, some form of population movement did occur between Britain and Ireland during the Late Neolithic or Bronze Age. The PCA indicates a homogenization of the two island populations, which had formed separate clusters during the Neolithic. Shared drift measured by outgroup f_3 statistic tests, indicate a more recent relationship between British Bronze samples and Irish Neolithic compared to the British Neolithic. Moreover, f_3 admixture results indicate the Irish Bronze samples approximate best as the result of admixture between the Irish Neolithic and a population akin to the Neolithic British, Bronze British or Bronze Netherlands. This result in itself may be masking the addition of a “ghost” population, likely *B. primigenius*, which will be discussed later. However what it does suggest is a new genetic component into the cattle of the Irish Bronze Age.

When using the *D*-statistic to test for clade integrity the Irish Neolithic and Bronze Age cluster excludes introgression of the Bronze Netherlands samples. However mirroring the outgroup f_3 results the British Neolithic and Bronze Age clade integrity breaks in numerous situations. Perhaps these results are due to admixture from the European auroch in the later Neolithic samples which is interrupting the statistic (Patterson et al. 2012). Alternatively movement of cattle and subsequent introgression into Britain could explain these results. The archaeological record would support such a homogenisation in population and movement of cattle during the Bronze Age. The Bronze Age was a period of trade and interaction across the Atlantic network, the Bell Beaker culture present across Europe developed a high level of connectivity, while the distribution of seafaring artifacts indicate a strong importance of maritime culture across this Atlantic Zone (Cunliffe 2013). Many parallels can be drawn between the cultures of Britain and Ireland during this period, such as, hoard depositions and burial rites, suggesting high interconnectivity between the islands (Mallory 2013).

It should be mentioned that within the Netherlands time series there is a slight difference in the position of the Neolithic and Bronze samples, the Neolithic sample sits closer to some Roman and the British Neolithic samples, whilst the Bronze Age, Iron Age and Medieval samples site closer to the continental Neolithics. It is possible that this is indicative of admixture between the Neolithic sample and an unknown population, or perhaps the more recent samples have had influence from the East. However, it is also possible that the single Neolithic sample from the Netherlands is not representative of the Neolithic population of the Netherlands. Perhaps the Roman samples that cluster towards the British Neolithics may be admixed and descended from imports of British cattle, Win1 for example shows higher drift values with British samples than with some of its contemporaries.

Post-Bronze Age a pattern of continuity remains within the British/Irish group and to a degree within the Netherlands. The Iron Age samples through to the Medieval period show high shared drift values within geographical regions rather than within time period. However, it is possible that during the Iron Age there was interaction and admixture between cattle from Britain and Netherlands. Whilst the integrity of the *D*-statistic clades of the different British Iron Age hold, the Netherlands Iron Age samples can break the clades of the British Bronze and British Iron. Showing perhaps a continued relationship between the island and the continental Atlantic Seaboard, however archeology describes an early Iron Age period of isolation for both Britain and Ireland, although this improves through the Middle and Late Iron Age when the maritime connections were re-established (Cunliffe 2013)

Britain as an island was notably invaded and settled by the Romans, the Anglo Saxons, the Vikings and the Normans, all of which are believed to have contributed to the genetic diversity of the islands. The Romans and Anglo Saxons did not invade or settle within Ireland, however Vikings raids were common and lead to the establishment of major cities such as Dublin (Cunliffe 2013). The invasion of the Normans and later the arrival of the English and Scottish planters are likely to have also contributed to the genetic diversity to the island (Gilbert et al. 2017). Modern human genetic studies have demonstrated an Anglo-Saxon component in South England, however it is argued that these post-Roman migrations have had less effect on genetic composition of Britain than previously thought (Leslie et al. 2015).

In addition to modern human data, the study of Romano British samples from York indicates continuity with British Iron Age but demonstrates genetic differences between these Romano British samples from the modern day humber estuary (Martiniano et al. 2016). Considering the human populations of the islands have been shaped by post Neolithic migrations (Leslie et al. 2015; Cassidy et al. 2016; Martiniano et al. 2016; Olalde et al. 2017) it is interesting that, at least within the context of modern cattle variation, there are no obvious time period groupings on the PCA. This may suggest that while humans may have migrated, they did not necessarily move cattle on a large scale. Perhaps in the context of ancient variation some difference may become apparent, or a more fine-scale approach such as a combination of chromopainter and fine structure (Lawson et al. 2012) may be necessary to tease out such differences in the British and Irish populations. However to enable the use of these programmes samples need to have the coverage to be called diploid.

To improve this regional time series a greater number samples from all cultural periods would be beneficial. However, in particular additional Neolithic animals from the Netherlands, Early Bronze Age animals from all three countries discussed and more early AD animals from the UK and Ireland would be benefical to track changes in population variation. Additionally, Neolithic and Bronze samples from Iberia would help to distinguish if differences between Neolithic Ireland and Britain are due to input from Iberia.

5.4.1.3. Bos primigenius - Auroch

The placement of the European aurochs on the PCA and the composition of ancestry components at K3 in ADMIXTURE analysis is similar to previous analysis of the British auroch specimen CPC using both the 50K (Park et al. 2015) and the 770K (Upadhyay et al. 2016) SNP array variation. Unlike the European taurine modern population there is little variation within the European auroch population, likely due to the lack of auroch-specific and rare variation present in the 770K SNP array. Therefore variation between these temporally and geographically diverse samples cannot be established. However, it is interesting that the Fal7 a more recent Medieval sample, sits closer to the European taurine animals than the other aurochsen. Additionally the reduced f_4 ratio value in comparison to earlier European aurochs is perhaps suggetive of taurine admixture.

5.4.1.4. Kir4 - An Anomaly?

The late German Neolithic sample Kir4 is somewhat of an anomaly. On the PCA rather than sitting with an early Neolithic German sample Hxh1 it sits between the Anatolian cluster and the European cluster, while its mitochondrial haplotype is an outgroup to other European *Bos* species (Chapter 4). Compared with other ancient domestic and wild samples the proportion of European auroch within Kir4s genome does not suggest that this is a sample with recent admixture from European auroch population analysed here. Other mitochondrial haplotypes of European auroch have been identified in Europe and Asia during the Neolithic and Mesolithic (Edwards et al. 2007; Lari et al. 2011; Zhang et al.

2013), and perhaps Kir4 is a relict of another auroch population, or more likely the descendent of a female auroch enveloped into domestic herd at some point as has been suggested for previous P haplotypes in domestic cattle (Schibler et al. 2014). Bone deposits from the Kirschbaumhöhle cave include human and animal remains, with late Neolithic human remains believed to be connected with Corded Ware Culture. Domestic animals formed the majority of the animal assemblage as opposed to wild animals (Seregély 2014). Human aDNA studies have shown that the Corded Ware culture to be genetically similar to the Yamnaya steppe herders that migrated through Europe at the end of Neolithic/beginning of the Bronze Age (Haak et al. 2015). Perhaps Kir4 is descended from cattle that moved from the steppe as part of this human migration, but whose genetic signature has since been lost or diluted.

5.4.2. Demography

There are various methods available to analyse and understand demography of high coverage genomes. Programs such as PSMC (Li & Durbin 2011) and MSMC (Schiffles and Durbin 2014) enable the reconstruction of demographic histories using a hidden Markovian framework applied to coalescent theory. These two methods have been used in a variety of analyses including; inferring past populations size of humans populations, woolly mammoths and horses (Li & Durbin 2011; Orlando et al. 2013; Palkopoulou et al. 2015), as well as attempting to time dog domestication (Frantz et al. 2016). However, these approaches are limited in accuracy for computing more recent demographic history (last 10,000 years) (Li & Durbin 2011). One technique to assess more recent demographic change is to examine runs of homozygosity (ROH), which has been used to identify differences in demographic histories between both individuals and populations in both cattle and humans (Purfield et al. 2012; Gamba et al. 2014; Cassidy et al. 2016; Hofmanová et al. 2016).

The results of ROH analysis of ten high coverage *Bos* genomes identifies different demographic histories between wild and domestic populations. This difference between the Neolithic taurine samples and the wild auroch samples, a large number of small ROH in the domesticates compared to auroch, makes sense, as it would be expected that the domestic population underwent some degree of bottleneck during the domestication process. Previous mitochondrial analysis has suggested a domestication occurred from a small effective population size of 80 females (Bollongino et al. 2012),

Differences in small ROH between the early European and Anatolian Neolithic samples, are possibly due to a population bottleneck that may have occurred during the migration from Anatolia into Europe. However, mitochondrial analysis suggested a lack of bottleneck during this movement (Scheu et al. 2015). The late Neolithic sample Kir4, as previously discussed, is an anomaly in many analyses and remains one here. A pattern of short ROH similar to the wild aurochsen and a pattern of long ROH similar to Neolithic domesticates, suggests this sample has a different demographic history to the other wild and domesticate genomes. Perhaps, as previously mentioned, this sample does have ancestry linked with the introgression of a wild female into a domestic herd.

Interestingly, the Bronze Age sample from Cladh Hallan shows a very different pattern of ROH from all other samples. Cladh Hallan is a Bronze Age settlement on the island of South Uist, part of the Outer Hebrides, with an interesting archaeology of Late Bronze Age roundhouses, human mummification and animal burials (Pearson et al. 2005; Mulville et al. 2011). This sample has a lower amount of short ROH than the Neolithic samples, likely due to the breakdown of the short ROH present in the Neolithics by recombination, therefore leaving all later samples with fewer short runs. However, compared with all other samples, this has the highest number of longer ROH, suggesting substantial recent inbreeding within an enclosed population. While the site is believed to show secondary metalworking (Niall Sharples *per. comm.*) and therefore is in connection with the Atlantic Bronze Age, it would make sense that large bulky animals such as cattle were not regularly transported to the island.

The two Roman samples Win1 and Yor1 have similar recent demographic histories with no or few long ROH, however there is a difference in earlier demographic history. Excepting Kir4, the Netherland sample Win1 has the fewest runs of ROH in its genome suggesting it belonged a population that was large in size. The British Yor1 sample has the highest amount of short ROH of all the samples, signifying historical population restriction relative to the other post Neolithic genomes.

5.4.3. *Bos primigenius* Introgression

Post-domestication admixture between domestic livestock and populations of the wild progenitor have been shown to occur in pigs (Larson et al. 2007). Previous mitochondrial and Y chromosome studies remained inconclusive regarding European auroch introgression into domestic cattle (Gotherstrom et al. 2005; Beja-Pereira et al. 2006; Edwards et al. 2007; Achilli et al. 2008; Bollongino et al. 2008). However with the whole genome sequencing of the British Mesolithic auroch CPC, introgression of European auroch into modern European cattle breeds could be tested (Park et al. 2015). It was shown that British unimproved breeds, such as Highland and Galloway, contained more auroch derived alleles than commercial breeds (Park et al. 2015) and other continental unimproved breeds (Upadhyay et al. 2016).

5.4.3.1. Neolithic

As the range of the auroch spanned Eurasia and North Africa (Van Vuure 2005), the key question in the Neolithic was whether any of the ancient animals demonstrate more European P auroch alleles than the Anatolian samples, P being the most common haplogroup for European auroch. Results from the *D*-statistics with Neolithic domestics mirror that of modern cattle (Park et al. 2015; Upadhyay et al. 2016), the highest *D* and *Z*-scores are generally seen in the British Neolithic, although the late Polish Neolithic also show high *D* values, but poor coverage reduces the significance of these. The British Neolithic samples contain the highest proportion of auroch in their genome, as calculated by f_4-ratios. Broadly, the level of European auroch introgression increases the greater the distance from the centre of domestication. This increase in introgression is likely due to increased contact with European auroch as domestic cattle migrated across Europe.

Previous studies using modern cattle and a single auroch specimen have suggested that auroch introgression occurred more than once (Park et al. 2015; Upadhyay et al. 2016). Through comparing the number of shared auroch derived alleles in different populations, by swapping the Anatolian samples for the different populations in turn, the results show that in the Neolithic it is likely introgression did occur more than once and at different points in the migratory journey across Europe. This pattern may simply be due to the increased opportunity for contact with different P mtDNA auroch populations with the progression of domesticates initially across the European continent (Edwards et al. 2007; Mona et al. 2010; Schibler et al. 2014).

The Late Neolithic British population from The Ness of Brodgar (NessX) and Durrington Walls (DurX) show an increase in auroch alleles compared to British Early Neolithic samples and Middle/Late Neolithic Irish samples, as well as an increase in f_4-ratio values. Indeed Durrington Walls show the highest reliable levels of auroch introgression from the Neolithic samples analysed. Therefore while the cattle that first entered Britain and Ireland had more derived auroch alleles than the Anatolian and other European samples, further auroch introgression also likely occurred within Britain.

The high levels shown in the Durrington Walls sample Dur2 may be a local phenomenon, however the high f_4-ratios in other British samples suggests it is indicative of the British population. Durrington Walls, the largest Henge monument in Britain, is situated in the southwest of Britain near Stonehenge. Both are believed to be part of the same complex (Pearson & Ramilisonina 1998; Pearson et al. 2007). A feasting site (Craig et al. 2015; Albarella & Serjeantson 2002) dominated by pig followed by cattle contains few remains identified as auroch (Albarella et al. 2011). Previous stable isotope analysis has indicated that cattle were brought from all over Britain, as far as Orkney (Viner et al. 2010), therefore these cattle tested from Durrington Walls are not necessary local to the area. However, other analysis such as PCA suggest a British origin. The increase in auroch introgression in Britain could be due to wild auroch bulls interacting with domestic herds, as all samples show domestic cattle mitochondrial haplotypes (Chapter 4). Perhaps intentional breeding could have been influenced by population collapse of the domestic population. The collapse of the Neolithic in Britain and Ireland has been suggested by some, but is usually linked more to crops than animals (Stevens & Fuller 2012; Whitehouse et al. 2014).

Thus, the results indicate an increasing trend of introgression with distance from Anatolia and with a decrease in age. The Neolithic arrived in Britain and Ireland over 5,000 years after its beginnings in the Near East. The oldest British cattle sample (Bis2) is approximately 3,000 years younger than Anatolian sample Sub1. While the samples from Bulgaria and Serbia are a maximum of 2,000 years younger than Sub1. It is plausible that in areas such as the Baltic region that through time auroch introgression did increase, but without the samples this can not be explored.

These results do not preclude introgression occurring from a different auroch population in samples closer to Anatolia which may have been genetically similar to the initial domesticates. Morphological analysis has demonstrated differences in size variation linked to both geographical and temporal parameters (Wright & Viner-Daniels 2015) as well as differences in horn shape (Van Vuure 2005). Additionally, mtDNA analysis has indicated different Eurasian *B. primigenius* haplogroups (Edwards et al. 2007; Edwards et al. 2010; Zhang et al. 2013) and differences in the ranges of different auroch populations (Mona et al. 2010) perhaps linked to refugia during the Last Glacial Maximum and/or the Younger Dryas event. The population genomics of European and Southwest Asian aurochs has so far not been explored to its full potential, therefore a complete understanding of introgression from the wild into domestic cattle is not achievable at this present time.

5.4.3.2. The Atlantic Edge

The local population history of auroch differs in each of the three countries analysed. The species did not inhabit Ireland, was extirpated in Britain by approximately 1,500 BC, and lasted until 4th century AD in the Netherlands (Van Vuure 2005). In the British time series no samples show a greater proportion of auroch introgression than the late Neolithic Durrington Walls sample. The Bronze Age samples have similar f_4-ratio proportions to the Neolithic samples. Additionally, there is no significant difference in derived auroch alleles between the Bronze Age cattle and the cattle from the late Neolithic Orcadian site of The Ness of Brodgar. The overall trend is a post-Neolithic decrease of auroch introgression in Britain, an expected trend when the local population of auroch was extirpated in the Bronze Age (Van Vuure 2005).

In comparison to Britain, Ireland and the Netherlands show a different pattern. Post-Neolithic samples show an increase in derived auroch alleles and larger proportions of auroch within their genome than the Neolithic samples. In addition, the f_3-statistics detect admixture between auroch and the Irish Neolithic samples in the Irish Bronze Age samples; a similar result is not achieved for the Netherlands Bronze Age samples. The increase present in the Irish samples cannot be due to introgression with auroch, but suggests admixture with a population with greater levels of auroch introgression. This links with the homogenisation of the British and Irish cattle communities post-Neolithic, as previously discussed.

Within the Netherlands samples it is clear there is an increase in auroch introgession between the Neolithic and the Bronze Age. However, samples from post-Bronze Age show similar size auroch components to the Bronze Age, suggesting that auroch introgression did not make a significant impact on the cattle population post-Bronze Age. As populations of auroch declined in post Neolithic Europe, due the reduction of habitat (Van Vuure 2005), encounters between domestic and wild auroch would have declined.

This time series of the Atlantic Edge is not complete, therefore pinpointing exact points in time of introgression is difficult. Nor may one sample or site be taken to be representative of a population of an entire land mass such as Britain or Ireland. However, it is clear that levels of European P auroch introgression differed through geography and time.

5.4.4. Dairying

In the present day cattle are bred and reared for three different commodities; meat, traction and milk. Lipid analysis of archaeological pots has indicated milk processing as early as 7th millennium BC Anatolia and South East Europe (Evershed et al. 2008); early European Neolithic cheese strainers indicate processing of milk into lactose-low products (Salque et al. 2013). In present day Western European human populations lactase persistence, the ability to drink raw milk into adulthood, is prevalent. However through the study of ancient human genomes it has become apparent that the major spread of the lactase persistence allele occurred in the Bronze Age as opposed to the Neolithic (Gamba et al. 2014; Allentoft et al. 2015; Cassidy et al. 2016). Therefore, during the Neolithic it is likely that processed milk products such as cheese would have been consumed by adults as opposed to raw milk.

The majority of variation analysed in modern cattle for milk phenotypic traits are quantitative trait loci; minor constituents of the cumulative effect of multiple genes on a phenotype. However, one particular missense mutation (K232A) in the DGAT1 gene on Chr14 has a significant effect on milk fat and milk yield. The two wild alleles AA code for lysine which is associated with high fat content and lower milk yields, while the derived alleles of GC code for alanine which is associated with lower fat content, but higher protein and milk volume yields. (Grisart et al. 2002; Spelman et al. 2002; Kaupe et al. 2004; Schennink et al. 2007; Schennink et al. 2008; Berry et al. 2010; Schopen et al. 2011). Fixation of the ancestral type alleles has been noted for some modern indicine and buffalo breeds, while modern taurine breeds show differences in allele frequencies. Beef breeds such as Belgium Blue and Hereford show fixation of the derived alleles, on the other hand dairy breeds such as British Friesians show high allele frequencies of the derived alleles but not fixation, possibly due to more recent selection for fat content (Kemper et al. 2014). While the Jersey Breed famed for its creamy milk has both alleles present in the population but with the ancestral at a higher frequency (Kaupe et al. 2004; Tantia et al. 2006).

The results for the allele presence of the DGAT1 mutation indicates that both European auroch and Anatolian Neolithic samples were wild type, helping to support the argument that the DGAT1 mutation spread due to selection for increase milk yields post-domestication (Kaupe et al. 2004). This time series indicates that the DGAT1 mutation is an ancient mutation, likely older than 5000 BC. The first sample to show allele presence of the mutation is Plo2, from the site Plocknick in Serbia and has an estimated date of 5,200-5,000BC, but confirmation by a direct radiocarbon date is required. It is apparent that the derived allele was also present in the Neolithic and Bronze populations likely at low frequencies. The allele presence then increases in the Iron Age, possibly indicating selection for higher milk yields. It is interesting that the increases in allele frequency of the derived allele occur post the appearance of lactose tolerance in ancient European human genomes. This suggests that once adult humans were able to drink raw milk cattle with a greater milk yield may have been selected for. Modelling allele trajectories to pinpoint selection, as published for chickens (Loog et al. 2017), would be a desirable next step.

5.4.5. Improvements and Limitations

SNP array data is limited by ascertainment bias, usually a bias toward the breed or breeds in which the variance was discovered (Kijas et al. 2009; Matukumalli et al. 2009). The Illumina Bovine HD Genotyping chip (770K) used a more diverse set of SNP discovery breeds, both taurine and indicine, than previous Bovine chips, hoping to overcome this issue.

While the 770K SNP dataset is useful for understanding ancient samples in the context of modern variation, this does not account for genetic variation present in the ancient samples which has subsequently been lost. Future analysis should focus on sequencing a greater number of ancient samples to a greater coverage to allow for both diploid calls and SNP discovery. These genomes could then be integrated with the increasing number of published modern cattle genomes (Daetwyler et al. 2014; Kim et al. 2017). Another approach would be to implement genome wide analysis through ANGSD where genotype likelihoods are utilised as opposed to genotype calls (Korneliussen et al. 2014), however to ensure inflation of likelihoods do not occur there should be a minimum target genome coverage of 2x.

The majority of statistics used here were performed using pseudodiploid data, taking one read and duplicating it to create a diploid call. While this technique helps to circumnavigate the issues of low coverage data, it simplifies the genetic diversity within an individual and population as no heterozygotes are present. Additionally, it likely biases towards the reference genome allele (Martiniano et al. 2016). While there are some programs written to attempt to alleviate this problem, such as Atlas (Link et al. 2017) and ANGSD (Korneliussen et al. 2014), there may be a greater issue of reference bias being introduced at the alignment stage (Gopalakrishnan et al. 2017; Cahill et al. 2017). Reads with a greater number of mismatches due to damage and true variation are less likely to align to the reference genome and may be discarded, consequently a larger proportion of the reads left behind will contain reference alleles as opposed to alternate alleles.

Moreover it is unknown what the effects of aligning auroch samples to a *B. taurus* has on subsequent analysis, possibly losing variation present in the *B. primigenius* population. A recent study from (Gopalakrishnan et al. 2017) aligned wolf samples to both a *de novo* wolf genome and the current dog genome demonstrating that genomic coverages, number of SNPs and heterozygosity were dependent on the reference genome used. However, they concluded that the majority of evolutionary genomic analyses yielded similar results, and that choice of reference genome should be dependent on the aims of the study. These issues do not negate the results presented here, however a greater understanding of cattle domestication and auroch may be gained with an improvement in methodology and perhaps a *de novo B. primigenius* genome to act as a reference genome. Understanding the biases introduced by both alignments and SNP calling will be a challenge that the aDNA field will have to confront going forward.

In addition to the ancient data analysed here, more ancient data is necessary to fully understand cattle domestication. In particular pre-domesticate samples from Southwest Asia would enable selection scans of wild versus early domestic animals. As it is believed that the Neolithic traveled through Europe via two different routes, the Danubian and Mediterranean pathways (Zeder 2008), the Mediterranean route needs to be sampled. Domestic and wild samples from Italy, France and Iberia would enable a greater understanding of the population genetics of auroch, the of dispersal of domestic cattle and interaction of domestic herds with wild populations. Post-Neolithic samples from central and Southeastern Europe are necessary to understand if the migrations of the steppe herders did affect domestic cattle herds in areas of the continent.

5.4.6. Conclusion

To conclude, the data presented in this study supports the long held view of taurine cattle domestication in Southwest Asia and a subsequent migration into Europe. The genetic differentiation of the Neolithic cattle, likely caused by a number of population bottlenecks and European auroch admixture events, echoes their migration from the Near East through to the Atlantic Edges. Post-Neolithic populations in the Atlantic Edge demonstrate a greater geographical affinity than temporal, differing from human populations genetics from this region (Cassidy et al. 2016; Olalde et al. 2017). Admixture with European auroch contributed significantly to the genomes of the ancient cattle, likely several different admixture events occurred with highs of introgression demonstrated in Neolithic Britain, a pattern still evident in the genomes of modern breeds today (Park et al. 2015). Preliminary results of allele frequencies suggest selection for increase milk yields occuring post the appearance of lactose tolerance in European humans (Gamba et al. 2014). Future analysis will focus on whole genome analysis to achieve a greater understanding of the genetic diversity and selection for phenotypic traits.

5.5. References

Achilli, A. et al., 2008. Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Current biology: CB*, 18(4), pp.R157–8.

Achilli, A. et al., 2009. The multifaceted origin of taurine cattle reflected by the mitochondrial genome. *PloS one*, 4(6), p.e5753.

Albarella, U. et al., 2011. Feeding Stonehenge: Feasting in Late Neolithic Britain. In U. Albarella & A. Trentacoste, eds. *EthnoZooarchaeology: the Present and Past of Humananimal Relationships.* Oxbow Books, Oxford, pp. 143–159.

Albarella, U. & Serjeantson, D., 2002. A passion for pork: meat consumption at the British Late Neolithic site of Durrington Walls. *Consuming passions and patterns of consumption*, pp.33–49.

Alexander, D.H., Novembre, J. & Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, 19(9), pp.1655–1664.

Allentoft, M.E. et al., 2015. Population genomics of Bronze Age Eurasia. *Nature*, 522(7555), pp.167–172.

Arbuckle, B.S., 2014. Pace and process in the emergence of animal husbandry in Neolithic Southwest Asia. *Bioarchaeol. Near East*, 8, pp.53–81.

Beja-Pereira, A. 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), pp.8113–8118.

Berry, D.P. et al., 2010. Associations between the K232A polymorphism in the diacylglycerol-O-transferase 1 (DGAT1) gene and performance in Irish Holstein-Friesian dairy cattle. *Irish Journal of Agricultural and Food Research*, 49(1), pp.1–9.

Bollongino, R. et al., 2006. Early history of European domestic cattle as revealed by ancient DNA. *Biology letters*, 2(1), pp.155–159.

Bollongino, R. et al., 2012. Modern taurine cattle descended from small number of near-eastern founders. *Molecular biology and evolution*, 29(9), pp.2101–2104.

Bollongino, R. et al., 2008. Y-SNPs do not indicate hybridisation between European aurochs and domestic cattle. *PloS one*, 3(10), p.e3418.

Bovine Genome Sequencing and Analysis Consortium et al., 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, 324(5926), pp.522–528.

Bovine HapMap Consortium et al., 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science*, 324(5926), pp.528–532.

Bradley, D.G. et al., 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), pp.5131–5135.

Cahill, J.A. et al., 2017. Genomic evidence of globally widespread admixture from polar bears into brown bears during the last ice age. *bioRxiv*, p.154773. Available at: https://www.biorxiv.org/content/early/2017/06/25/154773 [Accessed October 21, 2017].

Cassidy, L.M. et al., 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), pp.368–373.

Chung, N.C. et al., 2017. Population Structure Analysis of Bull Genomes of European and Western Ancestry. *Scientific reports*, 7, p.40688.

Craig, O.E. et al., 2015. Feeding Stonehenge: cuisine and consumption at the Late Neolithic site of Durrington Walls. *Antiquity*, 89(347), pp.1096–1109.

Cunliffe, B., 2013. *Britain Begins*, OUP Oxford.

Daetwyler, H.D. et al., 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature genetics*, 46(8), pp.858–865.

Decker, J.E. et al., 2014. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS genetics*, 10(3), p.e1004254.

Edwards, C.J. et al., 2010. A complete mitochondrial genome sequence from a mesolithic wild aurochs (Bos primigenius). *PloS one*, 5(2), p.e9255.

Edwards, C.J. et al., 2007. Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1616), pp.1377–1385.

Evershed, R.P. et al., 2008. Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. *Nature*, 455(7212), pp.528–531.

Frantz, L.A.F. et al., 2016. Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science*, 352(6290), pp.1228–1231.

Gamba, C. et al., 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nature communications*, 5, p.5257.

Gilbert, E. et al., 2017. The Irish DNA Atlas: Revealing Fine-Scale Population Structure and History within Ireland. *Scientific reports*, 7(1), p.17199.

Gopalakrishnan, S. et al., 2017. The wolf reference genome sequence (Canis lupus lupus) and its implications for Canis spp. population genomics. *BMC genomics*, 18(1), p.495.

Gotherstrom, A. et al., 2005. Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe. *Proceedings of the Royal Society B: Biological Sciences*, 272(1579), pp.2345–2351.

Grisart, B. et al., 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome research*, 12(2), pp.222–231.

Haak, W. et al., 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. Available at: http://dx.doi.org/10.1038/nature14317.

Helmer, D. et al., 2005. Identifying early domestic cattle from Pre-Pottery Neolithic sites on the Middle Euphrates using sexual dimorphism. In J. D. Vigne, J. Peter, & D. Helmer, eds. *The First Steps of Animal Domestication*. Oxbow Books, Oxford, pp. 86–95.

Hofmanová, Z. et al., 2016. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proceedings of the National Academy of Sciences of the United States of America*, 113(25), pp.6886–6891.

Jones, E.R. et al., 2015. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nature communications*, 6, p.8912.

Kaupe, B. et al., 2004. DGAT polymorphism in Bos Indicus and Bos Taurus cattle breeds. *Journal of Dairy Research*, 71, pp.182–187.

Kemper, K.E. et al., 2014. Selection for complex traits leaves little or no classic signatures of selection. *BMC genomics*, 15, p.246.

Kijas, J.W. et al., 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PloS one*, 4(3), p.e4668.

Kijas, J.W. et al., 2012. Genome-wide analysis of the world’s sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS biology*, 10(2), p.e1001258.

Kim, J. et al., 2017. The genome landscape of indigenous African cattle. *Genome biology*, 18(1), p.34.

Korneliussen, T.S., Albrechtsen, A. & Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC bioinformatics*, 15, p.356.

Lari, M. et al., 2011. The complete mitochondrial genome of an 11,450-year-old aurochsen (Bos primigenius) from Central Italy. *BMC evolutionary biology*, 11, p.32.

Larson, G. et al., 2007. Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences of the United States of America*, 104(39), pp.15276–15281.

Larson, G. & Burger, J., 2013. A population genetics view of animal domestication. *Trends in genetics: TIG*, 29(4), pp.197–205.

Lawson, D.J. et al., 2012. Inference of population structure using dense haplotype data. *PLoS genetics*, 8(1), p.e1002453.

Leslie, S. et al., 2015. The fine-scale genetic structure of the British population. *Nature*, 519(7543), pp.309–314.

Li, H. & Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), p.nature10231.

Link, V. et al., 2017. Atlas: analysis tools for low-depth and ancient samples. *bioRxiv*, p.105346.

Loftus, R.T. et al., 1994. Mitochondrial genetic variation in European, African and Indian cattle populations. *Animal genetics*, 25(4), pp.265–271.

Loog, L. et al., 2017. Inferring allele frequency trajectories from ancient DNA indicates that selection on a chicken gene coincided with changes in medieval husbandry practices. *Molecular biology and evolution*. Available at: http://dx.doi.org/10.1093/molbev/msx142.

MacLeod, I.M. et al., 2013. Inferring Demography from Runs of Homozygosity in Whole Genome Sequence, with Correction for Sequence Errors. *Molecular biology and evolution*. Available at: http://mbe.oxfordjournals.org/content/early/2013/07/10/molbev.mst125.abstract.

Mallory, J.P., 2013. *The origins of the Irish*, Thames & Hudson.

Mannen, H. et al., 1998. Mitochondrial DNA variation and evolution of Japanese black cattle (Bos taurus). *Genetics*, 150(3), pp.1169–1175.

Martiniano, R. et al., 2016. Genomic signals of migration and continuity in Britain before the Anglo-Saxons. *Nature communications*, 7, p.10326.

Matukumalli, L.K. et al., 2009. Development and characterization of a high density SNP genotyping assay for cattle. *PloS one*, 4(4), p.e5350.

McKenna, A. et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), pp.1297–1303.

Mona, S. et al., 2010. Population dynamic of the extinct European aurochs: genetic evidence of a north-south differentiation pattern and no evidence of post-glacial expansion. *BMC evolutionary biology*, 10, p.83.

Mulville, J. et al., 2011. Flesh on the bones: animal bodies in Atlantic roundhouses. Available at: http://eprints.bournemouth.ac.uk/18428/1/RitualFINALDRAFTHOPEFULLY.pdf.

Murray, C. et al., 2010. Cattle demographic history modelled from autosomal sequence variation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1552), pp.2531–2539.

Niemi, M. et al., 2015. Temporal fluctuation in North East Baltic Sea region cattle population revealed by mitochondrial and Y-chromosomal DNA analyses. *PloS one*, 10(5), p.e0123821.

QGIS Development Team, 2009. QGIS Geographic Information System, Open Source Geospatial Foundation. Available at: http://qgis.osgeo.org.

Olalde, I. et al., 2017. The Beaker phenomenon and the genomic transformation of northwest Europe. *bioRxiv*, p.28. Available at: http://repository.nms.ac.uk/1864/ [Accessed October 17, 2017].

Orlando, L. et al., 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, 499(7456), pp.74–78.

Palkopoulou, E. et al., 2015. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Current biology: CB*, 25(10), pp.1395–1400.

Park, S. et al., 2015. Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and evolution of cattle. *Genome Biology*, 16(1), p.234. Available at: http://genomebiology.com/2015/16/1/234.

Patterson, N. et al., 2012. Ancient admixture in human history. *Genetics*, 192(3), pp.1065–1093.

Pearson, M.P. et al., 2005. Evidence for mummification in Bronze Age Britain. *Antiquity*, 79(305), pp.529–546.

Pearson, M.P. et al., 2007. The age of Stonehenge. *Antiquity*, 81, pp.617–639.

Pearson, M.P. & Ramilisonina, 1998. Stonehenge for the ancestors: the stones pass on the message. *Antiquity*, 72, pp.308–326.

Purcell, S. et al., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*, 81(3), pp.559–575.

Purfield, D.C. et al., 2012. Runs of homozygosity and population history in cattle. *BMC genetics*, 13, p.70.

R Core Team, 2015. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Salque, M. et al., 2013. Earliest evidence for cheese making in the sixth millennium BC in northern Europe. *Nature*, 493(7433), pp.522–525.

Schennink, A. et al., 2007. DGAT1 underlies large genetic variation in milk-fat composition of dairy cows. *Animal genetics*, 38(5), pp.467–473.

Schennink, A. et al., 2008. Milk fatty acid unsaturation: genetic parameters and effects of stearoyl-CoA desaturase (SCD1) and acyl CoA: diacylglycerol acyltransferase 1 (DGAT1). *Journal of dairy science*, 91(5), pp.2135–2143.

Scheu, A. et al., 2015. The genetic prehistory of domesticated cattle from their origin to the spread across Europe. *BMC genetics*, 16, p.54.

Schibler, J. et al., 2014. Incorporation of aurochs into a cattle herd in Neolithic Europe: single event or breeding? *ncbi.nlm.nih.gov*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4107343/.

Schiffels, S. & Durbin, R., 2014. Inferring human population size and separation history from multiple genome sequences. Nature genetics, 46(8), pp.919–925.

Schopen, G.C.B. et al., 2011. Whole-genome association study for milk protein composition in dairy cattle. *Journal of dairy science*, 94(6), pp.3148–3158.

Sempéré, G. et al., 2015. WIDDE: a Web-Interfaced next generation database for genetic diversity exploration, with a first application in cattle. *BMC genomics*, 16, p.940.

Seregély, T., 2014. Human and animal remains from three eras. New documentation methods of a vertical cave of the Northern Franconian low mountain range and its inventory. *The European Archaeologist*, (41), pp.17–19.

Skoglund, P. et al., 2012. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science*, 336(6080), pp.466–469.

Speller, C.F. et al., 2013. Ancient mtDNA Analysis of Early 16th Century Caribbean Cattle Provides Insight into Founding Populations of New World Creole Cattle Breeds M. T. P. Gilbert, ed. *PloS one*, 8(7), p.e69584.

Spelman, R.J. et al., 2002. Characterization of the DGAT1 gene in the New Zealand dairy population. *Journal of dairy science*, 85(12), pp.3514–3517.

Stevens, C.J. & Fuller, D.Q., 2012. Did Neolithic farming fail? The case for a Bronze Age agricultural revolution in the British Isles. *Antiquity*, 86(333), pp.707–722.

Tantia, M.S. et al., 2006. DGAT1 and ABCG2 polymorphism in Indian cattle (Bos indicus) and buffalo (Bubalus bubalis) breeds. *BMC veterinary research*, 2, p.32.

Troy, C.S. et al., 2001. Genetic evidence for Near-Eastern origins of European cattle. *Nature*, 410(6832), pp.1088–1091.

Upadhyay, M.R. et al., 2016. Genetic origin, admixture and population history of aurochs (Bos primigenius) and primitive European cattle. *Heredity*. Available at: http://dx.doi.org/10.1038/hdy.2016.79.

Van Vuure, C., 2005. *Retracing the aurochs: history, morphology and ecology of an extinct wild ox*, Pensoft Pub.

Viner, S. et al., 2010. Cattle mobility in prehistoric Britain: strontium isotope analysis of cattle teeth from Durrington Walls (Wiltshire, Britain). *Journal of archaeological science*, 37(11), pp.2812–2820.

Wang, C. et al., 2015. Improved ancestry estimation for both genotyping and sequencing data using projection procrustes analysis and genotype imputation. *American journal of human genetics*, 96(6), pp.926–937.

Whitehouse, N.J. et al., 2014. Neolithic agriculture on the European western frontier: the boom and bust of early farming in Ireland. *Journal of archaeological science*, (0). Available at: http://dx.doi.org/10.1016/j.jas.2013.08.009.

Whittle, A.W.R., Healy, F.M.A. & Bayliss, A., 2011. *Gathering Time: Dating the Early Neolithic Enclosures of Southern Britain and Ireland*, Oxford: Oxbow Books.

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Available at: http://ggplot2.org.

Wright, E. & Viner-Daniels, S., 2015. Geographical variation in the size and shape of the European aurochs (Bos primigenius). *Journal of archaeological science*, 54, pp.8–22.

Zeder, M.A., 2008. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the National Academy of Sciences of the United States of America*, 105(33), pp.11597–11604.

Zeder, M.A., 2011. The Origins of Agriculture in the Near East. *Current anthropology*, 52(S4), pp.S221–S235.

Zhang, H. et al., 2013. Morphological and genetic evidence for early Holocene cattle management in northeastern China. *Nature communications*, 4, p.2755.

Chapter 6. Analysis of Ancient Whole Genome Data of *Ovis aries*

6.1. Introduction

Archaeozoological and genetic studies indicate the domestication of sheep occurred in the Fertile Crescent region of Southwest Asia approximately 11,000BP as part of the Neolithic revolution (Lawson Handley et al. 2007; Peter et al. 2007; Zeder 2008; Chessa et al. 2009; Meadows et al. 2011). Domesticated primarily for meat, wool and milk are believed to have been selected for as secondary products (Chessa et al. 2009). Through human mediated movement sheep migrated into Europe, Asia and Africa, possibly through two large migration events, where more primitive initial populations were displaced by improved secondary product populations – as identified by retroviral insertion analysis (Chessa et al. 2009). Over time, selection for different phenotypic traits has resulted in a multitude of breeds specialised for different production qualities and adapted to a range of environments, from the hairy sheep of the Caribbean, the fine wool quality of the Merino to the milking Friesian breeds. Much of the selection for phenotypic traits has occurred over the last 300 years with the development of breeds books, some of which have their origins in Britain (Ryder 1964).

The diversity of sheep has been studied at a morphological level with size, presence of horns and fleece quality forming arguments for how unimproved or primitive a breed is (Ryder 1981; Ryder 1984). Domestication is inferred from size diminution and evidence for human management such as kill-off patterns in assemblages with respect to age and sex (Zeder et al. 2006). More recently genetic studies have added to the understanding of sheep domestication. Mitochondrial analyses have detailed the global dispersal of two major haplogroups with the restriction of another three minor haplogroups discovered in the Southwest Asia tying in with this region as a centre of origin and hence genomic diversity (Meadows et al. 2005; Pereira et al. 2006; Tapio et al. 2006; Meadows et al. 2011; Lv et al. 2015). This increase in genetic diversity in Southwest Asia is replicated in microsatellite data (Lawson Handley et al. 2007; Peter et al. 2007). Heightened genetic diversity in the regions of primary domestication have been identified in other domesticates, such as cattle (Troy et al. 2001). Kijas et al. (2012) produced the first global genome-wide study of sheep breeds, in which 50,000 SNP were analysed to uncover phylogeographic structure in modern breeds, identifying clear genetic divisions between European, Asian and African breeds. Additionally signals of selection spanning the RXFP2 gene, linked to the presence or absence of horns (Johnston et al. 2011), and pigmentation genes such as KIT, ASIP and MITF were identified.

Previous genetic studies have primarily focused on modern mtDNA, retroviral insertions, microsatellite data, Y chromosome, and a variety of SNP datasets (Wood & Phua 1996; Tapio et al. 2006; Meadows et al. 2007; Chessa et al. 2009; Lawson Handley et al. 2007; Peter et al. 2007; Kijas et al. 2009; Meadows & Kijas 2009; Meadows et al. 2011; Kijas et al. 2012). In addition to modern data, ancient DNA studies have published mitochondrial DNA sequences for individuals and populations from Europe and Asia (Olivieri et al. 2012; Niemi et al. 2013; Niemi et al. 2015; Eve Rannamäe et al. 2016). Furthermore, a recent aDNA study on retroviral insertions in the Caucasus regions supports the findings of Chessa and colleagues (2009) of a secondary introduction of wool sheep from the Near East between Late Bronze Age and the Iron Age (Schroeder et al. 2017).

This study presents whole genome data for 19 sheep that constitute a time series of ancient sheep through Britain and Ireland, supplemented with 9 ancient Eurasian samples. Sheep have historically played an important role in the British economy, as non-native species introduced to Britain and Ireland during the Neolithic. They were used for grazing the poor agricultural uplands, their skins were a primary source for British parchment and most importantly their wool was the substrate vital for the English textile industry (Ryder 1964; Fiddyment et al. 2015) . From the Middle Ages to the industrial revolution wool was arguably the most important raw material in the English economic system (Bowden 2013). The aim of this study was to sequence samples from different archaeological periods in order to gain insight into the genetic structure present in ancient samples prior to the formation of the breed books and its variation through time.

6.2. Materials

All sheep samples presented in this chapter were sampled from the same bone element, the petrous portion of the temporal bone. The samples were screened as described in section 2.2.1 and results presented 3.4. In total 19 bones from Britain and Ireland were chosen for further sequencing (Fig6.1)(Table 6.1). This chapter is supplemented with another nine samples from Eurasia, processed by Andrew Hare and Kevin Daly (Table 6.2) (*per. comm.*).

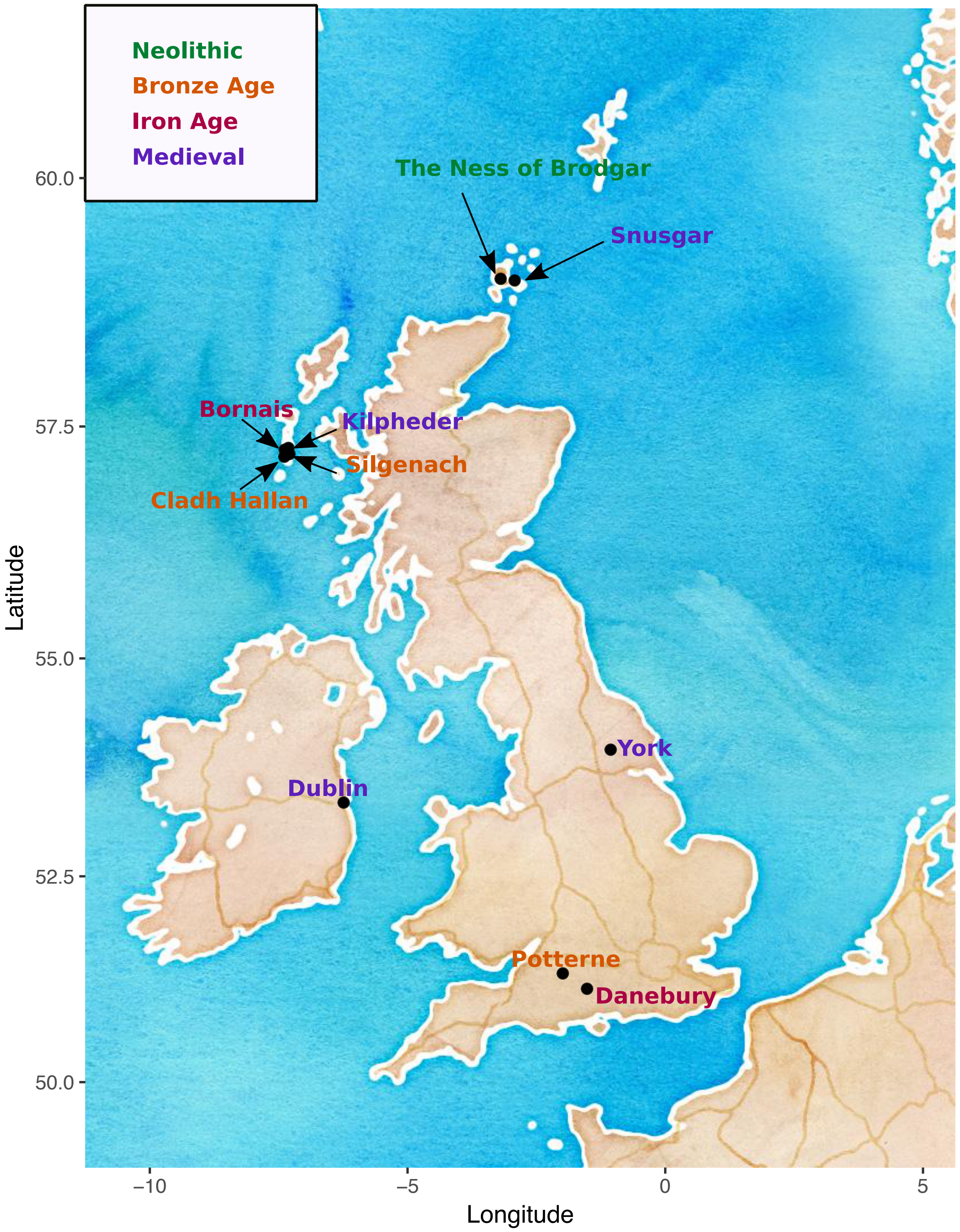


Fig 6.1. *Distribution of the British and Irish sites from which samples in this chapter originate from. Names of the sites are colour coded for predominate cultural period.*

Table 6.1. *Samples from Britain and Ireland analysed in this chapter.*

Name	Cultural Period	Site	Country
Pot7	Late Bronze Age	Potterne	England
Da8	Iron Age	Danebury	England
Da7	Iron Age	Danebury	England
Yor13	Medieval (Norse)	York	England
Dub3	Medieval (Norse)	Dublin	Ireland
Dub5	Medieval (Norse)	Dublin	Ireland
Dub4	Medieval (Norse)	Dublin	Ireland
Ness7	Neolithic	Ness of Brodgar	Scotland (Orkney)
Ness6	Neolithic	Ness of Brodgar	Scotland (Orkney)
Ork1	Medieval (Norse)	Snusgar	Scotland (Orkney)
Ork2	Medieval (Norse)	Snusgar	Scotland (Orkney)
Sil1	Bronze Age	Silgenach	Scotland (South Uist)
Cla12	Bronze Age	Cladh Hallan	Scotland (South Uist)
Cla13	Bronze Age	Cladh Hallan	Scotland (South Uist)
Cla11	Bronze Age	Cladh Hallan	Scotland (South Uist)
Cla10	Bronze Age	Cladh Hallan	Scotland (South Uist)
Cla14	Iron Age	Cladh Hallan	Scotland (South Uist)
Bor2	Iron Age	Bornais	Scotland (South Uist)
Kil2	Medieval (Norse)	Kilpheder	Scotland (South Uist)

Table 6.2. *Samples analysed in this chapter processed by Andrew Hare and Kevin Daly (per. comm.).*

Name	Cultural Period	Site	Country
Dz1	Neolithic	Dzhulyunitsa	Bulgaria
Bub1	Neolithic	Bubanj-Nevo	Serbia
Her8	Neolithic	Herxheim	Germany
Her3	Neolithic	Herxheim	Germany
San1	Iron Age	Sangir Tepe	Uzbekistan
Pad1	Iron Age	Padayatak	Uzbekistan
Geo3	Medieval	Kazbeg	Georgia
Alz1	Roman	Alzey	Germany
Mai2	Roman	Heim	Germany

6.3. Methods

6.3.1. Read Processing

Illumina HiSeq^{TM} reads were processed as described in section 2.2.2.3.

6.3.2. SNP calling and Modern Dataset Merge

To analyse the samples in the context of modern variation SNPs from the curated 50K Hapmap dataset (section 2.2.2.4.3) were called in the samples (section 2.2.2.5.1). Ancient samples were then merged with the modern dataset, which had been previously filtered for a minimum minor allele frequency of 1%, using PLINK (Purcell et al. 2007). Additional filtering of SNP missingness of 2% was applied and a second set of PLINK files with all moderns pseudo-diplodised was created using an in-house script (E. Jones *per. comm.*).

6.3.3. Population Genetic Analysis

6.3.3.1. PCA

In order to visualise autosomal variation among the ancient sheep genomes in the context of the modern breeds, a LASER PCA (Wang et al. 2015) (section 2.2.3.1) was performed and PC1-PC3 plotted using ggplot2 in R (Wickham 2009; R Core Team 2015).

6.3.3.2. Ancestry Estimation

Prior to estimating ancestry fractions, the pseudo-diplodised dataset was filtered for linkage disequilibrium using PLINK option *--indep-pairwise 50 5 0.1.* Ancestry coefficients of K=1-10 were computed using ADMIXTURE (Alexander et al. 2009) (section 2.2.3.2). Each value of K was run ten times. For each K the replicate with the lowest likelihood was chosen to be visualized in R (R Core Team 2015).

6.3.3.3. Shared Genetic Drift

To test for shared drift between samples f_3-outgroup statistics were also performed using AdmixTools (Patterson et al. 2012) (section 2.2.3.4.2), grouping ancient samples into populations where possible. The inbreed option was performed when the target population was an ancient pseudo-diplodised population. Tests were performed using the general formula (C;A,B), where C represents the outgroup of the test, while A and B represent the two test populations.

6.3.3.3. Testing for Admixture

For estimating possible admixture two different statistics from the AdmixTools (Patterson et al. 2012), the *D*-statistic and f_3-statistic, were implemented (sections 2.2.3.3 & 2.2.3.4.1). Results visualized using ggplot2 in R (Wickham 2009; R Core Team 2015), and compiled in table format.

To test for admixture from an Asian or Southwest Asian animal into Europeans several different breeds were used to represent this component. In particular, the modern Asian breeds of Bangladeshi Garole, Deccani and the Southwest Asian breeds of Local Awassi and Afshari sheep were selected due to their geographical origins.

In an attempt to understand gene flow between ancient samples and primitive/Northern European breeds, such as the Scandinavian breeds and some British breeds, tests for admixture were also performed.

6.3.3.3.1. D-statistics

The outgroup for the *D*-statistics was goat (SRA051557) aligned to the sheep genome (OviAri 3.1) by M.Teasdale (*per. comm.).* Diploid calls were generated for the goat genome as described in section 2.2.2.5.2.

Modern samples were left diploid and where possible ancient samples were grouped into populations (Appendix Table 17). The grouping of populations increases the coverage of the genome across the population and therefore helps to improve sensitivity of the statistic. Tests were performed using the general formula (Ref1,Ref2),X,Goat) where X is the introgressor and Ref1 and Ref2 represent ancient and modern samples or populations.

6.3.3.3.2. f_3-statistics

Modern samples were left diploid and where possible ancient samples were grouped into populations. The inbreed:YES option was applied when the target population was a group of ancient samples. Tests were performed using the general formula (C;A,B) where C is the target, while A and B represent the admixing populations.

6.3.4. Mitochondrial Haplogroup Assignment

Mitochondrial reads were extracted from BAM files previously filtered for removal of duplicates and rmdup (section 2.2.2.3). Pileup files were produced using GATK *pileup*

(McKenna et al. 2010) and subsequently filtered to create a consensus fasta file an in house script, using a minimum base quality of 20 and a coverage of two. Consensus fasta files were then run through the programme MitoToolPy (https://github.com/kizcas/MitoToolPy) for haplogroup assignment.

6.4. Results

6.4.1. Overview

The average genome coverage of the 19 sheep sampled from Britain and Ireland ranged from 0.15x to 4.03x, with these individuals originating from a range of geographic and temporal sites (Table 6.3). To contextualise these ancient British and Irish samples both temporally and geographically nine mainland continental European ancient sheep samples are co-analysed, with average genome coverages ranging from 0.18x to 2.12x (Table 6.4).

Table 6.3. *Sample information post HiSeq Sequencing. Nuclear coverage, mitochondrial haplotype and the number of 50K SNPs that were successfully called.*

Name	Cultural Period	Site	Country	Nuclear Coverage	Mitochondrial Haplotype	Number of 50K SNPs
Pot7	Late Bronze	Potterne	England	0.15	B	6874
Da8	Iron	Danebury	England	0.55	B	20426
Da7	Iron	Danebury	England	0.65	B	23370
Yor13	Medieval	York	England	0.62	B	21752
Dub3	Medieval	Dublin	Ireland	0.29	B	12520
Dub5	Medieval	Dublin	Ireland	0.35	B	14981
Dub4	Medieval	Dublin	Ireland	0.75	B	24775
Ness7	Neolithic	Ness of Brodgar	Scotland (Orkney)	0.35	B	15244
Ness6	Neolithic	Ness of Brodgar	Scotland (Orkney)	2.66	B	41965
Ork1	Medieval(Norse)	Orkney	Scotland (Orkney)	0.34	B	15012
Ork2	Medieval(Norse)	Orkney	Scotland (Orkney)	1.21	B	33220
Sil1	Bronze	Silgenach	Scotland (South Uist)	0.3	B	12687
Cla12	Bronze	Cladh Hallan	Scotland (South Uist)	0.39	B	16202
Cla13	Bronze	Cladh Hallan	Scotland (South Uist)	0.41	B	16941
Cla11	Bronze	Cladh Hallan	Scotland (South Uist)	4.03	B	43820
Cla10	Bronze	Cladh Hallan	Scotland (South Uist)	0.21	B	8365
Cla14	Iron	Cladh Hallan	Scotland (South Uist)	0.33	B	14179
Bor2	Iron	Bornais	Scotland (South Uist)	0.26	B	12311
Kil2	Medieval(Norse)	Kilpheder	Scotland (South Uist)	0.52	B	20105

Table 6.4. *Sample information post Hiseq sequencing for samples processed by AH and/or KD. Nuclear coverage, mitochondrial haplotype and the number of 50K SNPs that were successfully called.*

Name	Cultural Period	Site	Country	Nuclear Coverage	Mitochondrial Haplotype	Number of 50K SNPs
Dz1	Neolithic	Dzhulyunitsa	Bulgaria	0.72	B	19571
Bub1	Neolithic	Bubanj-Nevo	Serbia	0.96	B	22243
Her8	Neolithic	Herxheim	Germany	1.83	B	34545
Her3	Neolithic	Herxheim	Germany	2.12	B	37096
Alz1	Roman	Alzey	Germany	1.56	B	33010
Mai2	Roman	Heim	Germany	1.6	B	33432
San1	Iron Age	Sangir Tepe	Uzbekistan	0.19	B	7399
Pad1	Iron Age	Padayatak	Uzbekistan	0.2	A	8304
Geo3	Medieval	Kazbeg	Georgia	0.18	B	7201

6.4.2. Principal Components Analysis

The modern geographic structure of sheep (Kijas et al. 2012) is reproduced in Figure 6.2 (Appendix Fig 4). PC1 separates European breeds from African, Southwest Asia and Asian breeds. Through PC2 a geographic pattern is shown splitting African, Southwest Asian and Southern European derived animals from the extremities of Northern Europe and Asian breeds. On the European arm a stepwise pattern through the mediteran, the Swiss Alps and into the British Isles is shown. PC3 highlights admixture patterns in the some African breeds and breeds from the Americas, as well as distinguishing the isolated Hebridean “primitive” breeds of Soay and Boreray from other breeds.

When the ancient samples are projected onto the modern day variation (Fig 6.3), we immediately see a split between the ancient southwest Asian samples and the European archaeological samples on PC1. The Iron Age/Medieval Southwest Asian samples situate in between the modern diversity of Asia and Southwest Asia. Within the European ancients, there is an observable split between the Neolithic samples, with the earlier Neolithic Serbian and Bulgarian Neolithic samples (~6000-5000 BC) grouping near modern day swiss sheep breeds and are the closest to the modern Southwest Asia. One German Neolithic (Her8 ~5000 BC) falls between the Balkan Neolithics and the British Neolithic (~2500 BC); another (Her3) groups with British Medieval samples.

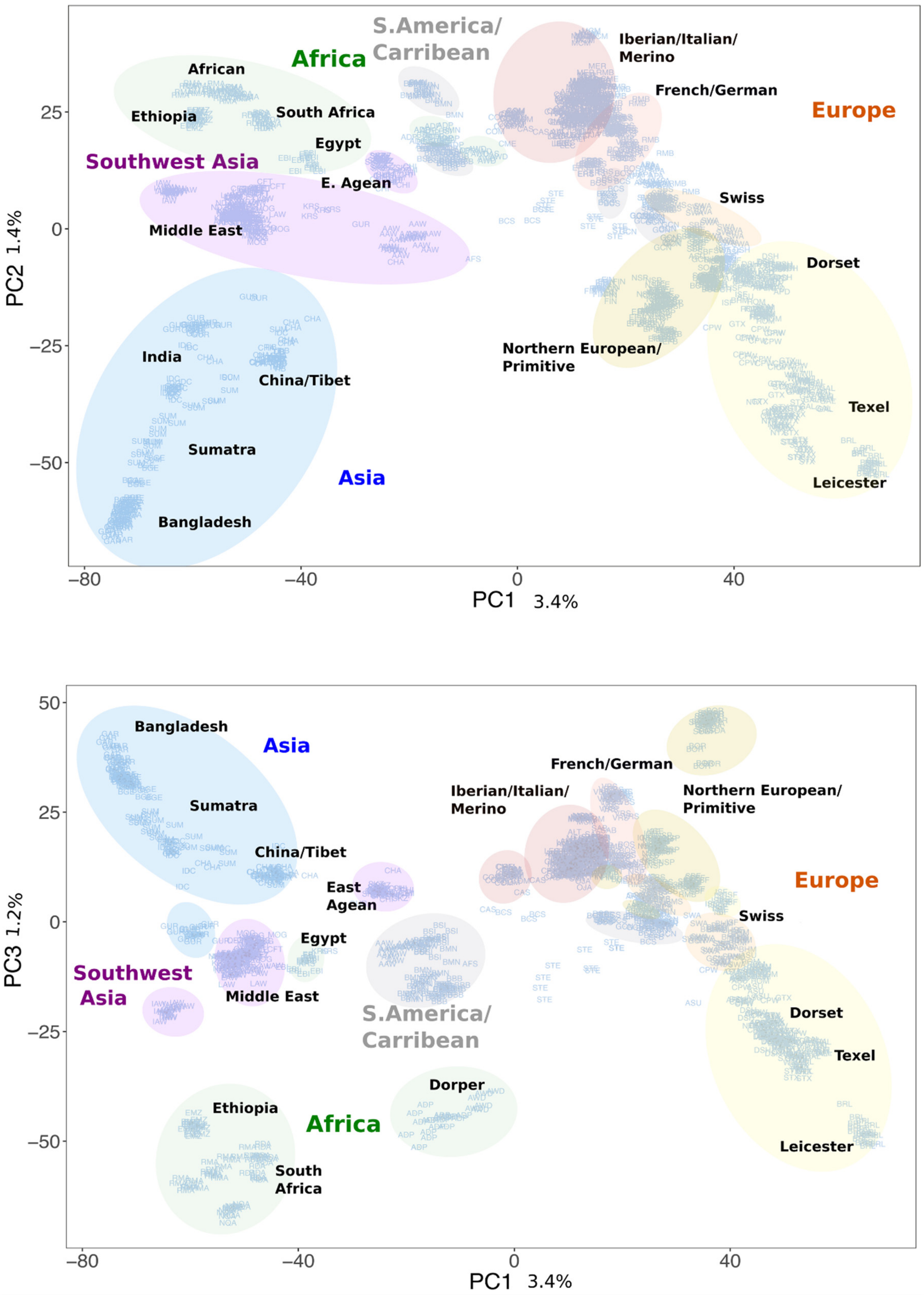


Fig 6.2. *Principal Components Analysis of modern sheep breeds using the 50K Hapmap dataset. PC1 and PC2 (top panel) reveals the split of Europe from Asia, Southwest Asia and Africa. Additionally the structure within Europe and Asia is apparent. PC1 and PC3 (bottom panel) highlights the admixture of breeds from the Americas and splits out the primitive breeds of Soay and Boreray. Colour coding identifies regional geographic clustering, while the labelling is similar to Kijas et al. 2012, showing regional clusters. Appendix Fig 4. demonstrates breed level clusters.*

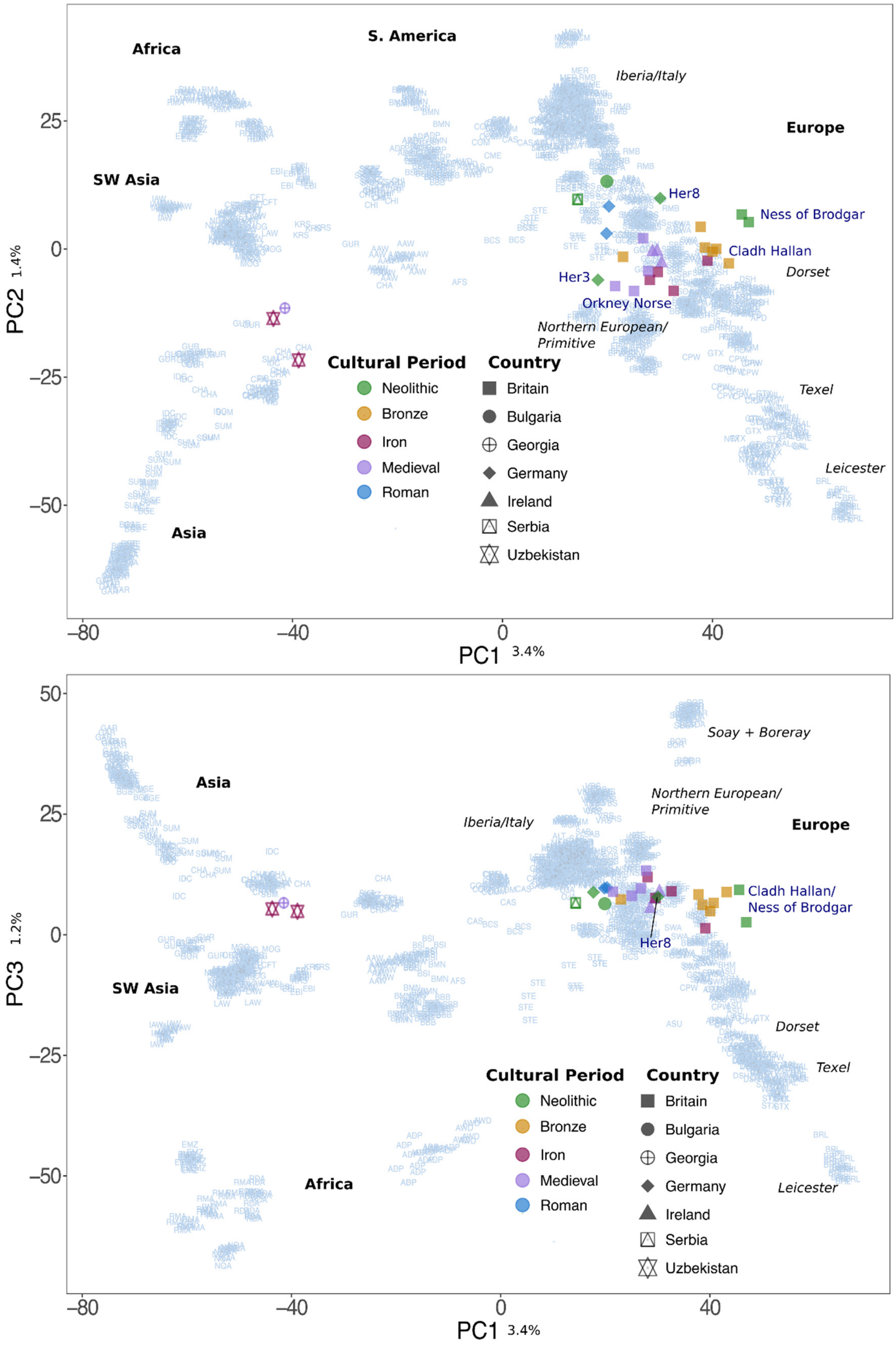


Fig 6.3. *Principal Components analysis with projection of ancient samples. PC1 vs PC2 (top panel) reveals a split between the ancient European samples and Asian samples, as well as structure within ancient Europe. PC1 vs PC3 (bottom panel) reveals a split between the Scottish Neolithic and Bronze group and the rest of Europe. Other British/Irish samples are sandwiched between mainland European samples and this Scottish group. Modern labels correspond with Fig 6.2, while some ancient samples are labelled in blue.*

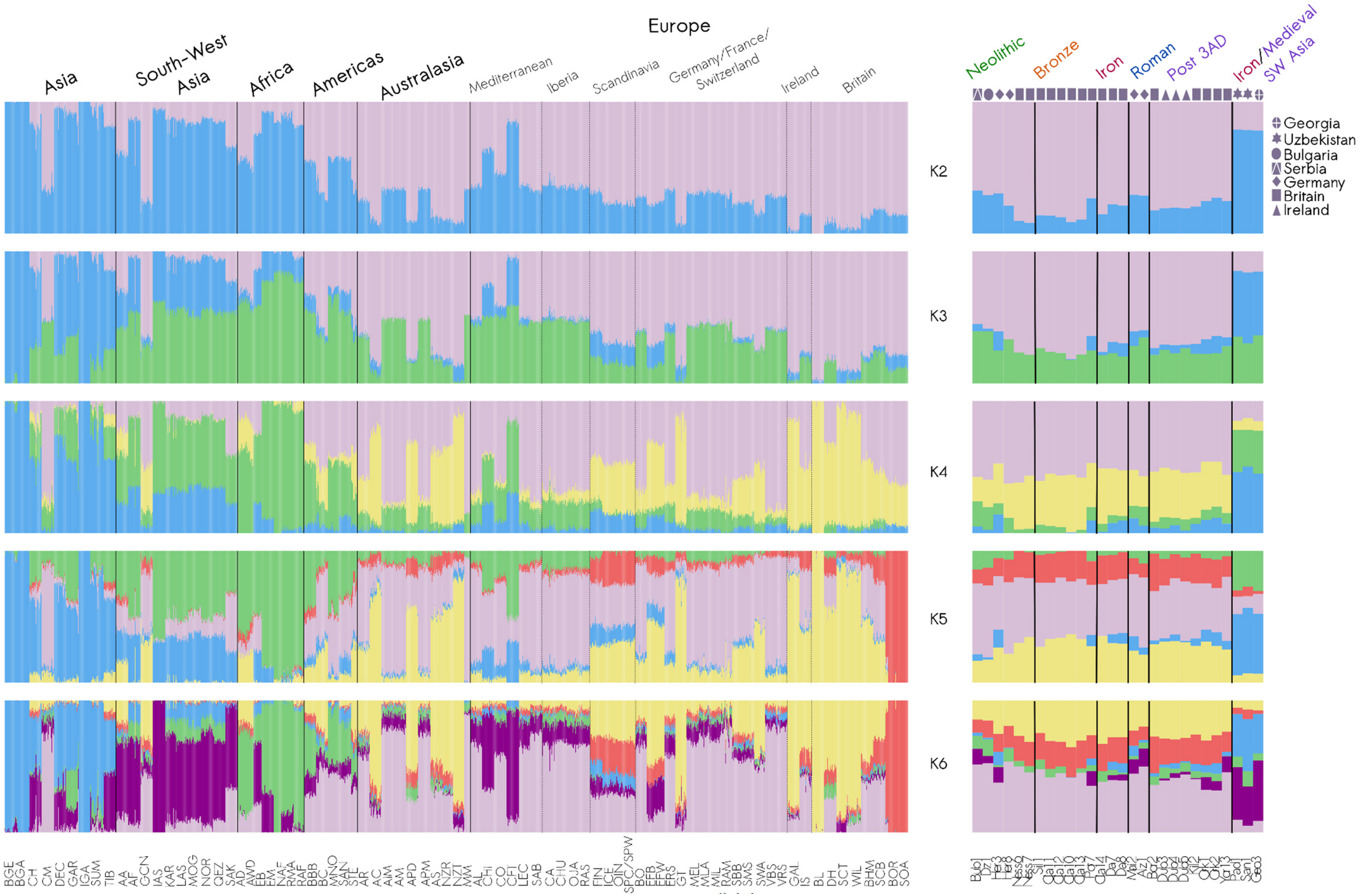
Within the British and Irish sheep there is a clear distinction between the group including the Scottish Island samples of Neolithic The Ness of Brodgar and the Bronze/Iron Cladh Hallan and the group of Late Bronze Age, Iron Age (511-43BC) and Medieval samples from England and Ireland. The two Roman German (~1-200 AD) samples then situate between the Balkans Neolithic samples and these later British/Irish samples. The majority of the later period British and Irish samples cluster near the upland breed Scottish Blackface. Interestingly the two Orkney Norse samples (Ork1&Ork2)(1000-1200 AD) fall towards the Scandinavian modern breeds, *e.g.* Finnsheep, Icelandic and Old Norwegian Spaelsau, whilst the Boranis Iron Age sample from the outer Herbidian Islands clusters towards the “primitive” isolated Boreary and Soay sheep. However, PC3 does not separate the ancient samples and therefore no ancient samples cluster with the primitive british Boreary and Soay breeds.

6.4.3. Ancestral Estimation- ADMIXTURE

Modern geographic structure emerges clearly in the ADMIXTURE analysis (Fig 6.4a&b) with K=2 separating European from Asian breeds. With K=3 the blue Asian component splits in two yielding three components. The European component strongest in Europe, the Asian component remaining in East Asia, and a third component strongest in Southwest Asia and Africa with some presence in Europe. As the K values increase, modern breeds start to form their own clusters: at K=4 Leicester falls out; at K=5 = the primitive breeds of Soay and Boreary gain their own component. While African breeds (K=6), Dorset horn (K=7), Friesian (K=8), Scandinavian breeds (K=9) and the British Wiltshire Breed (K=10) emerge as components.

Modern subgroups are uninformative for assessing ancient variation so only up to K=10 is considered. At K=2 all ancient samples show two components with the Europeans sharing a majority European component and the Southwest Asian ancient samples sharing a majority Asian coefficient. However at K=3 we start to see difference within the European ancient samples. The British Neolithic, German Neolithic sample Hxh8 and the Bronze Age/Iron Age Cladh Hallan samples all lack this Asian component, whilst later British, Irish and Roman German samples all possess this. At K=5 the British and Irish animals have a greater Boreray/Soay component than the early Neolithic Balkans samples and the German Roman samples. With K=6 the Middle Eastern component is present in all samples bar the Neolithic German sample Hxh8, Neolithic British and the Bronze/Iron Cladh Hallan samples. The Northern European animals share larger components of Friesian and Dorset Horn (K=7&8) than the Southern European Neolithic animals and the

Asian animals, whilst at K=9 the Scandinavian component is seen most strongly in the Norse samples from the Orkney Isles.


Europe
Asia
South-West Asia
Africa
Americas
Australasia
Mediterranean
Iberia
Scandinavia
Germany/France/Switzerland
Ireland
Britain
Neolithic
Bronze
Iron
Roman
Post 3AD
Iron/Medieval SW Asia
Georgia
Uzbekistan
Bulgaria
Serbia
Germany
Britain
Ireland
K2
K3
K4
K5
K6
BGE
BGA
CH
CM
DEC
GAR
IGA
SUM
TIB
AA
AF
GCN
IAS
KAR
LAS
MOG
NOR
QEZ
SAK
AD
AWD
EB
EM
NAF
RMA
RAF
BBB
BC
MNO
SAN
STE
AR
AC
AIM
AM
APD
APM
AS
NZR
NZT
MM
AL
CHI
CO
CFT
LEC
SAB
CA
CHU
OJA
RAS
FIN
ICE
OIN
SPC/SPW
BO
EFB
EFW
ERS
GT
MEL
MLA
MIL
RAM
SBB
SMS
SWA
VBS
VRS
GAL
IS
BL
DH
SCT
WIL
BHM
SCB
BOR
SOA
Bub1
Dz1
Her3
Her8
Ness9
Ness7
Sil1
Cal11
Cal12
Cal10
Cal13
Pol7
Cal14
Da7
Da8
Mel2
Az1
Bar2
Dub3
Dub4
Dub5
Kil2
Ok1
Ok2
Yar13
Pad1
San1
Geo3

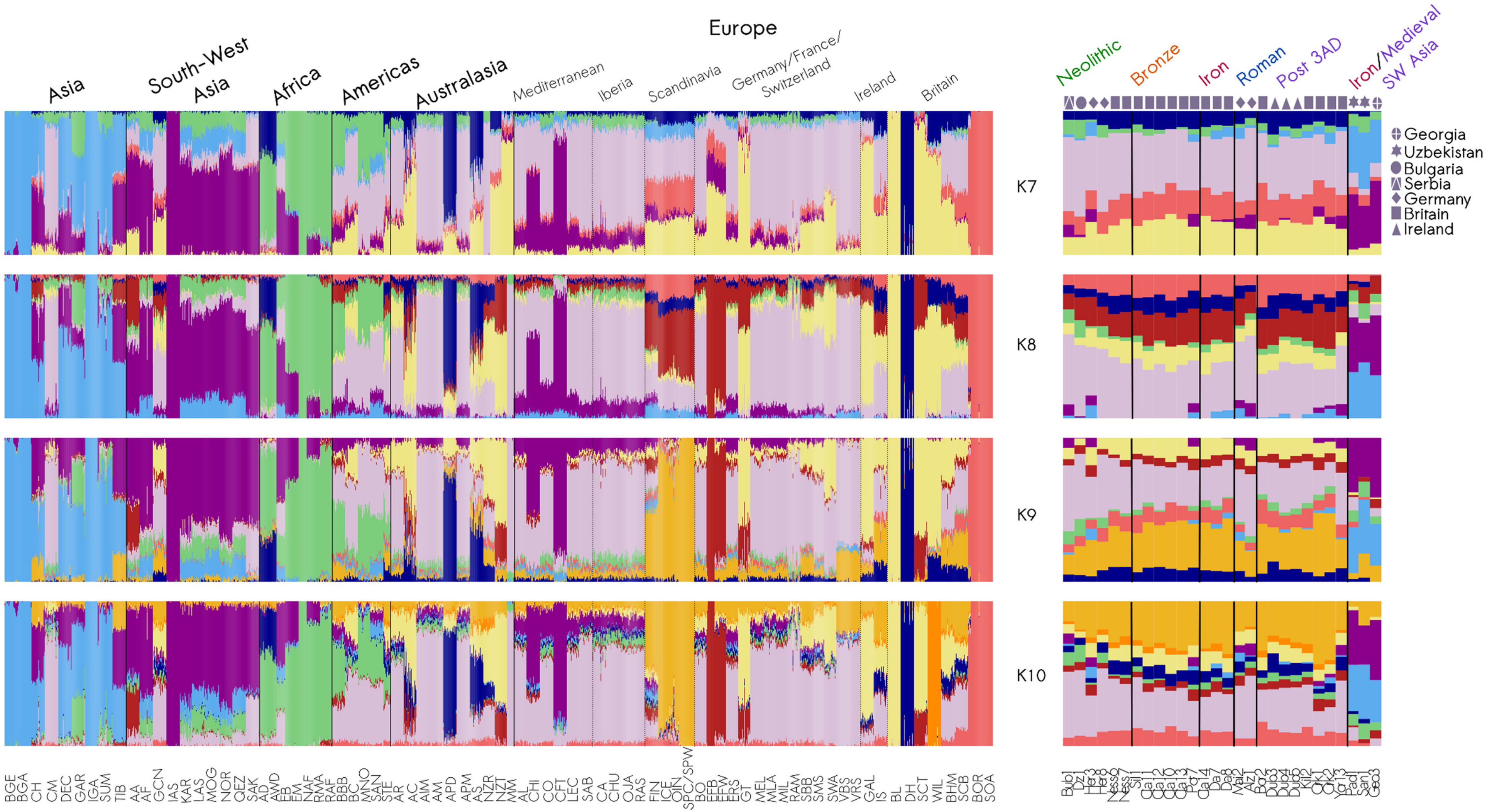


Fig 6.4a&b. *Unsupervised ADMIXTURE analysis of modern and ancient sheep populations using variance from the 50K sheep HapMap. Ancestry coefficient K2-10 are shown, with modern samples on the left and ancient samples on the right. The unmistakable split between Asian and European breeds is present at K=3, while. Ancient samples order via age and a shape denotes geographical origin*

6.4.4. Asian/Southwest Asian Admixture into European Sheep

6.4.4.1. D-statistics

To further explore the ADMIXTURE result of differing Asian and Southwest Asian components in European samples, *D*-statistics were employed using a variety of breeds as a proxy for these two components. In general, the *D*-statistic results confirm the ADMIXTURE results. Differing numbers of derived Asian/Southwest Asian alleles in the European modern and ancient samples, support the hypothesis of an introgression event from an Asian/SouthWest Asian type of sheep.

6.4.4.1.1. Asian Admixture

6.4.4.1.1.1. D(Neolithic Britain,Ref2),Asia,Goat)

When testing for admixture of the Asian component in samples in comparison to Neolithic British sample *D*(Ness6,Ref2),Bangladeshi Garole, Goat) (Fig 6.5), a clear pattern of a significant increase in the derived Asian alleles in modern European sheep emerges. This includes many of the British and Irish breeds. Within the ancient samples there is no significant difference (Z- score ≥ 3) in shared derived Asian alleles between the different Neolithic animals, except Her3 which is Neolithic by context. However we do see a pattern of increasing *D* and *Z*-scores indicating an increase of Asian component through the Bronze Age, Iron Age, Roman and Medieval time periods. However, only the Medieval and Roman periods show highly significant Z-scores (Z- score ≥ 3, equivalent to p-values > 0.001).

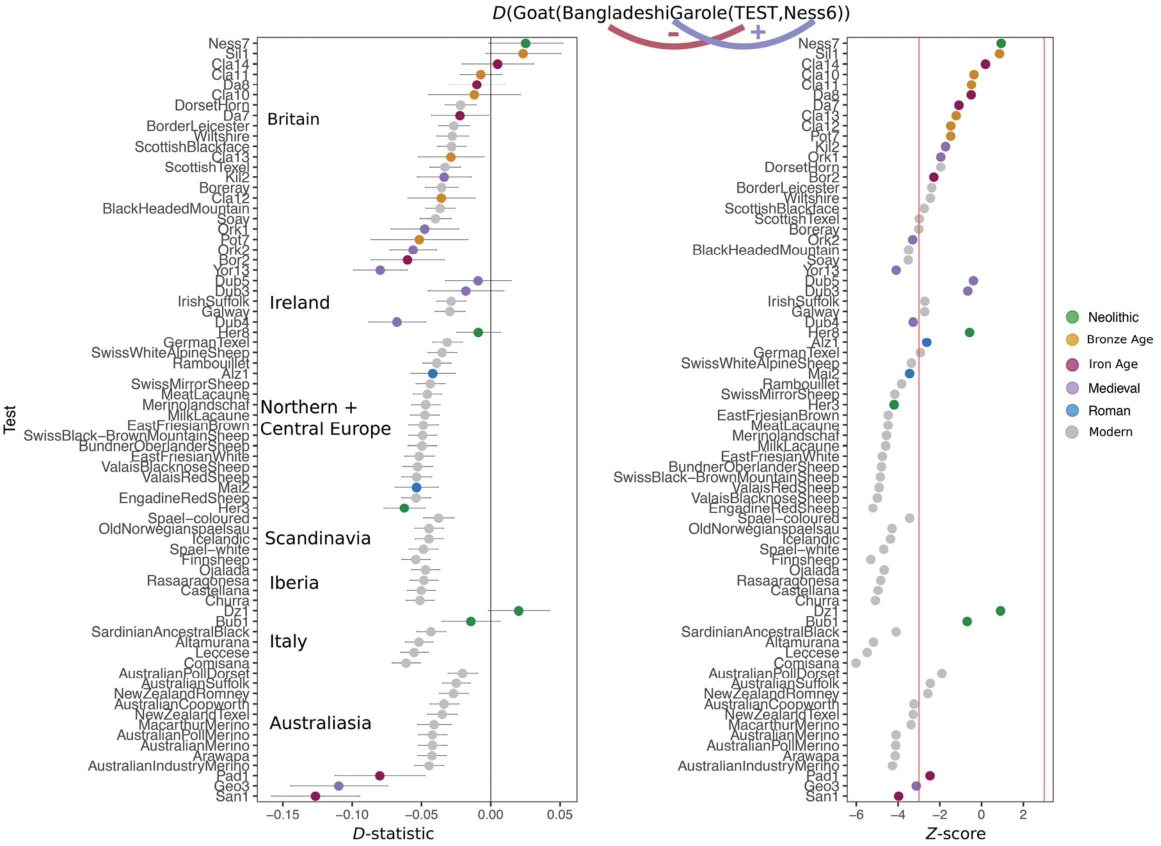


Fig 6.5. *D-statistics to measure introgression from the Bangladeshi Garole breed, into European derived breeds and ancient samples in comparison to a Neolithic British sample, Ness6. Colour denotes cultural period. The ancient samples from Asia are included in these statistics as a reference point. Red line indicate Z-score of 3. Note: the Australian breeds are derived from European breeds, hence they have been included in the test.*

Due to the low SNP count in the individual analysis samples were grouped, where possible, into cultural groups of ancients and the Asian admixture was re-tested. The *D* tests *D*(Neolithic GBR,Ref2)Bangladeshi Garole, Goat) and *D*(Neolithic GBR, Ref2), Deccani, Goat) convey the general pattern of an increase in Asian ancestry in the modern breeds and more recent ancient populations (Fig 6.6 & 6.7) (Appendix Table 17). *Z*-scores for many of the modern breeds, the Norse Orcadian samples, the Roman Germanic samples are of high significance (Z- score ≥ 3). While the trend in Bronze and Iron Age samples for increase in Asian ancestry still exists the *D* score is not highly significant, *Z*-score < 3. There are fluctuations in the *D* and *Z*-scores when the Asian component is represented by the different breeds, however the pattern of admixture in the more recent ancient samples holds.

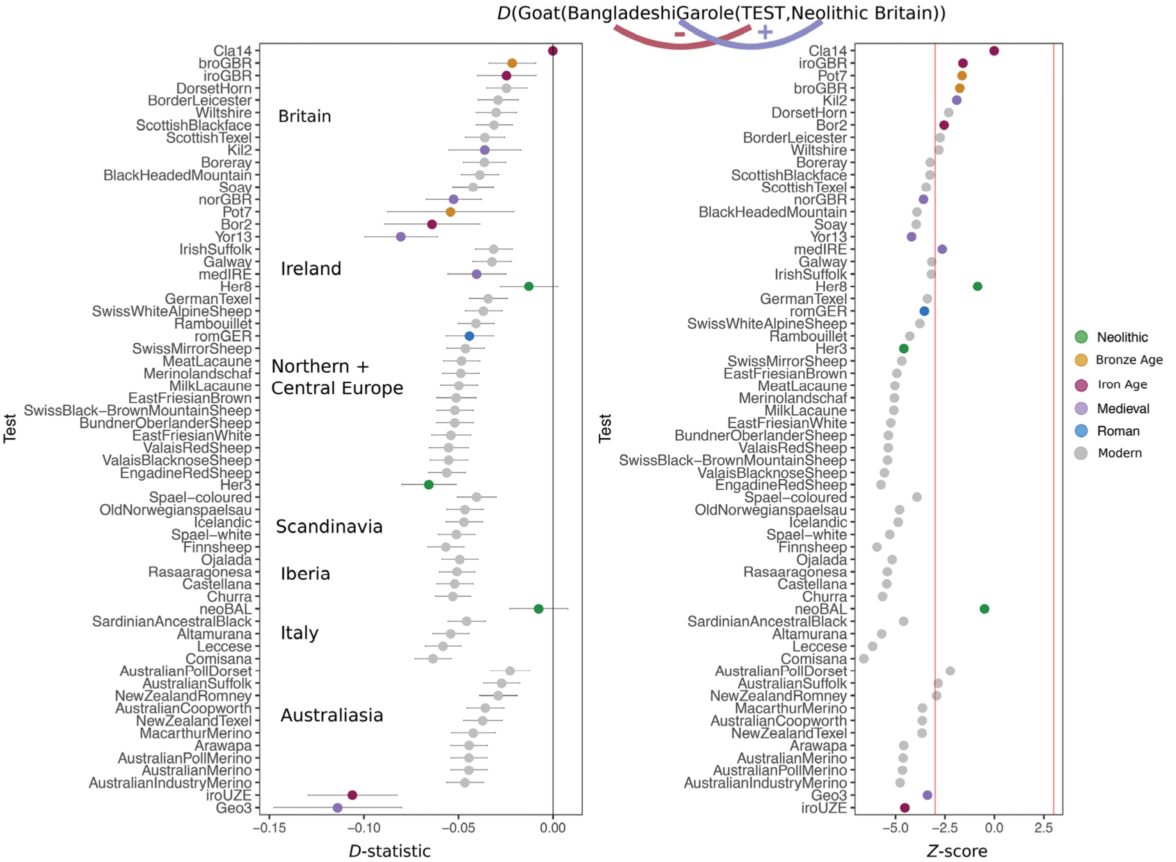


Fig 6.6. *D-statistics to measure introgression from the Bangladeshi Garole breed, into European derived breeds and ancient samples in comparison to a Neolithic British population. Where possible ancient samples were combined into populations based upon temporal and geographical information. Abbreviations are as follows: neoBAL= Neolithic Balkans, broGBR= Bronze Age Cladh Hallan (GBR), iroGBR= Iron Age Danebury (GBR), medIRE= Medieval Dublin (IRE), norGBR= Medieval (Norse) Orkney (GBR), iroUZE= Iron Age Uzbekistan and romGER= Roman Germany. Colour denotes cultural period geographic labels denote the country/region of the modern breeds. The ancient samples from Asia are included in these statistics as a reference point. Note: the Australian breeds are derived from European breeds, hence they have been included in the test.*

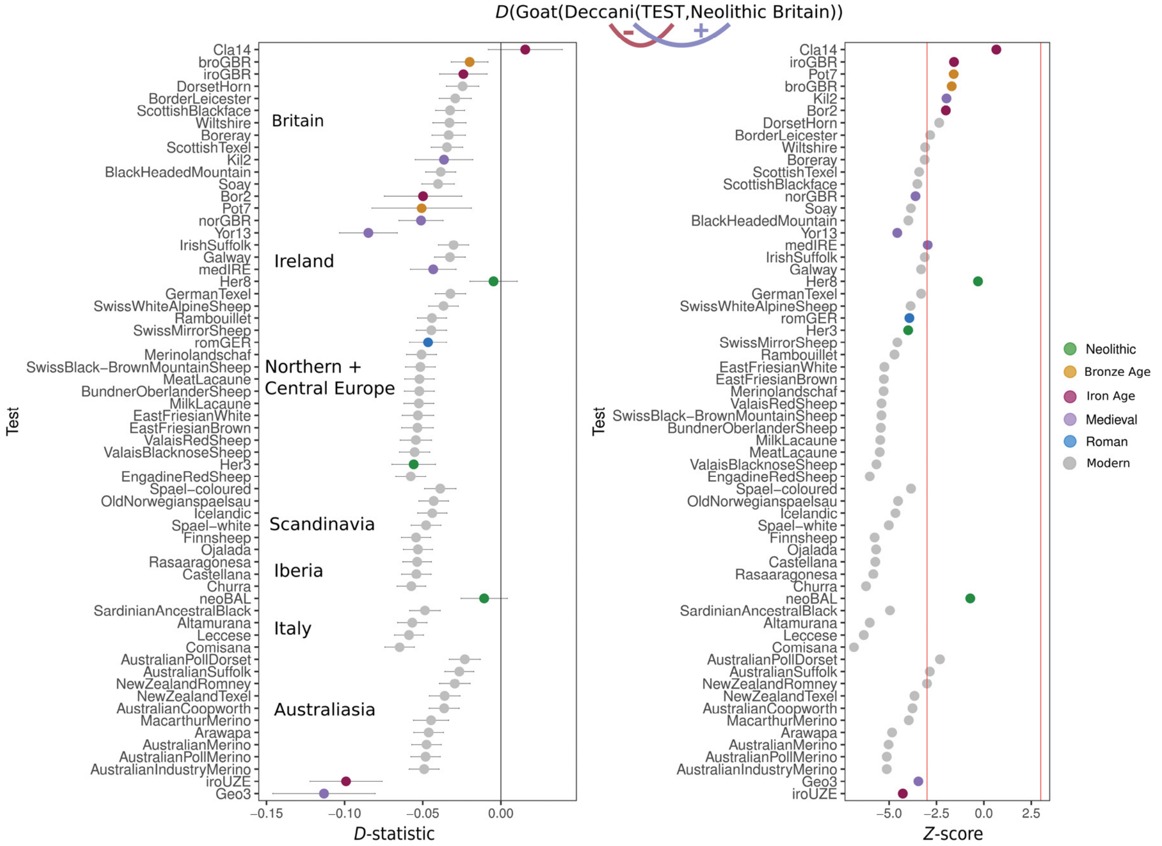


Fig 6.7. *D-statistics to measure introgression from the Deccani breed, into European derived breeds and ancient samples in comparison to a Neolithic British population. Where possible ancient samples were combined into populations based upon temporal and geographical information. Abbreviations are as follows: neoBAL= Neolithic Balkans, broGBR= Bronze Age Cladh Hallan (GBR), iroGBR= Iron Age Danebury (GBR), medIRE= Medieval Dublin (IRE), norGBR= Medieval (Norse) Orkney (GBR), iroUZE= Iron Age Uzbekistan and romGER= Roman Germany. Colour denotes cultural period and geographic labels denote the country/region of the modern breeds. The ancient samples from Asia are included in these statistics as a reference point. Note: the Australian breeds are derived from European breeds, hence they have been included in the test.*

6.4.4.1.1.2. D(Roman Germany, Ref2), Asia, Goat)

In another comparison of the number of derived Asian alleles in ancient populations, this time using the Roman German samples *D*((Roman DEU, Ref2), Bangladeshi Garole, Goat) & *D*((Roman Deu, Ref2), Deccani, Goat)), there is no statistical difference in the number of derived Asian alleles between the modern mainland European samples, and the Roman samples (Fig 6.8 & 6.9). However, there is a trend of fewer South Asian derived alleles in modern British/Irish breeds than the Roman German samples, though non are highly significant (*Z*-score > 3). Moreover, there is a highly significant (*Z*-score > 3) decrease in the number of derived Asian alleles in the British Neolithic in comparison to the Roman samples, indicating introgression into the Roman samples. In the other European Neolithic samples and the British Bronze Age samples this decrease is also present (*Z*-scores < 2). In comparison there is no statistical difference between Roman samples and the British Iron Age and Medieval Irish.

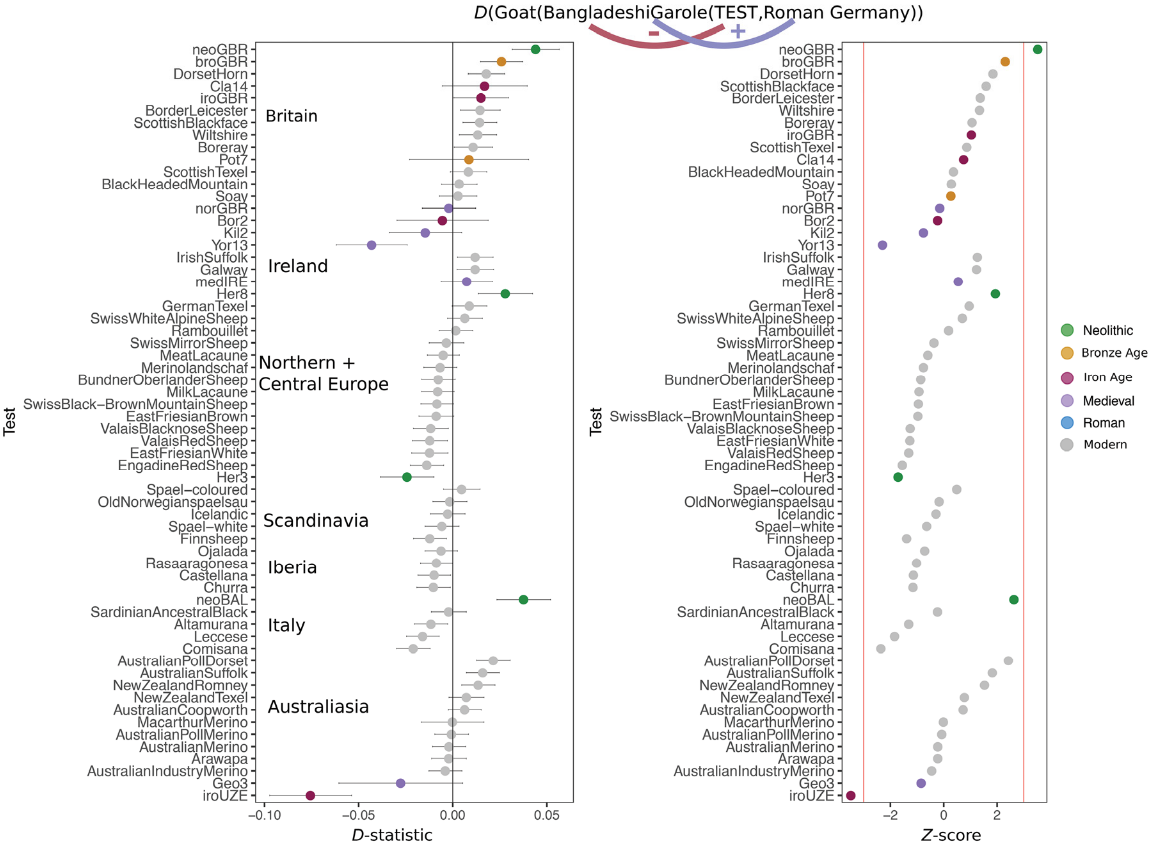


Fig 6.8. *D-statistics to measure introgression from the Bangladeshi Garole breed, into European derived breeds and ancient samples in comparison to a Roman German population. Where possible ancient samples were combined into populations based upon temporal and geographical information. Abbreviations are as follows: neoBAL= Neolithic Balkans, neoGBR= Neolithic Ness of Brodgar (GBR), broGBR= Bronze Age Cladh Hallan (GBR), iroGBR= Iron Age Danebury (GBR), medIRE= Medieval Dublin (IRE), norGBR= Medieval (Norse) Orkney (GBR), iroUZE= Iron Age Uzbekistan and romGER= Roman Germany. Colour denotes cultural period and geographic labels denote the country/region of the modern breeds. The ancient samples from Asia are included in these statistics as a reference point. Note: the Australian breeds are derived from European breeds, hence they have been included in the test.*

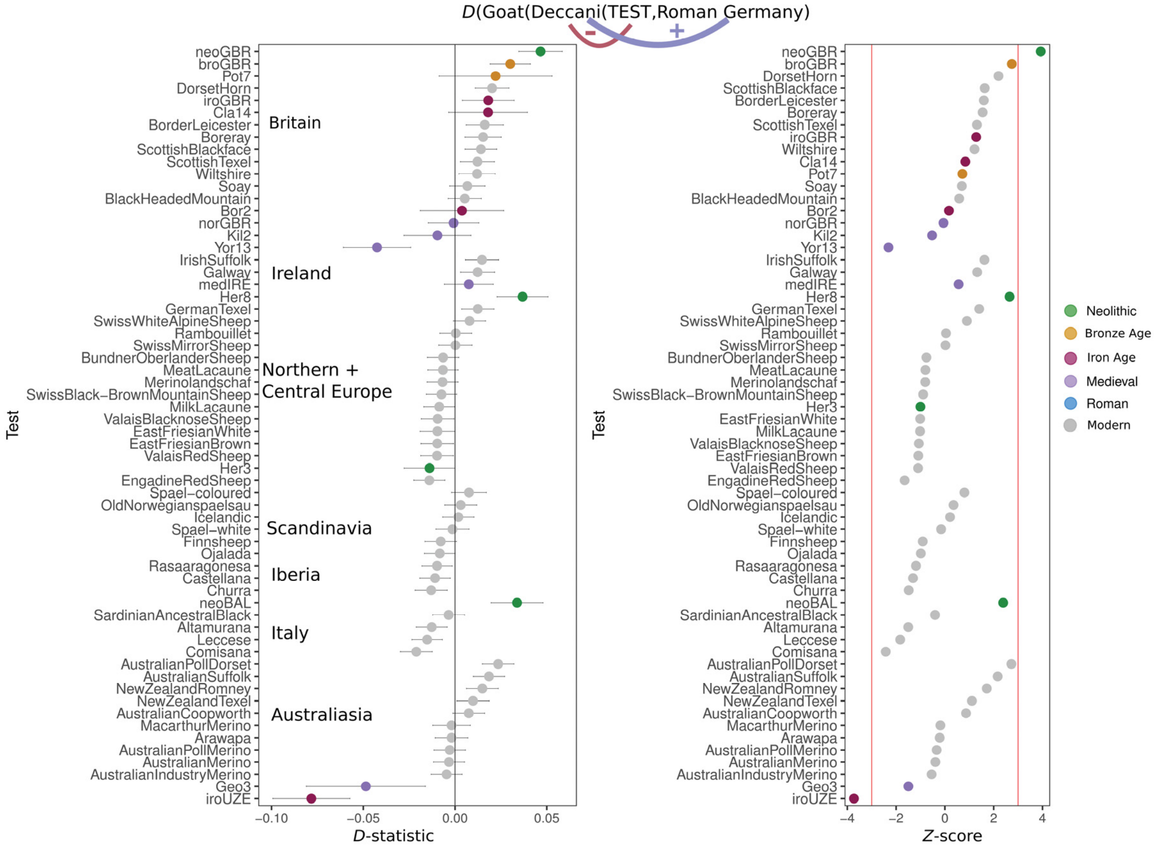


Fig 6.9. *D-statistics to measure introgression from the Bangladeshi Garole breed, into European derived breeds and ancient samples in comparison to a Roman German population. Where possible ancient samples were combined into populations based upon temporal and geographical information. Abbreviations are as follows: neoBAL= Neolithic Balkans, neoGBR= Neolithic Ness of Brodgar (GBR), broGBR= Bronze Age Cladh Hallan (GBR), iroGBR= Iron Age Danebury (GBR), medIRE= Medieval Dublin (IRE), norGBR= Medieval (Norse) Orkney (GBR), iroUZE= Iron Age Uzbekistan and romGER= Roman Germany. Colour denotes cultural period and geographic labels denote the country/region of the modern breeds. The ancient samples from Asia are included in these statistics as a reference point. Note: the Australian breeds are derived from European breeds, hence they have been included in the test.*

6.4.4.1.1.3. D(Scottish Blackface, Ref2), Asia, Goat)

To test whether a Northern European modern breed, as defined by (Kijas et al. 2012), had a significant difference in Asian admixture with other European breeds and ancients the test *D*(Scottish Blackface, Ref3),Bangladeshi Garole, Goat) was performed (Fig 6.10). Results indicate highly significant differences when Scottish Blackface tested against continental breeds, such as Swiss Black-Brown Mountain sheep, Scandinavian breeds such as Icelandic and Finnsheep and Merino breeds. Although there is a pattern of increasing shared derived Asian alleles in the British breeds, non has a significant difference in shared derived Asian alleles with the Scottish Blackface.

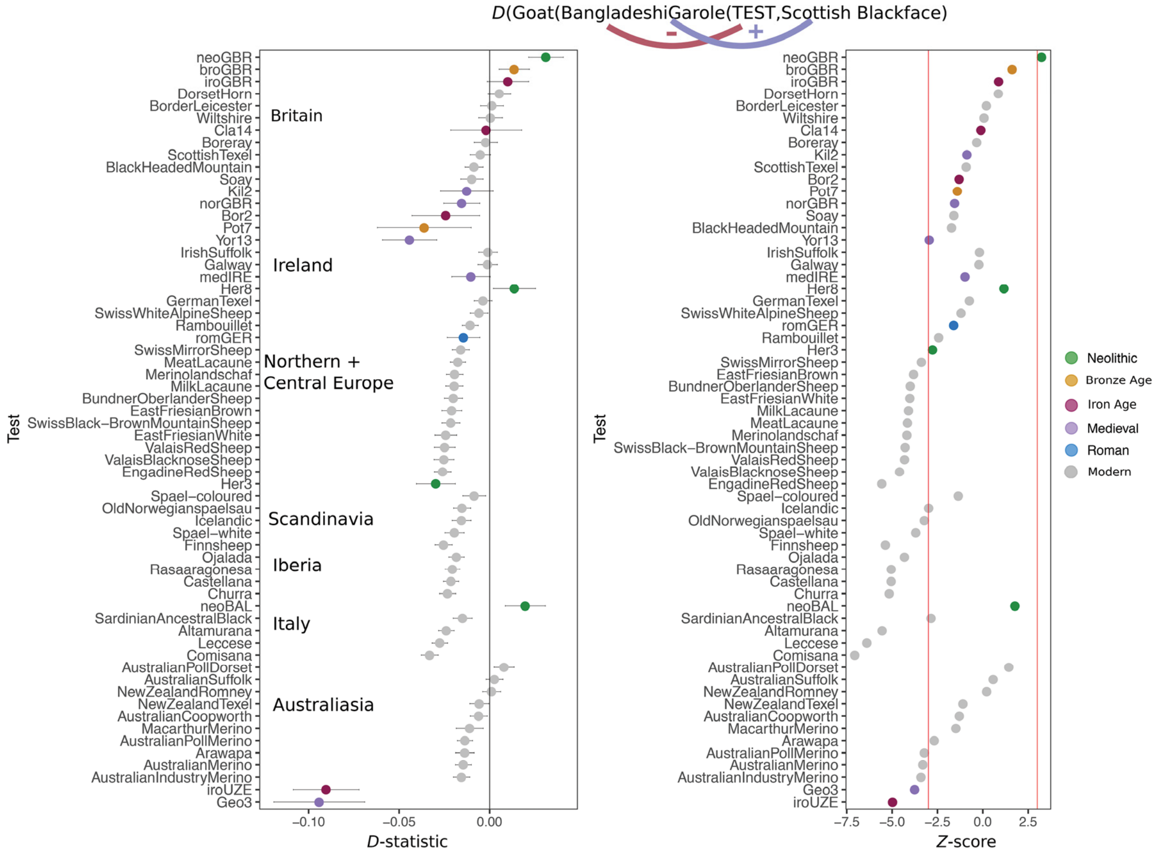


Fig 6.10. D-statistics to measure introgression from the Bangladeshi Garole breed, into European derived breeds and ancient samples in comparison to the Northern European breed Scottish Blackface. Where possible ancient samples were combined into populations based upon temporal and geographical information. Abbreviations are as follows: neoBAL= Neolithic Balkans, neoGBR= Neolithic Ness of Brodgar (GBR), broGBR= Bronze Age Cladh Hallan (GBR), iroGBR= Iron Age Danebury (GBR), medIRE= Medieval Dublin (IRE), norGBR= Medieval (Norse) Orkney (GBR), iroUZE= Iron Age Uzbekistan and romGER= Roman Germany. Colour denotes cultural period and geographic labels denote the country/region of the modern breeds. The ancient samples from Asia are included in these statistics as a reference point. Note: the Australian breeds are derived from European breeds, hence they have been included in the test.

6.4.4.1.2. Southwest Asian Admixture

6.4.4.1.2.1. (Ness6,Ref2),Local Awassi, Goat)

With the test (Ness6,Ref2),Local Awassi, Goat) we again see the modern European samples containing a higher proportion of derived Local Awassi alleles to the British Neolithic sample Ness6 (Fig 6.11). Significance of *Z*-score ≥ 3 are calculated for Continental European breeds, while *Z*-score > 2 are recorded for British and Irish breeds. Within the ancient samples, the Neolithic Balkan samples (Dz1 & Bub1) and the German Neolithic sample Her8 have once again similar levels of shared derived alleles as the British Neolithic sample Ness6, indicating no admixture from the Southwest Asia. Post-Neolithic there is a general pattern for increasing *D* and *Z*-scores indicating increasing shared derived alleles between test and Local Awassi. However the British Bronze and Iron age samples show no significant difference. Whereas *Z*-scores > 2 for the Roman

(Mai2 & Alz1) and the higher coverage Medieval samples (Dub4 & Ork2) indicate admixture between the samples and Local Awassi.

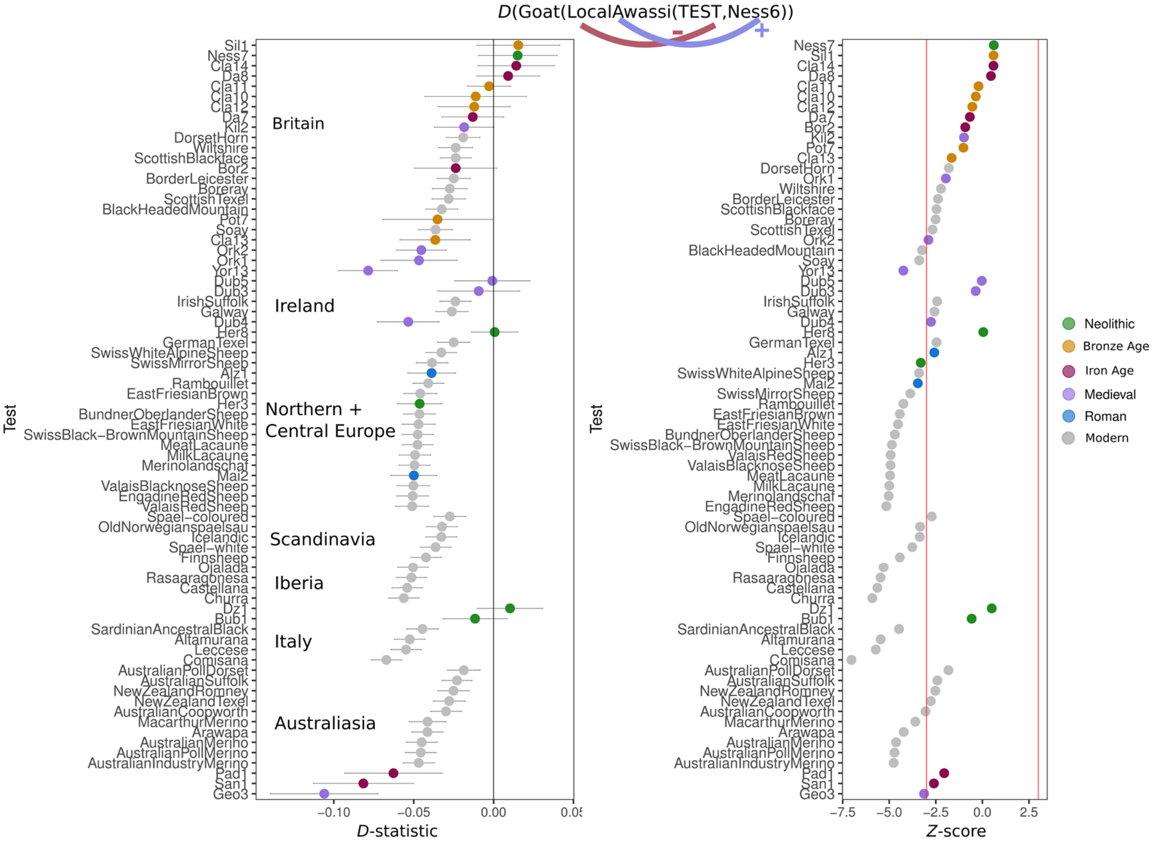


Fig 6.11. *D-statistics to measure introgression from the Local Awassi breed, into European derived breeds and ancient samples in comparison to a Neolithic British sample, Ness6. Colour denotes cultural period. The ancient samples from Asia are included in these statistics as a reference point. Note: the Australian breeds are derived from European breeds, hence they have been included in the test.*

6.4.4.1.2.2. D(Population, Ref2),Southwest Asia, Goat)

To test Southwest Asian introgression into ancient populations the Southwest Asian breed of Local Awassi was positioned as the introgressing population. Breeds and ancient populations (where possible) were tested in a pairwise fashion against Neolithic Britain and Roman German, respectively (Appendix Figure 5&6). The similar pattern of an increase of shared derived alleles, indicating introgression, is repeated; the modern breeds and the more recent ancient samples have past admixture with a Southwest Asian type animal, however there are fluctuations in *D* and *Z*-scores depending on which breed is used.

6.4.4.2. f_3-statistic – Admixture test

The f_3-statistic was performed to test for admixture of an Asian or Southwest Asian-like population with European ancient populations. The results (Table 6.6 and Table 6.7), albeit with low SNP numbers and *Z*-scores, indicate an admixture event in the ancestry of the more recent populations. For Iron Age Britain, Medieval Ireland and Roman Germany there is a signal of an admixture event occurring between Asian and/or Southwest Asian breeds with British Neolithic population and to some extent with the British Bronze Age population. Intriguingly this event can no longer be identified when using an ancient source post-Bronze Age. There is no trace of an admixture event prior to the British Bronze Age population tested here. In addition there is also no signal of an admixture event in the ancestry of the Orcadian Norse population.

Table 6.6. *f_3-statistic results testing admixture between Asian breeds and ancient populations. A negative f_3 value indicates that the Target has ancestry from Source 1 and Source 2. Z-scores are coloured to indicate significance level: green= $P \leq 0.001$, orange = $0.001 < P \leq 0.01$, red= $0.01 < P \leq 0.05$.*

Source 1	Source2	Target	f3	Std. err	Z score	SNP no.
Deccani	Neolithic GBR	Roman DEU	-0.014717	0.004486	-3.28	13348
Bangladeshi Garole	Neolithic GBR	Roman DEU	-0.013839	0.004688	-2.952	13085
Bangladeshi Garole	Neolithic GBR	Iron GBR	-0.018728	0.007267	-2.577	5626
Deccani	Neolithic GBR	Iron GBR	-0.018169	0.007183	-2.53	5730
Bangladeshi Garole	Bronze GBR	Iron GBR	-0.015993	0.006461	-2.475	6392
Deccani	Bronze GBR	Iron GBR	-0.013398	0.006349	-2.11	6532
Deccani	Neolithic GBR	Medieval IRE	-0.012324	0.006125	-2.012	8022
Bangladeshi Garole	Neolithic GBR	Medieval IRE	-0.011935	0.006256	-1.908	7863
Deccani	Bronze GBR	Roman DEU	-0.007064	0.004141	-1.706	15446
Deccani	Bronze GBR	Medieval IRE	-0.00875	0.005611	-1.559	9130
Bangladeshi Garole	Bronze GBR	Medieval IRE	-0.008914	0.00581	-1.534	8935
Bangladeshi Garole	Bronze GBR	Roman DEU	-0.006049	0.004323	-1.399	15090
Deccani	Neolithic GBR	Norse GBR	0.001341	0.00702	0.191	5939
Deccani	Neolithic GBR	Bronze GBR	0.001591	0.004003	0.397	17879
Bangladeshi Garole	Neolithic GBR	Norse GBR	0.003515	0.007373	0.477	5814
Bangladeshi Garole	Iron GBR	Medieval IRE	0.003626	0.006959	0.521	6200
Deccani	Iron GBR	Medieval IRE	0.003651	0.006865	0.532	6314
Deccani	Bronze GBR	Norse GBR	0.005356	0.006683	0.801	6845
Bangladeshi Garole	Neolithic GBR	Bronze GBR	0.003584	0.004199	0.854	17637
Bangladeshi Garole	Bronze GBR	Norse GBR	0.007218	0.006911	1.044	6676
Deccani	Iron GBR	Roman DEU	0.005841	0.005161	1.132	10643
Bangladeshi Garole	Iron GBR	Roman DEU	0.007168	0.005385	1.331	10463
Deccani	Iron GBR	Norse GBR	0.0137	0.008011	1.71	4732
Bangladeshi Garole	Iron GBR	Norse GBR	0.014387	0.008339	1.725	4631

Table 6.7. *f_3- statistic testing admixture between Soutwest Asian breeds and ancient populations. A negative f_3 value indicates that the Target has ancestry from Source 1 and Source 2. Z-scores are coloured to indicate significance level: green= $P \leq 0.001$, orange = $0.001 < P \leq 0.01$, red= $0.01 < P \leq 0.05$.*

Source 1	Source2	Target	f3	Std. err	Z score	SNP no.
Afshari	Neolithic GBR	Roman DEU	-0.013814	0.004511	-3.063	13454
Local Awassi	Neolithic GBR	Roman DEU	-0.013462	0.004461	-3.018	13486
Afshari	Neolithic GBR	Iron GBR	-0.018641	0.006852	-2.72	5782
Afshari	Bronze GBR	Iron GBR	-0.015539	0.006337	-2.452	6579
Afshari	Neolithic GBR	Medieval IRE	-0.012163	0.005976	-2.035	8082
Local Awassi	Bronze GBR	Iron GBR	-0.012454	0.006126	-2.033	6604
Local Awassi	Neolithic GBR	Iron GBR	-0.013402	0.006892	-1.945	5796
Afshari	Bronze GBR	Medieval IRE	-0.009348	0.005506	-1.698	9180
Local Awassi	Bronze GBR	Roman DEU	-0.006989	0.004194	-1.666	15609
Afshari	Bronze GBR	Roman DEU	-0.006807	0.004254	-1.6	15558
Local Awassi	Neolithic GBR	Medieval IRE	-0.008677	0.005986	-1.45	8096
Local Awassi	Bronze GBR	Medieval IRE	-0.007795	0.005583	-1.396	9220
Local Awassi	Iron GBR	Medieval IRE	-0.000093	0.006713	-0.014	6372
Afshari	Iron GBR	Medieval IRE	0.001366	0.006719	0.203	6335
Afshari	Neolithic GBR	Norse GBR	0.003111	0.007053	0.441	5987
Afshari	Bronze GBR	Norse GBR	0.003288	0.006554	0.502	6883
Local Awassi	Neolithic GBR	Norse GBR	0.003658	0.006723	0.544	6010
Local Awassi	Iron GBR	Roman DEU	0.004464	0.005078	0.879	10747
Afshari	Neolithic GBR	Bronze GBR	0.003798	0.004078	0.931	17984
Local Awassi	Neolithic GBR	Bronze GBR	0.003784	0.003973	0.952	18027
Local Awassi	Bronze GBR	Norse GBR	0.006209	0.006466	0.96	6919
Afshari	Iron GBR	Roman DEU	0.007488	0.005271	1.421	10713
Local Awassi	Iron GBR	Norse GBR	0.014546	0.007993	1.82	4781
Afshari	Iron GBR	Norse GBR	0.015954	0.008197	1.946	4770

6.4.5. Northern European Sheep Populations

6.4.5.1. Outgroup f_3-statistics

To test shared drift between the ancient samples and modern European breeds the outgroup f_3-statistic was utilised for a number of tests (Appendix Table 18).

6.4.5.1.1. Neolithic and Bronze Age Britain

The results of the outgroup f_3(X, Neolithic Britain;Neolithic Balkans) show Neolithic Britain shares the most genetic drift with the British Bronze Age and British Iron Age, respectively (Fig 6.12 & Appendix Table 18). When the modern breeds are substituted for X, the British breeds, British derived breeds and the Scandinavian short-tailed breeds demonstrate the highest values of f_3, and therefore of shared drift with Neolithic Britain.

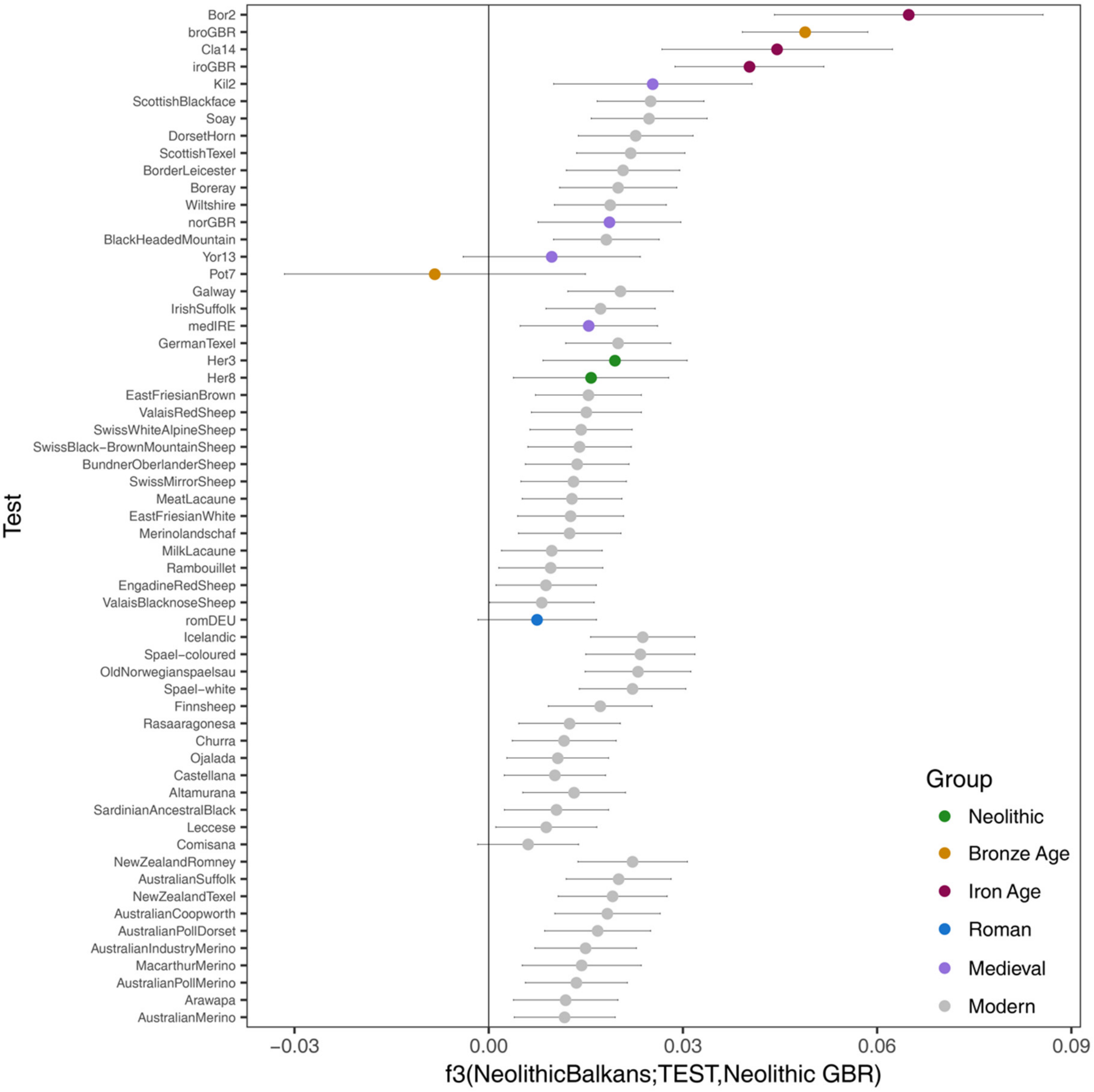


Fig 6.12 *Shared genetic drift, as calculate via the outgroup* f_3*-statistic test, of British Neolithic and variety of European and European derived breeds and ancient samples indicate high levels of drift with ancient samples from the Outer Hebridies and the Neolithic Orcadian samples.*

If Neolithic Britain is substituted for a population of Late British Bronze Age samples from Cladh Hallan (South Uist), the highest values of shared drift are displayed by the samples from the island of South Uist from the later Iron period (Cla14 & Bor2), followed by the Neolithic and Medieval (Norse) Britain populations (Fig 6.13 & Appendix Table 18). While the modern breeds that share the most drift are the Scandinavian short-tailed breeds, followed by British and British derived breeds from Australia. The Bronze Age sample Pot7 behaves strangely, however this sample is the lowest coverage sample in the analysis (Appendix Table 18).

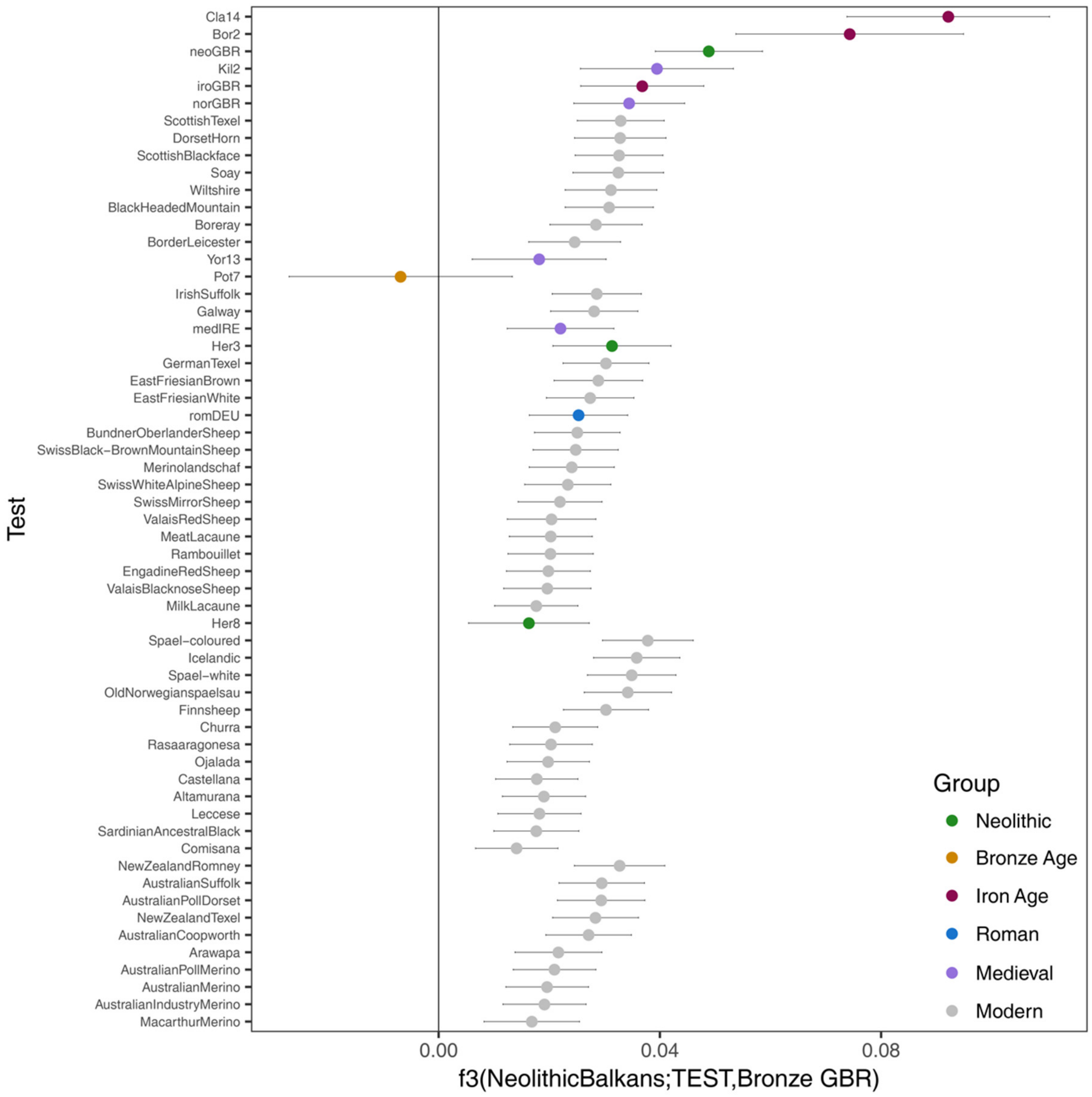


Fig 6.13. *Shared genetic drift, as calculate via the outgroup f_3-statistic test, of British Bronze Age and variety of European and European derived breeds and ancient samples indicate high levels of drift with Iron Age samples from the outer Hebridies and the Neolithic Orcadian samples.*

6.4.5.1.2. Iron Age Britain and Roman Germany

When comparing shared drift between the British Iron Age and other populations/breeds, f_3(X, Iron Age Britain;Neolithic Balkans), the highest f_3 values were for Late Scottish Iron Age sample Bor2 and the Medieval (Norse) sample Kil2 (Fig 6.14). Furthermore, the modern Scandinavian short-tailed breeds have higher values of shared drift with the British Iron Age than the earlier ancient samples, and even the modern British breeds.

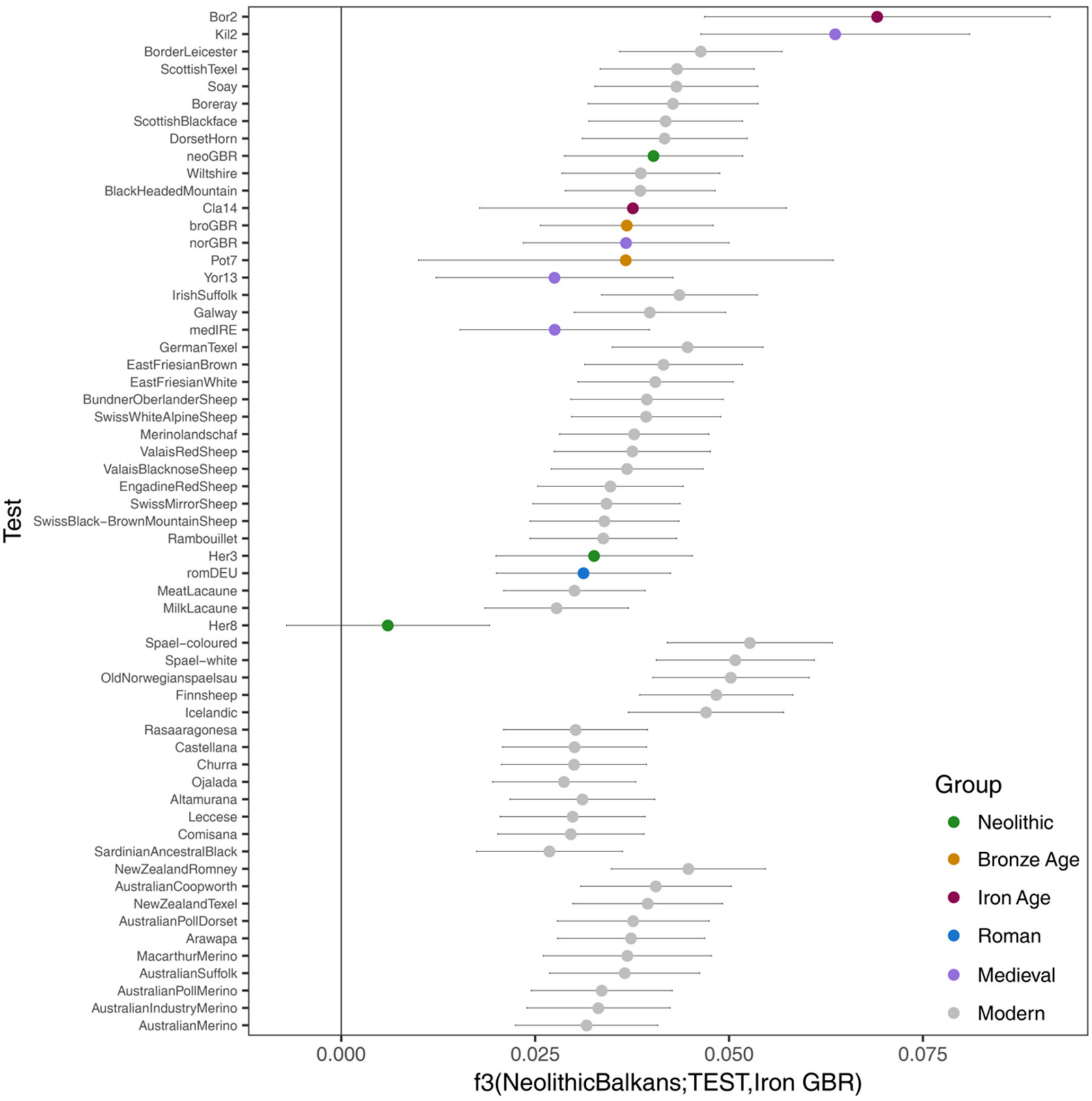


Fig6.14 *Shared genetic drift, as calculate via the outgroup f3-statistic test, of British Iron Age population and a variety of European and European derived breeds and ancient samples is higher with Scandinavian breeds than other modern breeds and the Iron Age and Medieval samples from the Outer Hebridies.*

Measuring the shared drift between the moderns and the German Roman samples does alter the pattern that we see from the ancient British and Irish samples (X, Roman German; Neolithic Balkans) (Fig 6.15 & Appendix Table 17). No specific population or breeds shares higher levels of drift with these samples. In contrast to other ancient populations it does not demonstrate higher levels of shared drift with the Scandinavian primitive breeds and British breeds.

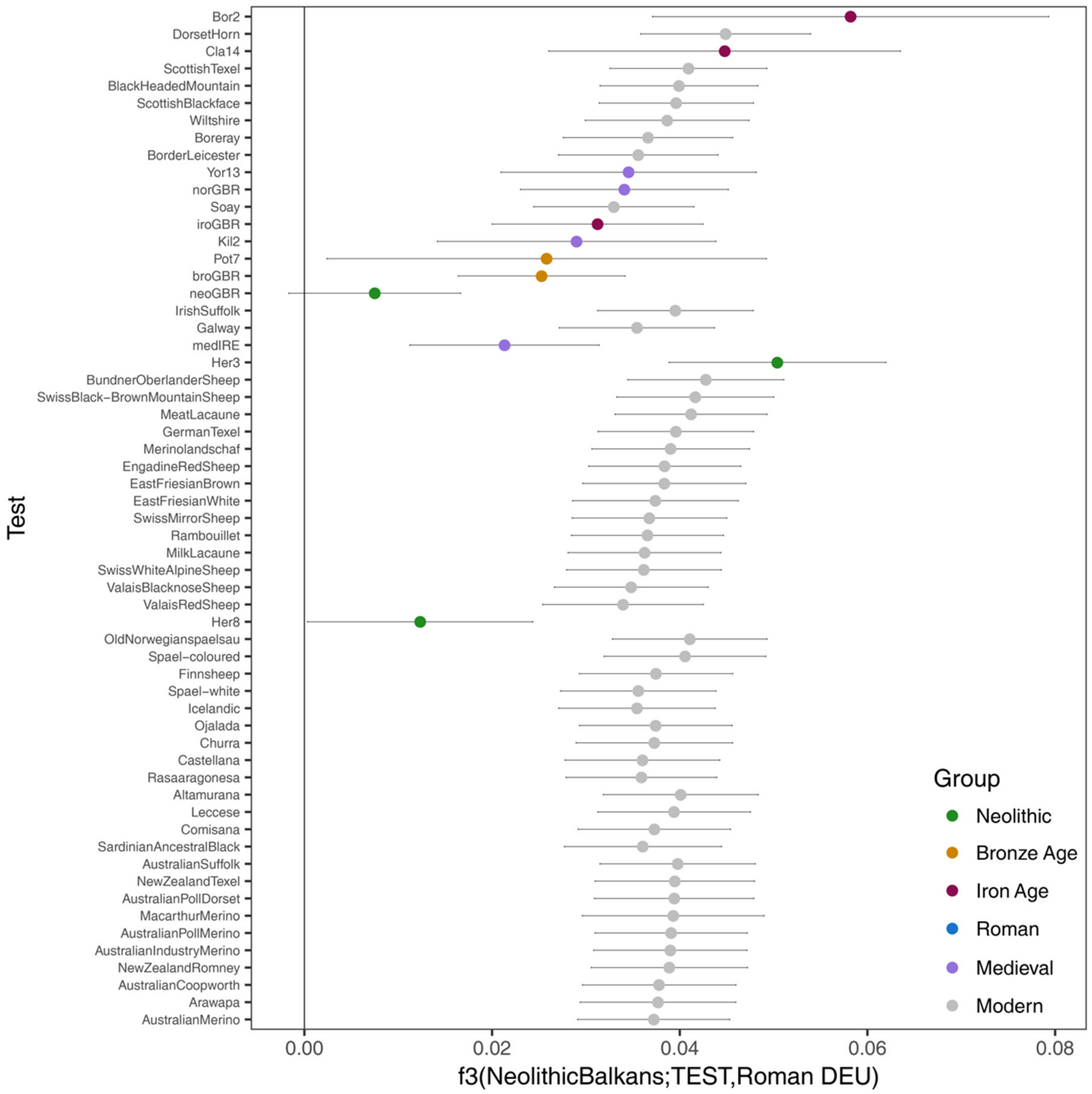


Fig 6.15. *Shared genetic drift, as calculated via the outgroup f3-statistic test, of Roman German and a variety of European and European derived breeds and ancient samples indicates similar levels of genetic drift with the modern breeds. In comparison to other ancient populations it does not demonstrate higher levels of shared drift with the Scandinavian primitive breeds.*

6.4.5.1.3. Medieval (Norse) Britain and Medieval (Norse) Ireland

Analysis centred on the Medieval Orcadian population from the Norse settlement of Snusgar (Ork1 & Ork2), (X, Medieval (Norse) Britain; Neolithic Balkans) (Fig6.16). Results indicate less shared drift with the older Neolithic and Bronze samples and higher values of $f_{3,}$ and therefore shared drift, with the ancient outer Hebridean samples from the Iron Age and Medieval (Norse) period (Cla14, Bor2 & Kil2). However, due to low SNP number some of the *Z*-scores are borderline to the high significance value of *Z*-score >3 (Appendix Table 18). When considering modern breeds, Scandinavian breeds gave the highest values of f_3, followed by the Friesian breeds and British/British derived breeds.

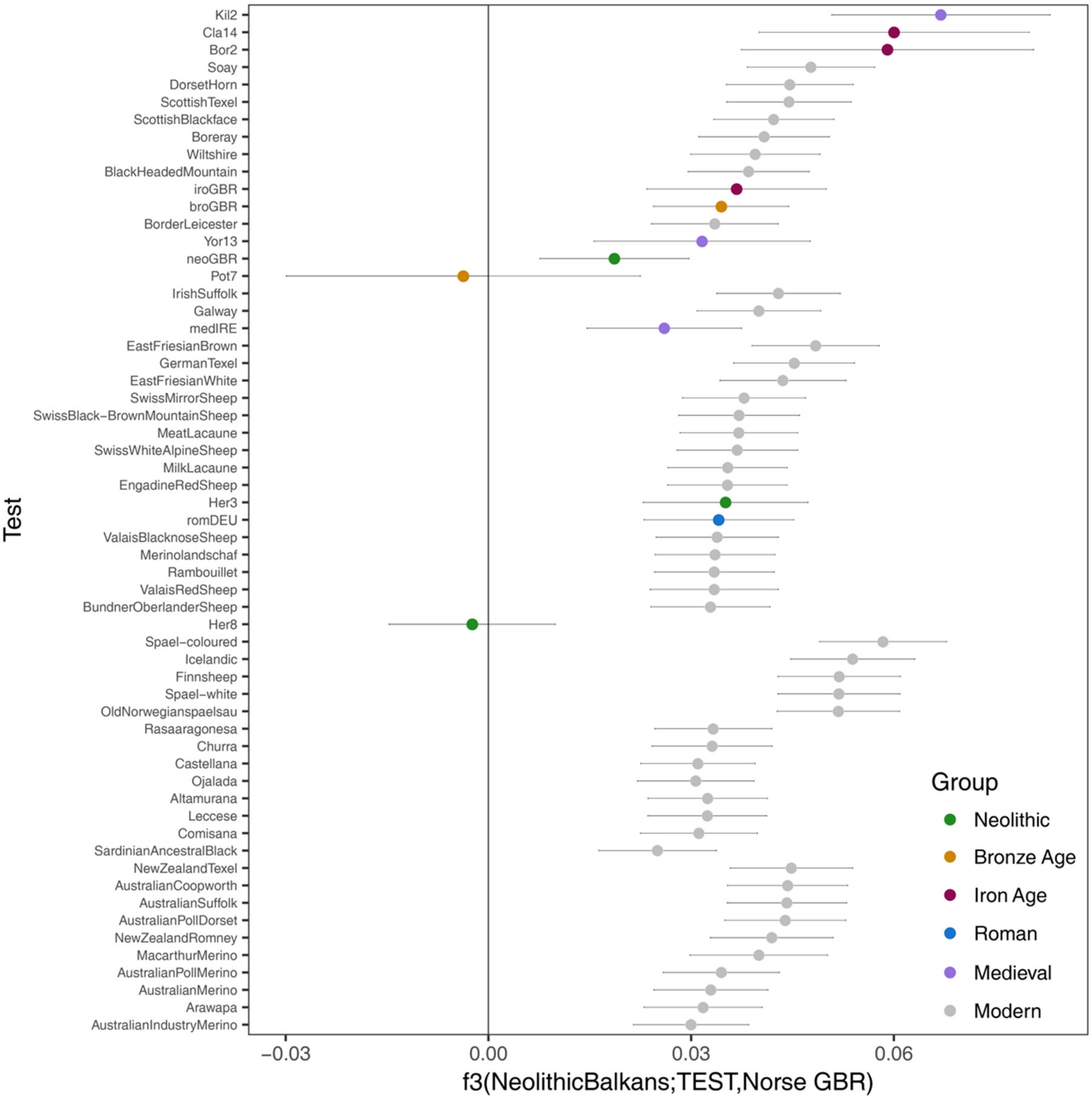


Fig 6.16. *Shared genetic drift, as calculated via the outgroup f3-statistic test, of Medeival Norse and a and variety of European and European derived breeds and ancient samples indicates high levels of drift with ancient Outer Hebridean Islands samples and the modern primitive Scandinavian breeds.*

In comparison, the test f_3(X, Medieval Irish; Neolithic Balkans) (Fig 6.17 & Appendix Table 18) indicates similar levels of shared drift between the Medieval Irish samples and the Irish, Scandinavian and British/British derived breeds, respectively. Therefore it is difficult to discern a clear pattern.

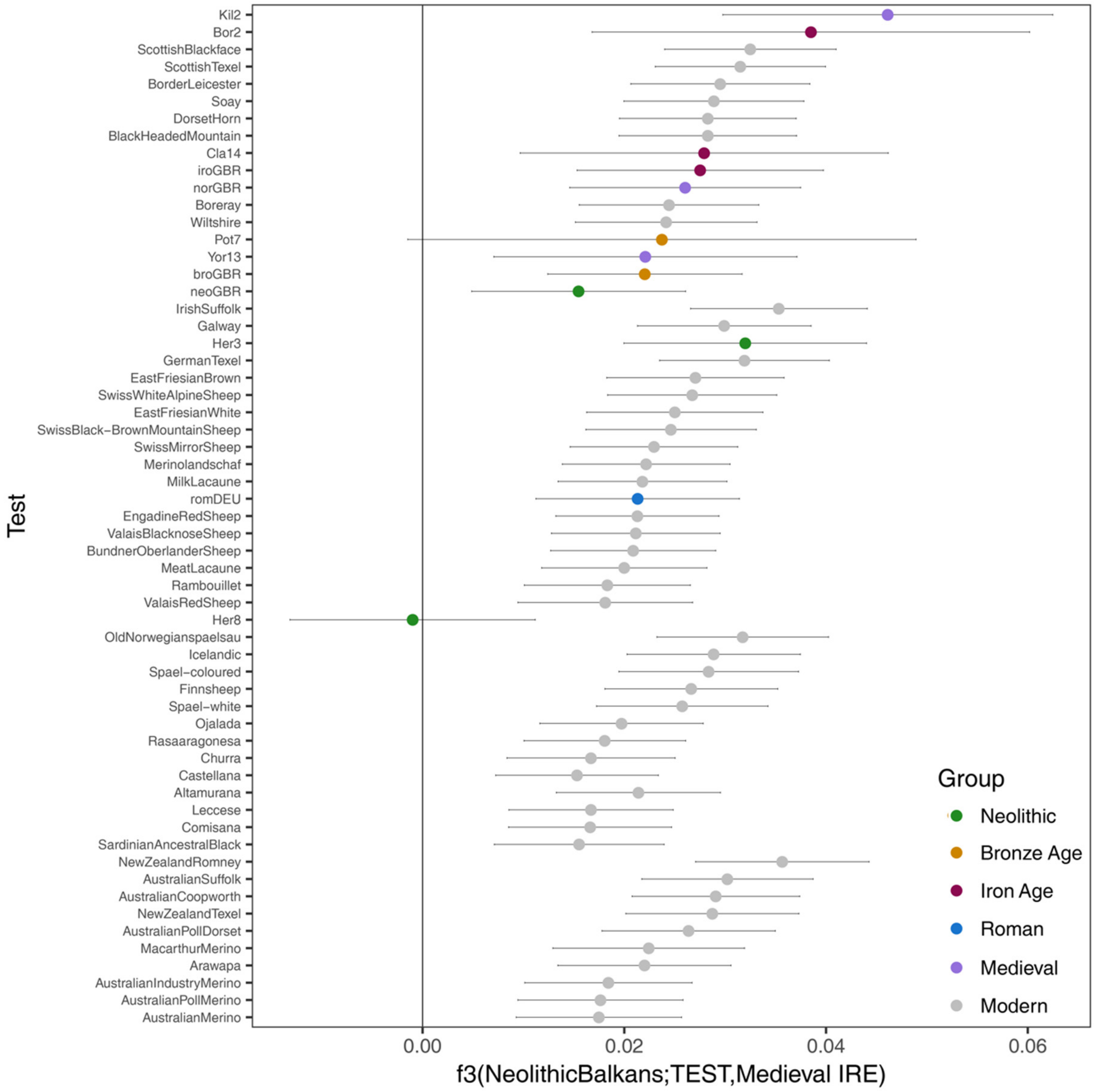


Fig 6.17. *Shared genetic drift, as calculated via the outgroup f3-statistic test, of Medeival Ireland and a and variety of European and European derived breeds and ancient samples indicates similar levels of shared drift between Irish, Scandinavian and British/British derived breeds.*

6.4.5.2. D-statistics

Using *D*-statistics to test the integrity of clades of sheep (Table 6.8 & 6.9), indicates that breeds such as the Soay and Icelandic share more British Neolithic and British Bronze derived alleles than other modern primitive/Northern European breeds. In comparison with other Scandinavian breeds the Icelandic breeds shares the most derived alleles with the Orcadian Medieval (Norse) samples (Table 6.10).

Table 6.8. *D(Pop4,Icelandic)Ancient,Goat), to test for shared derived alleles from ancient populations into modern Northern European breeds. Z scores are coloured for the equivalent significance p value – green= P ≤ 0.001, orange = 0.001 < P ≤ 0.01, red = 0.01 < P ≤ 0.05.*

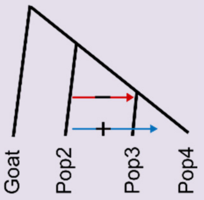
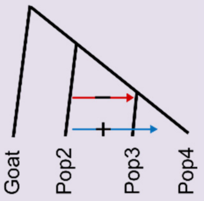
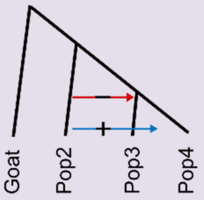
Test	Pop2	Pop3	Pop4	D	stdr error	Z	BABA	ABBA	Sites
	Neolithic GBR	Icelandic	Finnsheep	-0.0128	0.00612	-2.091	1704	1749	18725
			Boreray	-0.0119	0.00866	-1.374	1737	1779	18725
			Old Norwegian spaelsau	-0.007	0.00521	-1.343	1672	1696	18725
			Scottish Blackface	-0.006	0.00656	-0.915	1768	1789	18725
			Spael-coloured	-0.0053	0.00699	-0.758	1650	1668	18725
			Spael-white	-0.0007	0.00598	-0.117	1664	1666	18725
			Norse GBR	0.0074	0.01320	0.562	1431	1409	15608
			Soay	0.0103	0.00802	1.29	1763	1727	18725
			Iron Age GBR	0.0299	0.01381	2.168	1345	1266	13900
			Medieval IRE	0.0333	0.01312	2.54	1407	1316	14673
			Bronze Age GBR	0.0737	0.00997	7.394	1872	1615	18650
	Bronze Age GBR	Icelandic	Finnsheep	-0.0176	0.00554	-3.178	1758	1821	19299
			Old Norwegian spaelsau	-0.0144	0.00484	-2.977	1722	1772	19299
			Scottish Blackface	-0.0117	0.00595	-1.965	1819	1862	19299
			Spael-white	-0.0107	0.00548	-1.954	1709	1746	19299
			Spael-coloured	-0.0113	0.00649	-1.741	1700	1739	19299
			Boreray	-0.0119	0.00812	-1.466	1795	1838	19299
			Norse GBR	-0.0013	0.01161	-0.112	1458	1462	16060
			Soay	-0.0004	0.00816	-0.049	1807	1808	19299
			Iron Age GBR	0.0077	0.01250	0.619	1355	1334	14288
			Medieval IRE	0.0147	0.01182	1.247	1433	1392	15089
	Iron Age GBR	Icelandic	Spael-coloured	-0.0144	0.00823	-1.749	1249	1285	14345
			Boreray	-0.016	0.00930	-1.721	1318	1361	14345
			Old Norwegian spaelsau	-0.0068	0.00574	-1.185	1284	1301	14345
			Finnsheep	-0.0062	0.00657	-0.943	1320	1337	14345
			Scottish Blackface	-0.0053	0.00716	-0.74	1354	1368	14345
			Norse GBR	-0.0056	0.01447	-0.387	1088	1100	11954
			Spael-white	-0.0009	0.00672	-0.134	1274	1276	14345
			Soay	0.0009	0.00921	0.101	1338	1336	14345
			Medieval IRE	0.0021	0.01411	0.15	1063	1059	11302

Table 6.9. *D(Pop4,Soay)Ancient,Goat), to test for shared derived alleles from ancient samples into modern Northern European Breeds. Z scores are coloured for the equivalent significance p value – green= P ≤ 0.001, orange = 0.001 < P ≤ 0.01, red = 0.01 < P ≤ 0.05*

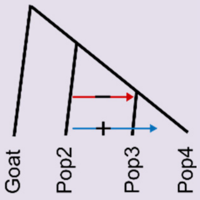
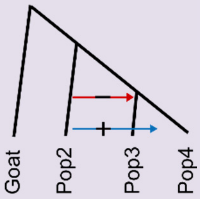
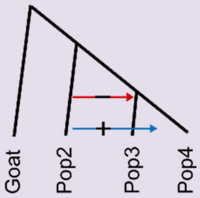
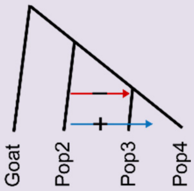
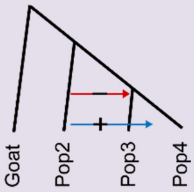
Test	Pop2	Pop3	Pop4	D	stdr error	Z	BABA	ABBA	Sites
	Neolithic GBR	Soay	Finnsheep	-0.0228	0.00799	-2.854	1718	1798	18725
			Boreray	-0.025	0.00920	-2.716	1520	1598	18725
			Old Norwegian spaelsau	-0.0171	0.00812	-2.107	1722	1781	18725
			Scottish Blackface	-0.0163	0.00786	-2.074	1736	1794	18725
			Spael-coloured	-0.0154	0.00931	-1.655	1718	1772	18725
			Spael-white	-0.011	0.00844	-1.304	1732	1771	18725
			Icelandic	-0.0103	0.00798	-1.29	1727	1763	18725
			Norse GBR	0.0028	0.01389	0.202	1458	1450	15608
			Medieval IRE	0.0223	0.01473	1.515	1380	1320	14673
			Iron Age GBR	0.0231	0.01460	1.584	1334	1273	13900
			Bronze Age GBR	0.0641	0.01118	5.734	1853	1630	18650
	Bronze Age GBR	Soay	Finnsheep	-0.0169	0.00734	-2.304	1791	1852	19299
			Old Norwegian spaelsau	-0.0135	0.00765	-1.765	1792	1840	19299
			Scottish Blackface	-0.0114	0.00712	-1.602	1802	1844	19299
			Boreray	-0.013	0.00839	-1.55	1596	1638	19299
			Spael-white	-0.0098	0.00795	-1.232	1797	1833	19299
			Spael-coloured	-0.0103	0.00863	-1.193	1790	1827	19299
			Icelandic	0.0004	0.00738	0.049	1808	1807	19299
			Norse GBR	0.0032	0.01314	0.245	1512	1503	16060
			Iron Age GBR	0.0098	0.01375	0.709	1354	1328	14288
			Medieval IRE	0.016	0.01292	1.242	1419	1374	15089
	Iron Age GBR	Soay	Boreray	-0.019	0.01026	-1.851	1174	1220	14345
			Spael-coloured	-0.0146	0.01056	-1.383	1318	1357	14345
			Old Norwegian spaelsau	-0.0075	0.00881	-0.851	1336	1356	14345
			Finnsheep	-0.007	0.00861	-0.813	1339	1358	14345
			Scottish Blackface	-0.0063	0.00869	-0.725	1340	1357	14345
			Spael-white	-0.0017	0.00881	-0.193	1339	1343	14345
			Icelandic	-0.0009	0.00891	-0.101	1336	1338	14345
			Medieval IRE	0.002	0.01582	0.126	1056	1052	11302
			Norse GBR	0.002	0.01546	0.128	1124	1120	11954

Table 6.10. *D(Pop4,Scandinavian breed)British (Norse) Medieval,Goat), to test for admixture from Orcadian Norse into modern Scandinavian Breeds. Z scores are coloured for the equivalent significance p value – green = P ≤ 0.001, orange = 0.001 < P ≤ 0.01, red = 0.01 < P ≤ 0.05*

Test	Pop2	Pop3	Pop4	D	stdr error	Z	BABA	ABBA	Sites
	Norse GBR	Icelandic	Old Norwegian spaelsau	-0.0204	0.00552	-3.694	1424	1484	16124
			Spael-coloured	-0.0169	0.00780	-2.166	1406	1455	16124
			Spael-white	-0.0152	0.00622	-2.444	1415	1459	16124
			Finnsheep	-0.0205	0.00619	-3.311	1465	1527	16124
	Norse GBR	Finnsheep	Old Norwegian spaelsau	0.0007	0.00536	0.129	1498	1496	16124
			Spael-coloured	0.0044	0.00748	0.586	1504	1491	16124
			Spael-white	0.0059	0.00633	0.93	1503	1485	16124
			Icelandic	0.0205	0.00620	3.311	1527	1465	16124

6.5. Discussion

6.5.1. Mitochondria

Previous studies of ovine mitochondrial DNA have identified five haplogroups (A,B,C,D,E), predominantly typed using small regions of the control region (Meadows et al. 2005; Pedrosa et al. 2005; Chen et al. 2006; Wang et al. 2007; Meadows et al. 2007) and confirmed by complete mitogenomes (Meadows et al. 2011; Lv et al. 2015). The diversity of ovine mtDNA has been hypothesised to be the result of either multiple domestication events, or alternatively, a single domestication event that captured highly divergent lineages (Hiendleder et al. 2002; Meadows et al. 2011). A later autosomal study suggests a common ancestral origin of all domestic breeds of sheep due to conservation of linkage disequilibrium (LD) (Kijas et al. 2012). These five haplogroups show some geographic distribution. Haplogroups A and B are dispersed globally, with A at high frequency in Asia and B at high frequency in Europe (Wood & Phua 1996; Hiendleder et al. 1998).

Haplotype C is limited to the Fertile Crescent, Caucasus mountains, Iberia and North China (Guo et al. 2005; Pedrosa et al. 2005; Tapio et al. 2006; Pereira et al. 2006; Lv et al. 2015). The more recently described haplotypes D and E are restricted to South West Asia (Meadows et al. 2005; Tapio et al. 2006; Lv et al. 2015). Previous ancient DNA studies have focused predominantly on mtDNA demonstrating the presence of haplotype A and B in Europe and Asia (Cai et al. 2007; Cai et al. 2011; E. Rannamäe et al. 2016; Eve Rannamäe et al. 2016). The assignment of the mtDNA haplogroups of the ancient European sheep in this chapter to B concurs with modern analysis of European sheep (Meadows et al. 2005; Meadows et al. 2007) as well as other ancient analyses exploring European samples (Olivieri et al. 2012; Niemi et al. 2013; O'Sullivan et al. 2016; Eve Rannamäe et al. 2016; E. Rannamäe et al. 2016).

6.5.2. Global Phylogeographic Structure

The global modern breed structure as displayed by results of the PCA and ADMIXTURE analyses is consistent with previous published work (Kijas et al. 2009; Kijas et al. 2012). However with the inclusion of ancient autosomal data for the first time, it is apparent that a proportion of this geographic structure is ancient and not solely due to breed formation over the last several hundred years. The split between Asia and Europe was certainly already defined by the Iron Age. Moreover, it is possible this geographic split existed prior to the Iron Age as Neolithic Serbia, Germany and Britain sit on the European arm of the PCA. Additionally, the ancestral components of these samples calculated by ADMIXTURE, indicate a higher proportion of European components for the ancient European samples and Asian components for the ancient Asian samples.

Taken together these analyses suggest a global phylogeographic structure was present in ancient sheep long before modern breeding. Arguments have been made for more than one domestication centres for sheep based in mtDNA (Hiendleder et al. 2002; Zeder 2006), as suggested for other species such as pigs and cattle (Loftus et al. 1994; Larson et al. 2005). Therefore, more detailed sampling from Neolithic Southwest Asia will be necessary to determine if this global structure resulted from several domestication events, or the domestication of a genetically diverse wild population followed by divergence by drift, as suggested by past publications (Meadows et al. 2011; Kijas et al. 2012).

6.5.3. European Population Structure and Migrations

6.5.3.1. Ancient European Population Structure

It is clear that even during the Neolithic there was population structure across the continent, however the extreme placement of the Neolithic Ness of Brodgar samples on the PCA could be due to an island bottleneck having taken place (Frankham 1997; Palkopoulou et al. 2015). Despite the geographic isolation of the Orkney islands, the islands have a rich Neolithic heritage with the UNESCO World Heritage – Heart of Neolithic Orkney which includes Skara Brae and the Ring of Brodgar, therefore geographic isolation should not be interpreted as cultural isolation. Higher genome coverages would allow for testing of runs of homozygosity which, when compared with other samples, can highlight past population bottlenecks (Purfield et al. 2012; Gamba et al. 2014; Cassidy et al. 2016). Nonetheless even if these samples underwent a drastic population bottleneck, the German Neolithic sample Her8 sits apart in between the Serbian and British Neolithic samples on the PCA, suggesting some differentiation in the Neolithic. The sequencing of more ancient Neolithic samples would help to clarify if there is a Neolithic pattern of movement out of the centre of domestication as seen in cattle (Chapter 5).

The British/Irish time series population differentiation could be interpreted as temporal or geographical. Within the British/Irish cluster the samples grouping further apart from the Medieval-Iron Age cluster are Scottish island populations from the Neolithic to the Norse period, while the main cluster is Southern England and Ireland. As previously noted, island populations may undergo reductions in population diversity and increases in relatedness. In turn, this would cluster samples together in analyses such as PCA. However this post-Neolithic/Bronze clustering pattern on the PCA, could also be partly explained by the admixture of the some of the later ancient samples with another type of sheep.

6.5.3.2. Ancient Admixture from Asia into Europe

The results of ADMIXTURE, *D*-statistics and f_3 admixture point toward the addition of a genetic component within the time series of British and Irish sheep, likely during the Bronze or Iron Age. Due to gaps in the series and lack of individual radiocarbon dates for these samples, an exact time point cannot be established. This pattern of admixture has been detected in ADMIXTURE, f_3-statistics and *D*-statistics using modern breeds from Asia and Southwest Asia as a proxy for an introgressing ancient population. It is widely accepted that sheep were first domesticated in the Neolithic for meat, followed by a secondary improvement for wool likely in Southwest Asia during the Bronze Age (Zeder 2008; Chessa et al. 2009; Meadows et al. 2011; Kijas et al. 2012). It is hypothesised that these improved wool animals then spread across Southwest Asia, Asia, Africa and Europe (Chessa et al. 2009). A recent aDNA retroviral insertions study supported a secondary sheep migration into the Caucasus mountains between the Bronze and Iron Age (Schroeder et al. 2017). Perhaps this admixture that occurs in the British time series, could be the introduction and subsequent admixture of this secondary population, perhaps an improved wooly sheep. Although the Late Bronze samples from Cladh Hallan do not possess this genetic component in ADMIXTURE, it is plausible that sheep on mainland Britain had already began to admix and this component had not reached the Outer Hebrides, however this can not be tested without additional ancient samples.

Some of the *D*-statistic results for the ancient samples/populations are of borderline significance, for example the *D*-statistics for the British Iron Age samples from Danebury in the South of Britain never have a *Z*-score ≥ 3, likely due to low SNP number in these analyses. However there is a general pattern of increase of shared Asian/Southwest Asian derived alleles post-Neolithic in Europe. Moreover there are several limitations to these D-statistics tests. The number of sites these tests are performed on are very low; below 20,000 for group tests and in general below 5,000 when ancient individuals are tested. Secondly both transitions and transversions had to be included to keep the number of SNPs at a reasonable level. Perhaps with an increase in the density of the data this pattern would be more easily resolved. While this data results from UDG treatment and reads are soft clipped, it is plausible that some DNA damage sequence error is still present which could affect the *D*-statistic, especially when low numbers of SNPs are used. Therefore the *D*-statistic results alone should not be interpreted as proof of introgression, however when combined with other analysis they can help to build a picture of admixture.

It is clear that a complete replacement of the European Neolithic/Bronze population by a new type of sheep did not occur, as PCA shows clustering of the ancient samples with modern European variation and no mitochondrial haplotype turnover is detected. Results of the f_3-statistic test for admixture results indicate the Iron Age (South Britain), Medieval (Irish) and Roman German populations to be descended from admixture between British Neolithic or Bronze and Modern breeds from Asia/Southwest Asia. As the f_3 admixture statistic, unlike *D*, is not impervious to drift the inability to detect admixture does not discount it from having occurred (Patterson et al. 2012).

One ancient sample that is likely mis-classified as Neolithic is Her3. Whilst the site itself is dated to the Neolithic (Amelie Scheu *per. comm.*) this sample behaves differently in the

PCA, ADMIXTURE and *D*-statistics to other Neolithic samples, and is likely a more recent individual. This sample needs to be radiocarbon dated before any conclusions are drawn.

It should be noted that while the ADMIXTURE analysis shows the two Serbian Neolithic samples to have an Asian component at K=3 and a Near Eastern component at K=6, this result is not replicated in *D*-statistics. It is possible that this component in the ADMIXTURE is merely noise in these samples. However, one could argue that the ADMIXTURE result is correct and the *D*-statistics lack power. Perhaps, the Neolithic Eastern European sheep populations closest to the centre of domestication retained a higher genetic diversity than Northern Europe, paralleling the pattern seen in cattle mitochondria (Chapter 4&5). Perhaps due to migration to and from the domestication centre in the Early Neolithic as seen in early cattle migrations (Scheu et al. 2015), or because they were not subject to later bottlenecks specific to the further migrations west. Diversity was likely reduced as sheep migrated with the Neolithic through Europe, likely via the Danubian route as modern mediterranean sheep have been shown to have high levels of heterozygosity (Kijas et al. 2012). Then at some point, post-Neolithic/Bronze Age, the Asia/Southwest Asian genetic component entered Northern Europe possibly via new wooly sheep populations formed in Southwest Asia (Chessa et al. 2009). More ancient genomes from a variety of time periods from continental Europe and Southwest Asia are needed to fully resolve this situation.

6.5.3.3. Relationship of Modern European Primitive Breeds to Ancient Britain and Ireland

Past publications have sought to understand possible sheep migrations through Europe and Britain using morphological characteristics (Ryder 1964; Ryder 1981), and more recently through the use of retroviral insertions (Chessa et al. 2009; Bowles et al. 2014) and SNP data (Kijas et al. 2012). The inclusion of some of the European primitive breeds in the 50K SNP data, allows for the analysis of these breeds with the ancient samples. The short-tail breeds of both Scandinavia and the island breeds of Soay and Boreray are all classed as primitive breeds. This classification is due to morphological characteristics such as possession of a primitive moulting coat, coloured coat either brown or a variety of colours and short tails. Genetically these breeds have been shown to have a lack of retroviral insertions or lack/low frequency of the most common retorviral haplogroup R2, indicative of a primitive status, believed to be a relicit of the Iron Age (Ryder 1981; Chessa et al. 2009). It is reasonable to propose that the Soay breed, which is now limited to the island of St Kilda in the Outer Hebrides, is a relict of the early European sheep population, and is regarded as a relicit of the Bronze Age (Ryder 1964; Ryder 1981). Out of the modern breeds it is the most closely related to the European mouflon, a modern wild sheep believed to be of feral early domestic decent (Poplin 1979; Chessa et al. 2009). In comparison, the other primitive breeds are suggested to be Iron Age relicts pushed to the extremities of Europe, possibly by the white wool sheep believed to have been moved by the Romans (Ryder 1981).

In addition to these primitive breeds, other hardy British breeds such as the Scottish Blackface and the Herdwick have been classed as Northern European along with the other primitive breeds (Kijas et al. 2012) and have been demonstrated to have low frequencies of the R2 retroviral haplotype (Bowles et al. 2014). Consequently the groupings of the ancient British and Irish samples with these Northern European and primitive breeds in outgroup f_3-statistics and the PCA helps to confirm the conclusion from previous studies that these breeds are more similar to early European sheep than other modern British derived breeds such as Wiltshire, Dorset Horn and Texel (Ryder 1981; Chessa et al. 2009; Kijas et al. 2012; Bowles et al. 2014).

However the hypothesis based on morphological and retroviral insertions, that Soay is more primitive than the other primitive breeds, cannot be fully resolved. If the breed is more primitive, one would expect it to group with the Neolithic and Bronze samples on a PCA and higher values of shared drift in the outgroup f_3-statistic. Intriguingly, out of the modern breeds it is the Scandinavian short-tailed animals that share the most shared drift with the Neolithic, Bronze Age and Iron Age, however this pattern is most apparent in the Iron Age when regarding the results of these three periods of prehistory. All of the primitive breeds show an increase in shared derived alleles of Asian/Southwest Asian origin in comparison to Neolithic British samples, and Asian genetic componets in the ADMIXUTRE anaylsis, suggesting that even if the Soay are representative of the ancestral type, admixture has occurred at some point. It should be mentioned, that the 50K SNP chip is not ideal for understanding the population ancestry of primitive breeds, especially island breeds like the feral Soay that undergoes regular population fluctuations (Coulson et al. 2001), as no primitive short-tailed sheep were used in SNP discovery (Kijas et al. 2012).

Nonetheless, using *D*-statistics to test clade integrities has shown the Soay and Icelandic breeds share more derived alleles with Neolithic and/or Bronze Age ancient British. Perhaps these results occur because they have less admixture with other modern breeds or an Asian type breed than the other primitive animals. For example these two breeds have very low or no frequency of the global R2 haplotype of retroviral insertions, suggested to have accompanied the second wave of sheep from Southwest Asia (Chessa et al. 2009). While in low frequencies the other primitive breeds do posses the haplotype, suggestive of admixture at some point in time. It is also plausible that low levels of diversity, nonrandom mating (Lawson Handley et al. 2007), small population size (McRae et al. 2005) and possibly even non-representative population sampling (Kijas et al. 2009), as seen in Soay, could skew these results. In previous microsatellite analysis of 29 different domestic sheep breeds both Soay and Icelandic have been classed as the most genetically pure (free from admixture) and distinctive breeds (Lawson Handley et al. 2007).

The British Medieval samples from the Norse settlements of Snusgar on Orkney (Ork1 & Ork2), Kilpheder on South Uist of the Outer Hebrides (Kil2) and York in England (Yor13), behave somewhat differently to the Medieval Norse samples from Dublin (Dub3, Dub4 & Dub5). The Orkney samples cluster closer to the primitive Scandinavian breeds than the samples from the other sites, which is mirrored in the outgroup f_3-statistics. It is documented that the Vikings moved animals; for example Icelandic breeds are descended from animals brought by the Vikings (as reviewed by McGovern 1990). Perhaps these two samples represent ancient Scandinavian sheep, or British sheep admixed with ancient Scandinavian sheep.

In particular the two Norse Orcadian samples show a different pattern of genetic components in ADMIXTURE when compared with other British samples. At K=9 the two samples look similar to Finnsheep, with a larger proportion of the Scandinavian component when compared with other ancient samples. Modern Orkney sheep are still short-tailed primitive animals, said to have Scandinavian influence, although it is suggested they could be relicts of ancient British sheep (Ryder 1981).

Moreover, not only does the Icelandic breed share more derived alleles with the Neolithic and Bronze than other Scandinavian breeds, but also with the Norse Orkney samples. Either suggesting admixture into the Icelandic breed from the Norse Orcadian samples, perhaps due to animals from Orkney contributing to Icelandic stock or simply the Icelandic animals again retain ancestral alleles that the other Scandinavian breeds do not due to more recent admixture. Thus it may be inferred that human migrations of Anglo-Saxons and Vikings likely resulted in genetic contributions to British sheep, as hypothesised for some upland breeds (Ryder 1981; Bowles et al. 2014).

Utilising modern breeds to understand domestication and past migrations, can be complicated by historic movement of livestock both pre and post breed. However, these results do support previous designations of these modern breeds as primitive in comparison to other modern breeds.

6.5.4. Limitations and Future Analysis

The 50K SNP chip was designed to separate modern breeds for global variation (Kijas et al. 2012), and the genotype data published is limited. Therefore the ability to resolve more closely related breeds such as British Isles or primitive breeds is reduced. If more modern breeds were genotyped for the 50K this would improve; especially inclusion of primitive breeds such as Orkney Sheep, the Mediterranean Mouflon and upland British sheep breeds such as Herdwick (Bowles et al. 2014). The study of (Kijas et al. 2012) notes that the variation accounted for by the 20 largest PCs only accounts for 16% of the total, suggesting sheep to have a weak population structure (Kijas et al. 2009). Thus the typing of more breeds may not overcome the SNP dataset's power to resolve closely related breeds. As SNP arrays do not include rare variation, to analyse primitive breeds and ancient data, there is a need to move away from genotyping modern animals on SNP chips toward analyzing whole genome data.

Moreover, modern SNP-restricted variation does not account for genetic variation present in the ancient samples which has subsequently been lost. Therefore ancient data should not only be analysed in the context of modern variation but in that of ancient variation also. Unfortunately the coverage of the ancient data presented here does not suffice for SNP discovery. Future analysis should focus on sequencing ancient samples to a coverage greater to allow for SNP discovery and diploid calls, as well as making full use of resources such as the Sheep Genome Database (http://sheepgenomesdb.org/), which provides whole genome variant calls for published modern genomes. Another approach would be to sequence ancient sheep to an average coverage of 2-4x and modern genomes from a variety of breeds to the same or greater coverage and implement analysis through ANGSD where genotype likelihoods are utilised rather than genotype calls (Korneliussen et al. 2014).

In addition to the ancient data shown here, more ancient data from Southwest Asia and Europe is necessary to ensure a full understanding of domestic sheep dispersal and subsequent migrations and admixture events. As currently the wild progenitor of the sheep is not fully established (Meadows et al. 2011), the sampling of pre-domestication *Ovis* populations would not only enable selection sweeps to be identified pre-breed formation, but also a greater understanding of sheep domestication.

6.6. Conclusion

To fully understand the domestication history of an animal such as the sheep, *Ovis aries*, the use of ancient genomes is imperative, as historic population movement and the formation of breeds can confound conclusions based solely on modern data. The ancient genomes presented illuminate sheep population structure prior to the formation of modern breeds. Despite the limitations of the 50K SNP dataset, it is apparent that prior to modern breeding, geographic structure similar to that of modern breeds existed in the global sheep population. Specifically, a phylogeographic divide between European and Asian populations was likely present prior to the Iron Age. Differentiation between European Neolithic populations indicates possible bottlenecks occurring during the migration to the west, while within the British and Irish time-series genetic structure is present between the different cultural periods. Additionally, the primitive breeds of Northern Europe show a genetic affiliation towards the ancient samples, supporting the long held view of their primitive status. Lastly, tentative conclusions of subsequent migrations of sheep from Southwest Asia into Europe post-Bronze Age, provide a foundation which future work with ancient genomes should build on.

6.7. References

Alexander, D.H., Novembre, J. & Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, 19(9), pp.1655–1664.

Bowden, P.J., 2013. *Wool Trade in Tudor and Stuart England*, Routledge.

Bowles, D., Carson, A. & Isaac, P., 2014. Genetic distinctiveness of the Herdwick sheep breed and two other locally adapted hill breeds of the UK. *PloS one*, 9(1), p.e87823.

Cai, D. et al., 2011. Early history of Chinese domestic sheep indicated by ancient DNA analysis of Bronze Age individuals. *Journal of archaeological science*, 38(4), pp.896–902.

Cai, D.-W. et al., 2007. DNA analysis of archaeological sheep remains from China. *Journal of archaeological science*, 34(9), pp.1347–1355.

Cassidy, L.M. et al., 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), pp.368–373.

Chen, S.-Y. et al., 2006. Origin, genetic diversity, and population structure of Chinese domestic sheep. *Gene*, 376(2), pp.216–223.

Chessa, B. et al., 2009. Revealing the history of sheep domestication using retrovirus integrations. *Science*, 324(5926), pp.532–536.

Coulson, T. et al., 2001. Age, sex, density, winter weather, and population crashes in Soay sheep. *Science*, 292(5521), pp.1528–1531.

Fiddyment, S. et al., 2015. Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting. *Proceedings of the National Academy of Sciences of the United States of America*, 112(49), pp.15066–15071.

Frankham, R., 1997. Do island populations have less genetic variation than mainland populations? *Heredity*, 78 (Pt 3), pp.311–327.

Gamba, C. et al., 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nature communications*, 5, p.5257.

Guo, J. et al., 2005. A novel maternal lineage revealed in sheep (Ovis aries). *Animal genetics*, 36(4), pp.331–336.

Hiendleder, S. et al., 1998. Analysis of mitochondrial DNA indicates that domestic sheep are derived from two different ancestral maternal sources: no evidence for contributions from urial and argali sheep. *The Journal of heredity*, 89(2), pp.113–120.

Hiendleder, S. et al., 2002. Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proceedings. Biological sciences / The Royal Society*, 269(1494), pp.893–904.

Johnston, S.E. et al., 2011. Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular ecology*, 20(12), pp.2555–2566.

Kijas, J.W. et al., 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PloS one*, 4(3), p.e4668.

Kijas, J.W. et al., 2012. Genome-wide analysis of the world’s sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS biology*, 10(2), p.e1001258.

Korneliussen, T.S., Albrechtsen, A. & Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC bioinformatics*, 15, p.356.

Larson, G. et al., 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science*, 307(5715), pp.1618–1621.

Lawson Handley, L.-J. et al., 2007. Genetic structure of European sheep breeds. *Heredity*, 99(6), pp.620–631.

Loftus, R.T. et al., 1994. Mitochondrial genetic variation in European, African and Indian cattle populations. *Animal genetics*, 25(4), pp.265–271.

Lv, F.-H. et al., 2015. Mitogenomic Meta-Analysis Identifies Two Phases of Migration in the History of Eastern Eurasian Sheep. *Molecular biology and evolution*, 32(10), pp.2515–2533.

McGovern, T.H., 1990. The Archeology of the Norse North Atlantic. *Annual review of anthropology*, 19(1), pp.331–351.

McKenna, A. et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), pp.1297–1303.

McRae, A.F., Pemberton, J.M. & Visscher, P.M., 2005. Modeling linkage disequilibrium in natural populations: the example of the Soay sheep population of St. Kilda, Scotland. *Genetics*, 171(1), pp.251–258.

Meadows, J. & Kijas, J.W., 2009. Re-sequencing regions of the ovine Y chromosome in domestic and wild sheep reveals novel paternal haplotypes. *Animal genetics*, 40(1), pp.119–123.

Meadows, J.R.S. et al., 2007. Five ovine mitochondrial lineages identified from sheep breeds of the near East. *Genetics*, 175(3), pp.1371–1379.

Meadows, J.R.S. et al., 2005. Mitochondrial sequence reveals high levels of gene flow between breeds of domestic sheep from Asia and Europe. *The Journal of heredity*, 96(5), pp.494–501.

Meadows, J.R.S., Hiendleder, S. & Kijas, J.W., 2011. Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel. *Heredity*, 106(4), pp.700–706.

Niemi, M. et al., 2013. Mitochondrial DNA and Y-chromosomal diversity in ancient populations of domestic sheep (Ovis aries) in Finland: comparison with contemporary sheep breeds. *Genetics, selection, evolution: GSE*, 45, p.2.

Niemi, M. et al., 2015. Temporal fluctuation in North East Baltic Sea region cattle population revealed by mitochondrial and Y-chromosomal DNA analyses. *PloS one*, 10(5), p.e0123821.

Olivieri, C. et al., 2012. Phylogenetic position of a copper age sheep (Ovis aries) mitochondrial DNA. *PloS one*, 7(3), p.e33792.

O'Sullivan, N.J. et al., 2016. A whole mitochondria analysis of the Tyrolean Iceman's leather provides insights into the animal sources of Copper Age clothing. *Scientific reports*, 6, p.31279.

Palkopoulou, E. et al., 2015. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Current biology: CB*, 25(10), pp.1395–1400.

Paradis, E., Claude, J. & Strimmer, K., 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* , 20(2), pp.289–290.

Patterson, N. et al., 2012. Ancient admixture in human history. *Genetics*, 192(3), pp.1065–1093.

Pedrosa, S. et al., 2005. Evidence of three maternal lineages in Near Eastern sheep supporting multiple domestication events. *Proceedings. Biological sciences / The Royal Society*, 272(1577), pp.2211–2217.

Pereira, F. et al., 2006. Genetic signatures of a Mediterranean influence in Iberian Peninsula sheep husbandry. *Molecular biology and evolution*, 23(7), pp.1420–1426.

Peter, C. et al., 2007. Genetic diversity and subdivision of 57 European and Middle-Eastern sheep breeds. *Genetics*. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2052.2007.01561.x/full.

Poplin, F., 1979. Origine du Mouflon de Corse dans une nouvelle perspective paléontologique: par marronnage. In *Annales de génétique et de sélection animale*. BioMed Central, p. 133.

Purcell, S. et al., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*, 81(3), pp.559–575.

Purfield, D.C. et al., 2012. Runs of homozygosity and population history in cattle. *BMC genetics*, 13, p.70.

Rambaut, A., 2007. FigTree, a graphical viewer of phylogenetic trees. *See http://tree. bio. ed. ac. uk/software/figtree*.

Rannamäe, E. et al., 2016. Maternal and paternal genetic diversity of ancient sheep in Estonia from the Late Bronze Age to the post-medieval period and comparison with other regions in Eurasia. *Animal genetics*, 47(2), pp.208–218.

Rannamäe, E. et al., 2016. Three Thousand Years of Continuity in the Maternal Lineages of Ancient Sheep (Ovis aries) in Estonia. *PloS one*, 11(10), p.e0163676.

R Core Team, 2015. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Ryder, M.L., 1981. A survey of European primitive breeds of sheep. *Annales de génétique et de sélection*, 13(4), pp.381–418.

Ryder, M.L., 1984. Medieval Sheep and Wool Types. *The Agricultural history review*, 32(1), pp.14–28.

Ryder, M.L., 1964. The History of Sheep Breeds in Britain. *The Agricultural History Review*, 12, pp.65–82.

Scheu, A. et al., 2015. The genetic prehistory of domesticated cattle from their origin to the spread across Europe. *BMC genetics*, 16, p.54.

Schroeder, O. et al., 2017. Endogenous Retroviral Insertions Indicate a Secondary Introduction of Domestic Sheep Lineages to the Caucasus and Central Asia between the Bronze and Iron Age. *Genes*, 8(6). Available at: http://dx.doi.org/10.3390/genes8060165.

Tapio, M. et al., 2006. Sheep mitochondrial DNA variation in European, Caucasian, and Central Asian areas. *Molecular biology and evolution*, 23(9), pp.1776–1783.

Troy, C.S. et al., 2001. Genetic evidence for Near-Eastern origins of European cattle. *Nature*, 410(6832), pp.1088–1091.

Wang, C. et al., 2015. Improved ancestry estimation for both genotyping and sequencing data using projection procrustes analysis and genotype imputation. *American journal of human genetics*, 96(6), pp.926–937.

Wang, X. et al., 2007. Genetic and phylogenetic studies of Chinese native sheep breeds (Ovis aries) based on mtDNA D-loop sequences. *Small ruminant research: the journal of the International Goat Association*, 72(2), pp.232–236.

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Available at: http://ggplot2.org.

Wood, N.J. & Phua, S.H., 1996. Variation in the control region sequence of the sheep mitochondrial genome. *Animal genetics*, 27(1), pp.25–33.

Zeder, M.A. et al., 2006. Documenting domestication: the intersection of genetics and archaeology. *Trends in genetics: TIG*, 22(3), pp.139–155.

Zeder, M.A., 2008. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the National Academy of Sciences of the United States of America*, 105(33), pp.11597–11604.

Chapter 7. General Conclusions

7.1. Summary Of Results

The process of domestication has long interested archaeologists and population geneticists, alike. A focus for both modern DNA and ancient DNA past publications, the contribution of genetics to our understanding of domestication has included identifying likely domestication centres, admixture events and selection processes (Loftus et al. 1994; Bradley et al. 1996; Troy et al. 2001; Edwards et al. 2007; Park et al. 2015; Librado et al. 2017; Loog et al. 2017). The improvements in sequencing technology, aDNA recovery and computational approaches has led to a revolution within the aDNA community, with many publications of hominid whole genomes (Rasmussen et al. 2010; Meyer et al. 2012; Olalde et al. 2014; Gamba et al. 2014; Allentoft et al. 2015; Cassidy et al. 2016). However, only a handful of publications have applied NGS technology to analyse nuclear aDNA in order to further understand the process of domestication (Schubert et al. 2014; Skoglund et al. 2015; Frantz et al. 2016; Librado et al. 2017; Park et al. 2015).

This thesis has demonstrated the value of the sequencing of multiple ancient whole genomes in order to understand the process of domestication and subsequent dispersal of domesticates. A total of 113 whole genomes and 127 mitogenomes have illuminated the past population histories of European cattle and Eurasian sheep, demonstrating migration and admixture events in both species. The aim of this following chapter is to provide a brief overview of the conclusions from each of these studies, and discusses future role of ancient DNA in this research area.

7.1.1. Screening of aDNA Samples

Chapter 3 presents the endogenous percentage results of 183 ancient animal remains. Through the targeting of the petrous bone, first identified as a potential bone element for superior DNA preservation by (Gamba et al. 2014), high percentages of endogenous DNA were recovered from Northern Europe. The British and Irish timesect of cattle and sheep demonstrated an uneven number of males and females, suggesting some form of herd management occurring in prehistory.

7.1.2. Mitochondrial Analysis of Ancient *Bos* Samples

Chapter 4 presents a mtDNA analysis on 127 new ancient genomes, demonstrating that while the matrilineal single locus has limitations, the study of ancient mtDNA is still warranted. With new haplotypes emerging from ancient samples it is apparent that the diversity once present in the wild population is still yet to be fully uncovered. Due to poor preservation of endogenous DNA the sequencing of whole genomes is not always cost effective, therefore target capture of mtDNA is a viable alternative.

Ancient mitogenomes from across Europe, the Near East and Africa confirm previous publications of modern and ancient D-loop sequences, showing the high haplogroup diversity from the Near East dwindling as domestic cattle moved into Europe (Loftus et al. 1994; Troy et al. 2001; Scheu et al. 2015). Two Bronze Age African cattle advocate for ancient haplogroup diversity in Egypt as opposed to this being a modern phenomenon, additionally confirming the presence of the T1 haplogroup in Africa dates to at least the Egyptian Bronze Age.

The Bayesian analysis (BEAST) dated for the first time a number of branches, including the C/R/Kir4 vs P/Q/T split and the P vs Q/T, dating to 60,055 yrBP and 19,259 yrBP respectively. The taurine vs indicine split was dated to 127,446 yrBP supports the long held view that the modern *B. taurus* and *B. indicus* were domesticated from diverse populations of auroch (Loftus et al. 1994). The TMRCA of the predominate domesticate macro haplogroup, T, dates to 10,704 yrBP coincides with domestication of cattle (Helmer et al. 2005), suggesting much of the modern European cattle haplotype diversity has occurred post-domestication. The dates presented here, while shifted towards a more recent timeframe, overlap with previously published data (Ho et al. 2008; Edwards et al. 2007; Massilani et al. 2016), and can be linked to environmental conditions.

The preliminary analysis of the successful capture and subsequent sequencing of DNA from a 200,000 year old British auroch, only furthers the argument for continuing of target capture and sequencing of mtDNA especially for low endogenous DNA samples. As demonstrated here mtDNA still has a role to play in understanding past population events.

7.1.3. Whole Genome Analysis of Ancient Near Eastern and European *Bos* Samples

Chapter 5 analyses 113 whole genomes Bos ranging geographically from the Near East to Ireland and temporally from the Mesolithic to the Medieval Period. Building upon the mtDNA analysis, further support for a Near Eastern domestication centre of *Bos taurus* and subsequent migration into Europe is given, as previously hypothesized by others (Loftus et al. 1994; Troy et al. 2001; Edwards et al. 2007).

The analysis of 42 Neolithic genomes demonstrate genetic differentiation and population histories likely caused by a number of population bottlenecks and European auroch admixture events, echoing their migration from the Near East through to the Atlantic Edges. Through the PCA the distinctive visual clusters of the Neolithic samples demonstrate both geographical and temporal affinity, with samples furthest apart in both geography and time clustering the greatest distance from each other. A pattern that is repeated via the analysis of shared genetic drift and clade integrity. Admixture with European auroch contributed significantly to the genomes of the ancient cattle, likely several different admixture events occurred with highs of introgression demonstrated in Neolithic Britain, a pattern still evident in the genomes of modern breeds today (Park et al. 2015).

The time series through the Atlantic Edge demonstrates a greater geographical affinity than temporal, differing from human populations genetics from this region (Cassidy et al. 2016; Olalde et al. 2017). The populations of post-Neolithic Britain and Ireland form one large cluster of samples on the PCA, while testing for clade integrity and shared drift suggests a closer relationship of the British Bronze Age to the Irish Neolithic than to the British Neolithic. Additionally, through the study of auroch admixture it is apparent that an increase in auroch introgression occurred after the Irish Bronze Age. Due to the lack of auroch on the island of Ireland this has to have occurred via admixture with previously admixed animals, therefore indicating trade. Homogenisation of the two island populations is supported archaeology, during the Bronze Age the Atlantic Edge was connected by seafaring communities and many similarities of cultural practices across the islands (Cunliffe 2013). The pattern of auroch admixture does differ between Britain, Ireland and the Netherlands, with the height of auroch admixture occurring in the Neolithic in Britain, Bronze Age in the Netherlands and the Medieval period in Ireland, suggesting that while these populations were interwoven, there are still differences in the population histories of cattle within the Atlantic Edge.

Additionally, by assessing the allele presence of the known causative mutations in the DGAT1 gene on Chr14 in all the cattle genomes, the tentative conclusion of the selection for greater milk yields post the appearance of lactase persistence in ancient humans can be formed (Gamba et al. 2014).

7.1.4. Analysis of Ancient Whole Genome Data of *Ovis aries*

This chapter anaylsed 28 ancient sheep whole genomes from Eurasia, with the predominate focus region of Britain and Ireland. The ancient genomes presented illuminate sheep population structure prior to the formation of modern breeds, highlighting geographic structure similar to that of modern breeds (Kijas et al. 2012) existed in the ancient global sheep population. Specifically the Asia vs Europe split, as demonstrated by the PCA, was present by the Iron Age and likely before.

Previous publications of sheep retroviral insertions in modern animals, and more recently in Bronze/Iron Age Caucasus samples, have demonstrated a possible secondary migration of sheep from Southwest Asia into Europe and Asia (Chessa et al. 2009; Schroeder et al. 2017). It has been hypothesised that this was the introduction of a wool-type sheep. Results of ADMIXTURE, f3 and D-statistics add support for a secondary migration into Europe post-Bronze Age. However, future work is necessary including the sequencing of more ancient genomes before substantiated claims can be made.

A number of breeds genotyped for the 50K Sheep HapMap SNPs are deemed primitive breeds, reminiscent of prehistory. The Soay, a relict from the Bronze Age, while the Scandinavian breeds such as Icelandic and Norwegian Spaelsau are deemed Iron Age relicts (Ryder 1981; Chessa et al. 2009). Analysis of these breeds through the *D*-statistic and outgroup f_3-statistics of the British and Irish time series suggest these breeds do share a closer genetic relationship to the ancient samples than other modern breeds. In particular the Medieval Orcadian samples from the Norse settlement of Snusgar show highest shared drift with the Scandinavian breeds, relative to other ancient samples, while testing for admixture using the D-statistic suggests admixture into the Icelandic breed from the Norse Orcadian samples, perhaps due to animals from Orkney contributing to Icelandic stock or simply the Icelandic animals again retain ancestral alleles that the other Scandinavian breeds do not due to more recent admixture. In general, the results presented here support the long held view of the primitive status of these sheep breeds.

7.2. Limitations and Future Analysis

To understand the domestication of animals the utilisation of ancient DNA is an essential component. Breed formation and historical population movement can confound results solely based upon modern data. However, to date there are no published ancient genomes for domestic cattle and sheep, therefore the analysis of modern genomes and ancient/modern mitogenomes has been the genetic currency for the recent analysis of these two species. The analyses presented here look to change the focus of domestication studies of these two species onto ancient and modern whole genome data, supplemented by mitogenomes.

While the results of this thesis illuminate the prehistory of cattle and sheep there are a number of limitations. The study of ancient genomes in the context of modern SNP-restricted variation does not account for genetic variation present in the ancient samples which has subsequently been lost. For future analysis both modern and ancient data should be analysed in the context of ancient variation in addition to modern variation. To achieve this two routes can be followed; SNP discovery which requires high coverage diploid genomes or the use of programs such as ANGSD where genotype likelihoods are utilised rather than genotype calls (Korneliussen et al. 2014). In addition, many of the tests here were performed using pseudodiploid data ancient data, taking one read and duplicating it to create a diploid call. While this technique helps to circumnavigate the issues of low coverage data, it simplifies the genetic diversity within an individual and population as no heterozygotes are present and likely biases towards the reference genome allele (Martiniano et al. 2016). Again using a programme like ANGSD (Korneliussen et al. 2014) for some tests would help to alleviate this issue.

In the study of ancient and modern human populations, haplotype-based approaches have demonstrated higher power in identifying fine genetic structure than unlinked genetic loci (Lawson et al. 2012; Cassidy et al. 2016; Martiniano et al. 2016; Gilbert et al. 2017). Programs such as fineSTRUCTURE (Lawson et al. 2012) could be used to identify structure that at present cannot be identified by analysis such as PCA and model-based ancestry estimation such as ADMIXTURE (Alexander et al. 2009). However diploid SNP calls or imputed diploid SNP calls are necessary for this analysis. Therefore either more samples sequenced to a higher coverage are necessary or the testing of genome-wide imputation in ancient domesticates needs to occur.

Utilising ancient genomes for selection analysis has so far been limited (Park et al. 2015; Librado et al. 2017). The sequencing of domesticate populations from different geographical and temporal transects will help to inform how selection for different traits has shaped the genome modern and ancient domesticates. In cattle and sheep interesting questions surround milk and wool production, traditionally thought of as secondary products. It would also be interesting to model allele trajectories to understand the timings of these selection process (Loog et al. 2017), thereby indicating if traits were selected for in prehistory or during modern breed formation.

In cattle the sequencing of ancient genomes from Africa and Asia will help to understand the domestication and migrations events that surround the modern day population structure. While more, and higher coverage, ancient genomes from Europe will enable selection scans and demographic modelling of whole genome data to be performed. More ancient genomes of sheep, of higher coverages, from Europe and Asia will help to clarify ancient population structure and establish the different migration events that occurred. In addition, more higher coverage genomes will enable the exploration of selection for different secondary products such as wool and milk.

7.3. Conclusion

The use of ancient DNA recovered from archaeological material can enrich our understanding of the domestication process and subsequent migrations prior to the development of modern breeding. This thesis has demonstrated the power of next generation sequencing (NGS) of ancient domestic genomes in order to analyse past populations of two domesticate species, cattle (*Bos taurus*) and sheep (*Ovis aries*).

A total of 113 whole genomes and 127 mitogenomes have illuminated the past population histories of European cattle and Eurasian sheep. Migration and admixture events have been identified in both species. However, the nature of these events are different in the two species. Domestic cattle admixed with the local wild progenitor (*Bos primigenius*), whilst the domestic sheep likely admixed with an improved domestic sheep translocated by humans. With improvements in sequencing technologies and ancient DNA recovery, the contribution of ancient genomes to our understanding of the process of domestication has only just begun.

7.4. References

Alexander, D.H., Novembre, J. & Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome research, 19(9), pp.1655–1664.

Allentoft, M.E. et al., 2015. Population genomics of Bronze Age Eurasia. Nature, 522(7555), pp.167–172.

Bradley, D.G. et al., 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), pp.5131–5135.

Cassidy, L.M. et al., 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), pp.368–373.

Chessa, B. et al., 2009. Revealing the history of sheep domestication using retrovirus integrations. Science, 324(5926), pp.532–536.

Cunliffe, B., 2013. Britain Begins, OUP Oxford.

Edwards, C.J. et al., 2007. Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. Proceedings of the Royal Society of London B: Biological Sciences, 274(1616), pp.1377–1385.

Frantz, L.A.F. et al., 2016. Genomic and archaeological evidence suggest a dual origin of domestic dogs. Science, 352(6290), pp.1228–1231.

Gamba, C. et al., 2014. Genome flux and stasis in a five millennium transect of European prehistory. Nature communications, 5, p.5257.

Gilbert, E. et al., 2017. The Irish DNA Atlas: Revealing Fine-Scale Population Structure and History within Ireland. Scientific reports, 7(1), p.17199.

Helmer, D. et al., 2005. Identifying early domestic cattle from Pre-Pottery Neolithic sites on the Middle Euphrates using sexual dimorphism. In J. D. Vigne, J. Peter, & D. Helmer, eds. The First Steps of Animal Domestication. Oxbow Books, Oxford, pp. 86–95.

Ho, S.Y.W. et al., 2008. Correlating Bayesian date estimates with climatic events and domestication using a bovine case study. Biology letters, 4(4), pp.370–374.

Kijas, J.W. et al., 2012. Genome-wide analysis of the world’s sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS biology, 10(2), p.e1001258.

Korneliussen, T.S., Albrechtsen, A. & Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC bioinformatics, 15, p.356.

Lawson, D.J. et al., 2012. Inference of population structure using dense haplotype data. PLoS genetics, 8(1), p.e1002453.

Librado, P. et al., 2017. Ancient genomic changes associated with domestication of the horse. Science, 356(6336), pp.442–445.

Loftus, R.T. et al., 1994. Mitochondrial genetic variation in European, African and Indian cattle populations. Animal genetics, 25(4), pp.265–271.

Loog, L. et al., 2017. Inferring allele frequency trajectories from ancient DNA indicates that selection on a chicken gene coincided with changes in medieval husbandry practices. Molecular biology and evolution. Available at: http://dx.doi.org/10.1093/molbev/msx142.

Martiniano, R. et al., 2016. Genomic signals of migration and continuity in Britain before the Anglo-Saxons. Nature communications, 7, p.10326.

Massilani, D. et al., 2016. Past climate changes, population dynamics and the origin of Bison in Europe. BMC biology, 14(1), p.93.

Meyer, M. et al., 2012. A high-coverage genome sequence from an archaic Denisovan individual. Science, 338(6104), pp.222–226.

Olalde, I. et al., 2014. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature. Available at: https://www.nature.com/nature/journal/v507/n7491/abs/nature12960.html.

Olalde, I. et al., 2017. The Beaker phenomenon and the genomic transformation of northwest Europe. bioRxiv, p.28. Available at: http://repository.nms.ac.uk/1864/ [Accessed October 17, 2017].

Park, S. et al., 2015. Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and evolution of cattle. Genome Biology, 16(1), p.234. Available at: http://genomebiology.com/2015/16/1/234.

Rasmussen, M. et al., 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature, 463(7282), pp.757–762.

Ryder, M.L., 1981. A survey of European primitive breeds of sheep. Annales de génétique et de sélection, 13(4), pp.381–418.

Scheu, A. et al., 2015. The genetic prehistory of domesticated cattle from their origin to the spread across Europe. BMC genetics, 16, p.54.

Schroeder, O. et al., 2017. Endogenous Retroviral Insertions Indicate a Secondary Introduction of Domestic Sheep Lineages to the Caucasus and Central Asia between the Bronze and Iron Age. Genes, 8(6). Available at: http://dx.doi.org/10.3390/genes8060165.

Schubert, M. et al., 2014. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. Proceedings of the National Academy of Sciences of the United States of America. Available at: http://dx.doi.org/10.1073/pnas.1416991111.

Skoglund, P. et al., 2015. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. Current biology: CB, 25(11), pp.1515–1519.

Troy, C.S. et al., 2001. Genetic evidence for Near-Eastern origins of European cattle. Nature, 410(6832), pp.1088–1091.

THE END

"Above all, don't fear difficult moments. The best comes from them."
Rita Levi-Montalcini