Faunal response to abrupt climate change: the history of the British mammal fauna from the Lateglacial to the Holocene

A thesis submitted to
Royal Holloway, University of London
for the degree of
DOCTOR OF PHILOSOPHY

Melissa Margaret Marr

Department of Geography
December 2016

Declaration of Authorship

I, Melissa Margaret Marr, hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Signed:

Date: 31/01/2017

Abstract

Rapid changes in climate are known to be drivers of profound ecosystem change and adaptive evolution. Repetitive and abrupt switches between glacial and interglacial conditions are strongly associated with range shifts, isolation in refugia, extinctions, local extirpations, re-colonisation events, demographic oscillations and repeated bouts of secondary contact in both flora and fauna. This study utilises a synthesis of ancient DNA and both 2D and 3D geometric morphometrics to examine phylogenetic, ecomorphological and population-level responses to abrupt climate change in British mammals. The investigation focuses on a period of significant climatic variability; the closing stages of the last (Devensian) glaciation into the current Holocene interglacial. Four broad alternating cold and warm episodes are recognized in Britain – the end of the Dimlington Stadial, the Lateglacial Interstadial, the Younger Dryas Stadial and the Holocene Interglacial. Three species with differing ecologies and life history traits were selected in order to look for common trends or individualistic responses: common vole (*Microtus arvalis*), Eurasian beaver (*Castor fiber*) and the European wildcat (*Felis silvestris silvestris*).

Geometric morphometrics was shown to be a highly effective tool, both with which to discriminate between morphologically similar species using isolated molar teeth, and with which to identify subtle form changes in tooth morphology that can be related to climate. Whole mitogenomes and single coding and non-coding mtDNA sequences were successfully obtained, in all cases representing the oldest DNA yet sequenced from these species. A common geographical origin was identified for all species as the north-west coast of Europe. Evidence for population continuity over the Younger Dryas cold interval was uncovered for *M. arvalis* and *C. fiber* and was associated with a possible micro-refugial area in the south-west of England. Levels of genetic diversity and the degree of phylogenetic and population substructure could be tentatively attributed to the individual dispersal capabilities and ecological preferences of different species.

Overall, this study uncovered formerly unknown population histories from ancient British mammal species and highlighted the huge potential of ancient DNA and geometric morphometrics for unravelling the Late and Postglacial history of the British mammal fauna.

Acknowledgements

I would like to extend my sincere gratitude to my supervisors, Prof. Danielle Schreve, Prof. Ian Barnes and Prof. Norman MacLeod. To have had access to their considerable expertise in their respective fields has been a very privileged experience for me. I would like to express my deepest appreciation for being allowed the freedom to work and learn independently while having their fullest support and guidance. I take inspiration from the passion, dedication and professionalism they demonstrate towards their work and I hope to take these qualities forward as I embark on my own research career.

Heartfelt thanks are extended to towards Dr. Selina Brace who has served as both a mentor and friend. Her encouragement and reassurance kept my confidence and spirits high during trying times in the project.

This study would not have been possible without the support and expertise of the many curators and researchers from across the UK that allowed me access to their collections. I am indebted to the backing they have offered and hope that I have managed to justify their trust in me. In no particular order these are: Linda Wilson of the University of Bristol Speleological Society; Andrew Kitchener of the National Museum of Scotland; Philippa Brewer, Spyridoula Pappa and Roberto P. Miguez of the Natural History Museum London; Danielle Schreve of Royal Holloway, University of London; Dennis Parsons of Somerset County Museum; Matthew Riley of the Sedgwick Museum Cambridge; Matthew Lowe of the University Museum of Zoology Cambridge; Tom Lord (private collection); David Rice of Gloucester Museum; Nicky Milner and Becky Knight of the University of York; Roy Stephenson of the Museum of London and David Gelsthorpe of the University of Manchester Museum.

Funding for this project was primarily provided by a National Environment Research Council (NERC) studentship. I am very grateful for the additional investment

provided by NERC for radiocarbon dating and by the Quaternary Research Association (QRA) for travel.

I've been honored to have been able to spend this time with work colleagues who I am also lucky enough to call my friends; Leila D'Souza, Roseina Woods and Tom Booth. Thanks for sharing this experience with me, for all the laughs and the meandering conversations.

To my mum, Marlene Marr, who ignited in me the first spark of curiosity about the natural world and to my brother, William Dale; you have both offered unwavering love, support and belief in me. I wouldn't be here without you.

Finally, to Toby, my malodorous and disloyal companion throughout all these years. Thank you for the distractions, I needed them the most.

CONTENTS

LIST OF FIGURES

Chapter 1.

Chapter 2.

Chapter 3.

Chapter 4.

Chapter 5.

Chapter 6.

LIST OF TABLES

Chapter 1.

Chapter 2.

Chapter 3.

Chapter 4.

Chapter 5.

Chapter 6.

LIST OF EQUATIONS

Chapter 2.

Chapter 3.

Chapter 5.

LIST OF ABBREVIATIONS

aDNA: Ancient DNA

AMS: Accelerator Mass Spectrometry

BIIS: British-Irish Ice Sheet

bp: Base pairs

cal BP: Calibrated years before AD 2000

CR: Control Region

CS: Centroid size

CVA: Canonical Variates Analysis

cyt *b*: Cytochrome *b*

d.f: Degrees of freedom

GICC05: Greenland Ice Core Chronology 2005

GLS: Generalised Least Squares

ka: Thousand years ago

LGM: Last Glacial Maximum (*c.*26,500 – 19,000 cal BP)

logCS: Log Centroid Size

LP: Late Pleistocene (*c.*110, 000 – 11,700 cal BP)

Ma: Million years ago

MAZ: Mammal assemblage zone

MIS: Marine oxygen isotope stages

mtDNA: Mitochondrial DNA

NGRIP/NorthGRIP: North Greenland Ice Core Project

NGS: Next Generation Sequencing

NUMT: Nuclear-inserted mitochondrial DNA (pseudogene)

PCA: Principal Components Analysis

PCR: Polymerase Chain Reaction

x: x-fold coverage. Calculated as: $\frac{read\ count\ x\ average\ read\ length}{total\ genome\ size}$

LIST OF APPENDICES

Figures

Tables

Figures

Figures

Chapter 1: General Introduction

1.1 Introduction

1.1.1 Overview

Many interacting factors play a role in determining the distribution of species in space and time, such as the level of adaptive potential, life history traits, biotic and abiotic interactions and the prevailing environment (Pearson & Dawson 2003). Within the suite of variables that together comprise a species' environment, climate has been identified as constituting a dominant force in determining biotic distributions (Wisz *et al.* 2013). The 'climatic envelope' of a species, defined as the full range of climatic variables that it can tolerate, can provide an accurate proxy for its fundamental niche in circumstances where it is the *sole* limiting factor in a species' existence (*i.e.* in the absence of intraspecific competition and resource constraints; Soberón 2007; Polly & Eronen 2011). For example, the intensity, rapidity and scale of climatic fluctuations over the Pleistocene epoch (*c.* 2.6 – 0.0117 Ma) have been observed to have exerted extreme selection pressures on taxa. Repetitive and abrupt switches between glacial and interglacial conditions triggered range shifts, isolation in refugia, extinctions, local extirpations, recolonisation events, demographic oscillations and repeated bouts of secondary contact in both flora and fauna (Hewitt 1999, 2000, 2004; Avise 2000).

These rapid climatic fluctuations have played a major role in shaping contemporary patterns of mammalian species assemblages, phylogeography and population structure. The Last Glacial Maximum (LGM: 26,500 – 19,000 cal BP) saw widespread global glaciations associated with low sea levels, which created land-bridges between regions now separated by large water bodies (Clark *et al.* 2012). In Eurasia, the mid to high latitude regions were largely glaciated and south of the ice sheets, the climate and environment were arctic in nature with permafrost reaching just shy of the multiple Mediterranean peninsulae (Svendsen *et al.* 2004). The classic refugia concept states that during the last ice sheet advances, temperate-adapted mammal species were forced to retreat to southern glacial refugia, where they became isolated until the climate ameliorated and they could re-colonise mainland Europe (Hewitt 1999, 2004).

This standard model of refugium theory has been supported by a large number of studies on European mammals that show a characteristic genetic signature: broad geographical subdivision into lineages derived from the Iberian Peninsula (western), Italian Peninsula and/or Balkan Peninsula (Southern) and the Caucasus (Eastern; Hewitt 2000). However, recent research has revealed a greater level of complexity than previously hypothesised in the classic refugium expansion-contraction paradigm, with many species showing individualistic responses to Quaternary climate fluctuations and potential cryptic northern refugia detected for some temperate species (*e.g.* common vole *Microtus arvalis;* Stewart & Lister 2001; Hofreiter & Stewart 2009; Stewart *et al.* 2010).

The Late Pleistocene (LP; *c.* 126,000 – 11,700 cal BP) is associated with the global extinction of an estimated 121 genera of terrestrial mammals, with a striking bias towards megafauna over 45 kg (Barnosky *et al.* 2004; Cooper *et al.* 2015). The causal factors of this remarkable extinction event have been a matter of debate for some time, with both climatic and anthropogenic influences being cited as the main drivers (Koch & Barnosky 2006; Sandom *et al.* 2014; Cooper *et al.* 2015). The LP saw a notable increase in the global population size and distribution of modern humans. Evidence of widespread human-mediated habitat modification, habitat fragmentation, 'over-kill' hunting and the possible introduction of novel pathogens to new environments has led many authors to conclude that anthropogenic activity was the principal factor driving megafaunal extinctions (Martin *et al.* 1967; Alroy 2001; Lyons *et al.* 2004; Burney *et al.* 2005; Sandom *et al.* 2014). However, evidence of fluctuations in micromammal populations, and a strong association between warming events and regional extinction/replacement of major Holarctic megafaunal clades, indicates that climate played a key role in precipitating LP ecosystem instability (Blois *et al.* 2010; Brace *et al.* 2012; Cooper *et al.* 2015; Brace *et al.* 2016).

The Pleistocene – Holocene transition is notable for a series of particularly rapid climatic fluctuations that are clearly reflected in the Greenland Ice Core Chronology (GICC05; Svensson *et al.* 2005) and which precipitated huge environmental change. External orbital forces triggered a period of global warming beginning *c.* 21,000 cal BP (Peltier and Fairbanks 2006), which ultimately resulted in global deglaciation and

a rise in sea levels. By around 14,500 cal BP, a rapid climatic amelioration marked the onset of the Lateglacial Interstadial. An abrupt climatic deterioration *c.* 12,900 cal BP marked the end of this warm phase and the commencement of the Younger Dryas, a cold phase of full glacial re-advance that persisted until rapid warming commenced at the onset of the Holocene Interglacial *c.* 11,700 cal BP. These abrupt millennial-scale climatic oscillations would have had profound effects on biota. The rapid succession of warming and cooling events are reflected by clear changes in vegetation and both invertebrate and vertebrate communities (Hewitt 2004; Hofreiter & Stewart 2009).

Relatively few studies have explored the fine-scale regional effects of Terminal Pleistocene - Holocene climate change on mammal population structure and phylogeography. Blois *et al.* (2010) uncovered patterns of range shifts, species diversity loss, and changes in abundance in North American small mammal assemblages and Brace *et al.* (2016), using molecular methods, revealed patterns of temporal population replacement in water voles. Britain is well placed to examine questions of timing, pattern and process of post-ice age recolonisation in mammal species due to *i*) its sensitivity to climate on account of its proximity to the North Atlantic; *ii*) its exceptional fossil assemblages and, *iii*) its geography, formerly connected to the European mainland via a large contiguous landbridge, Doggerland, which persisted until *c.* 8, 200 cal BP (Weninger *et al.* 2008). Accordingly, this study aims to investigate the effects of abrupt climate change through the British Lateglacial to Early Holocene on mammal species using a novel synthesis of ancient DNA, geometric morphometrics and ^{14}C radiocarbon dating.

1.1.2 Chronology and subdivision of the Pleistocene and Holocene epochs

The GICC05, created from annual counting of layers from high-resolution Greenland ice cores, is utilised here to chronicle climatic events in the Late Pleistocene/Early Holocene (Svensson *et al.* 2008). This chronology is based on the counting of annual layers in $\delta^{18}O$ and δD from the DYE-3, GRIP and NorthGRIP ice cores (Vinther *et al.* 2006). All chronological dates given in the text are given in calibrated radiocarbon years (cal BP) and where conversions have had to be made from uncalibrated

radiocarbon years in the literature, these have been calculated using the OxCal program performed online at https://c14.arch.ox.ac.uk/oxcal. Reference to subdivisions of the Pleistocene and Holocene epochs are based on definitions from Gibbard *et al.* (2010) and Walker *et. al.* (2012) respectively. However, it is important to note that other indicators of palaeo-temperature change exist, such as the GRIP deuterium-excess record (Jouzel *et al.* 2007) but that due to inconsistencies and/or correlation controversies between these two proxies, only the profile based on $\delta^{18}O$ is adopted here. In the case of broad divisions of the Holocene into 'Early', 'Middle' and 'Late' these are defined as i) Early - *c.* 11.5kya to 8.2kya ii) Middle – *c.* 8.2kya to 4.2kya and iii) Late - *c.* 4.2kya to present (Walker *et al.* 2012).

1.1.3 The Quaternary Period - palaeoclimate and oxygen isotope dating

The Quaternary Period began *c.* 2.6 Ma at the Gauss/Matuyama palaeomagnetic chron boundary of the Gelasian GSSP (Global Boundary Stratotype Section and Point; Gibbard *et al.* 2005) and is associated with complex cyclical patterns of glacial advance and retreat. These alternating periods of global warming and cooling are thought to be initiated by external orbital forcing where changes in the nature of the Earth's rotation on its axis and its orbit around the Sun cause a redistribution of insolation across the southern and northern hemispheres (Dawson 1992). Milankovitch (1941) theorised that changes in the eccentricity, obliquity and precession of the Earth's orbit (and their interactions) would create variations in the amount of solar radiation reaching the earth and thus be responsible for major climatic events (Hays *et al.* 1976). These external orbiting forces are augmented by internal influences such as the concentration of atmospheric greenhouse gases and the extent of ice coverage (Wolff *et al.* 2010). The frequency and amplitude of global fluctuations in the Earth's climate increased dramatically during the Quaternary (Fig 1.1), with around 50 glacial-interglacial cycles observed during this period.

The early part of this period was characterised by obliquity-driven 41ka cycles until *c.* 1Ma when there was a transition to 100ka cycles. Termed the Mid-Pleistocene

Revolution (MPR), this had the effect of amplifying the range of climatic oscillations (Fig 1.1) and these are hypothesised to be eccentricity-driven, although this remains controversial (Maslin and Ridgwell 2005). Climatic instability over this period caused dramatic changes in environment that exerted major ecological and evolutionary pressures on biota (Bennett 1990). Many cold-adapted species have their origins in the Quaternary and particular faunal communities are strongly associated with specific climatic and ecosystem types. For example, the cold-adapted large mammal faunas ('mammoth fauna'; Vereschagin and Baryshnikov 1982 or '*Mammuthus-Coelodonta* faunal complex'; Kahlke 1999), are closely related with the steppe-tundra or 'Mammoth Steppe' (Guthrie 1982), This was a herbaceous ecosystem that was characteristic of Pleistocene glacial periods and which does not exist today. The Quaternary Period is therefore an excellent model within which faunal responses to abrupt climate change can be examined.

Evidence for Quaternary climate change comes largely from analysis of oxygen isotope measurements of the calcareous skeletal remains of deep sea foraminifera extracted from marine sediment cores and from Greenland and Antarctic ice cores. Found within sedimentary layers, foraminifera shells provide a mode of determining former deep water environments and, by proxy, a means of reconstructing the extent of ice masses and past climatic conditions (Stott *et al.* 2007). Two factors affect the stable oxygen isotope concentration of foraminifera shells *i*) sea water temperature and *ii*) the isotopic concentration of marine water throughout the lifetime of the micro-organisms (Dawson 1992). During the process of evaporation, the oxygen isotopes ^{16}O, ^{17}O and ^{18}O are released into the atmosphere, with a preferential evaporation of the lighter isotopes (fractionation). During glacial periods, the lighter ^{16}O becomes trapped within terrestrial ice sheets, resulting in an oceanic enrichment of ^{18}O. The calcium carbonate shells reflect the oxygen isotopic content of the surrounding water, thus analysis of the skeletal ^{18}O provides an indirect indicator of past global climate and ice sheet volume (Waelbroeck *et al.* 2002). Analyses of ice cores can provide a more direct measure. The preservation of precipitated oxygen isotopes within glaciers provide a record of the atmospheric composition from the period they became incorporated into the ice, while chronology can be established

from the counting of individual ice layers and from the geochronological dating of incorporated tephra layers. The study of $^{16}O/^{18}O$ ratios in ice cores using the $\delta^{18}O$ profile (based on relative deviations from the mean ratios of a standard value; Shackleton and Opdyke 1973) has provided a widely adopted proxy for dating climatic events during the Quaternary Period with data being gathered mainly from Antarctica (Lüthi *et al.* 2008) and Greenland (Dansgaard *et al.* 1993).

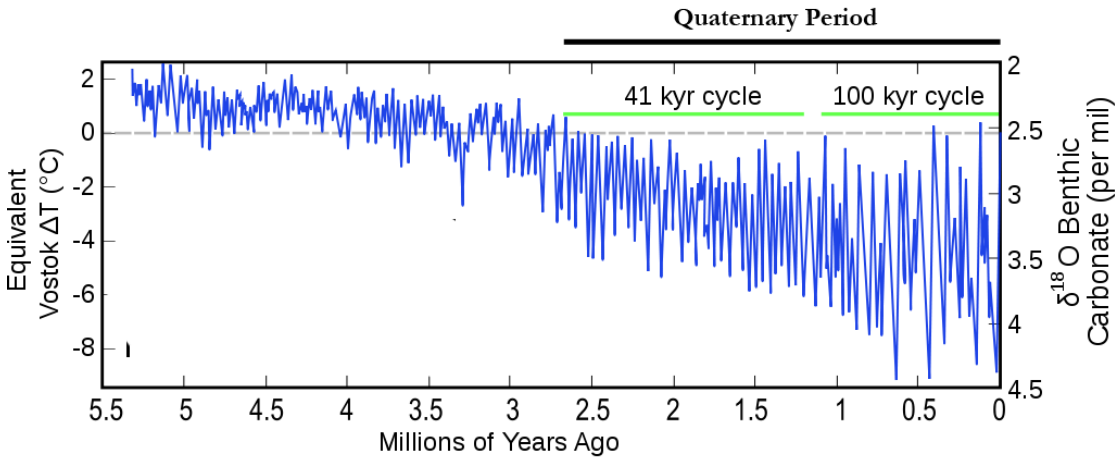


Figure 1.1. Climatic fluctuations over the past 5.5 million years. An increase in amplitude is shown from the beginning of the Quaternary Period. Temperature changes are represented through the $\delta^{18}O$ profile. Image redrawn from Lisiecki & Raymo (2005); data downloaded from http://www.ncdc.noaa.gov/

1.1.4 The Late Pleistocene and Early Holocene epochs - marine Oxygen Isotope Stages (MIS) and climatic fluctuations

The Quaternary period is typically subdivided into two epochs (or series; Fig 1.2) – the Pleistocene (*ca.* 2.6 - 0.0117 Ma) and the Holocene (*ca.* 0.0117 – 0Ma). Oxygen isotope signals have been utilised to produce a timescale that chronicles the alternating warm and cold periods throughout Quaternary palaeoclimatic history. These are represented by a succession of numbered marine oxygen isotope stages (MIS) in the deep ocean record and paralleled on land by long terrestrial records such as the ice core data from Antarctica and Greenland. Glacials are considered to be cold periods spanning many millennia in which temperatures in the mid and high latitudes are sufficiently low as to cause widespread glaciation. Further subdivision can be made into stadials - cool periods within interglacials, typically ≤ 10 000 years in duration and within which some glacial re-advance may be observed, and interstadials – warm

periods within glacials, typically lasting ≤ 5000 years and where temperatures are warm but not as high as the present day (Lowe and Walker 1997).

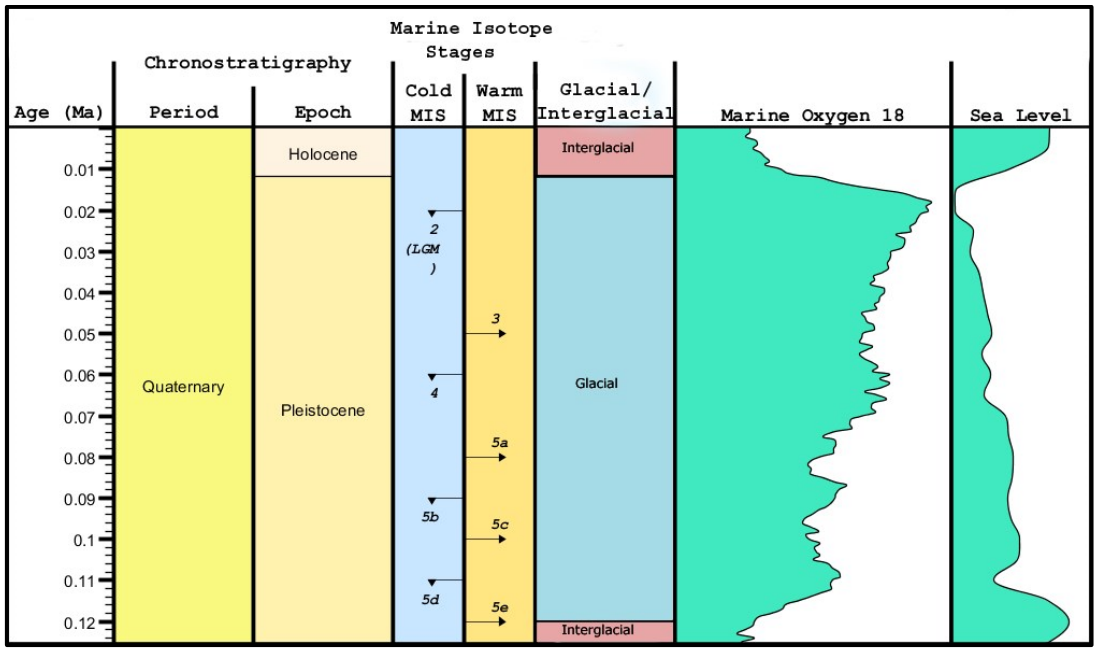


Figure 1.2. The Late Quaternary Period (0.14 Ma to present). Showing chrono-stratigraphic framework including Marine Isotope Stages (MIS), $\delta^{18}O$ profile, sea level and date of Last Glacial Maximum (LMG). Created using TimeScale Creator, Lugowski & Ogg (2009)

The study period that is the focus here encompasses the Lateglacial and Early-Postglacial in Britain, which span the closing phases of the LP and into the Early Holocene (*c.* 15,000 cal BP onwards), covering the closing stages of MIS 2 (Late Pleistocene) and the beginning of MIS 1 (Early Holocene). MIS 2 began *c.* 24,000 cal BP, close to the global Last Glacial Maximum (LGM) and is characterised by low sea levels, enrichment of ^{18}O in sea water and widespread glaciation (Fig 1.2). MIS 1, which began *c.* 14,500 cal BP and spans the current Holocene epoch, is characterised by a rise both in sea levels and global temperature and the retreat of the Late Pleistocene ice sheets. The fluctuating climate of this time is clearly reflected in oxygen isotope records from GICC05 (Svensson *et al.* 2008), which show a series of clearly-defined climatic events. The global LGM of MIS 2 is traditionally defined as having occurred *c.* 23,000 - 19,000 cal BP (EPILOG LGM chronozone level 1; Mix *et al.* 2001). However, more recent evidence has revised and extended this period so that the most accurate chronology of the LGM is now *c.* 26,000 – 21,000 cal BP

(Peltier and Fairbanks 2006), within which range most of the major global ice sheets experienced their maximum glacial limits. Around 21,000 cal BP, global temperatures started to increase, initiating a reduction in global ice volume and an associated rise in sea levels (Peltier and Fairbanks 2006). Close to the beginning of the study period, *c.* 14,700 cal BP, an episode of rapid warming occurred, which lasted for around 2000 years (Fig 1.3 and 1.4). This is the Lateglacial Interstadial, equated with Greenland Interstadial (GI) I, a complex period divided into a series of warm and cold sub-events (GI-1e-a) showing an alternating, low magnitude, pattern of warming and cooling (Fig 1.3; Björk *et al.* 1998; Walker *et al.* 1999; Svensson *et al.* 2008). Within the broad Lateglacial Interstadial, two temperate-climate episodes are indicated, the Bølling Interstadial (equated with GI-1e) and the Allerød Interstadial (correlated with GI-1abc). The interstadial was followed by a period of abrupt cooling, *c.* 12,900 – 11,600 cal BP, the Younger Dryas Stadial, corresponding to Greenland Stadial 1 (GS 1). Immediately following this, rapid warming commenced *c.* 11, 700 cal BP marking the onset of the Holocene interglacial.

1.1.5 The British-Irish Ice Sheet (BIIS) – extent and de-glaciation

During the global LGM (26, 000 – 21, 000 cal BP), the widespread extent of glaciation caused a eustatic sea level drop of *c.*130m (Clark *et al.* 2012), resulting in the exposure of land previously inundated by sea water and the (re)creation of land-bridges between formerly-isolated regions. The relatively small British-Irish Ice Sheet (BIIS) reached its maximum limits (in terms of total ice volume) slightly in advance of the global LGM at *c.* 27,000 cal BP, where it had a maximum reconstructed extent of *c.* 840,000km^3 of which around 300,000km^3 was situated on the British and Irish continental shelf and over the North Sea (Clark *et al.* 2012). Different regions of the ice sheet reached their maximum extent at different times (Figure 1.5). The southern-most maximum extent covered most of Wales (*c.* 23,000 cal BP), Derbyshire and Yorkshire (*c.* 23,000 – 21,000 cal BP) with a strip of ice extending down the North Sea coast and covering east Norfolk (*c.* 17,000 – 19,000 cal BP; Phillips *et al.* 1994; Clark *et al.* 2012). At the BIIS glacial maximum, sea levels were sufficiently low as to

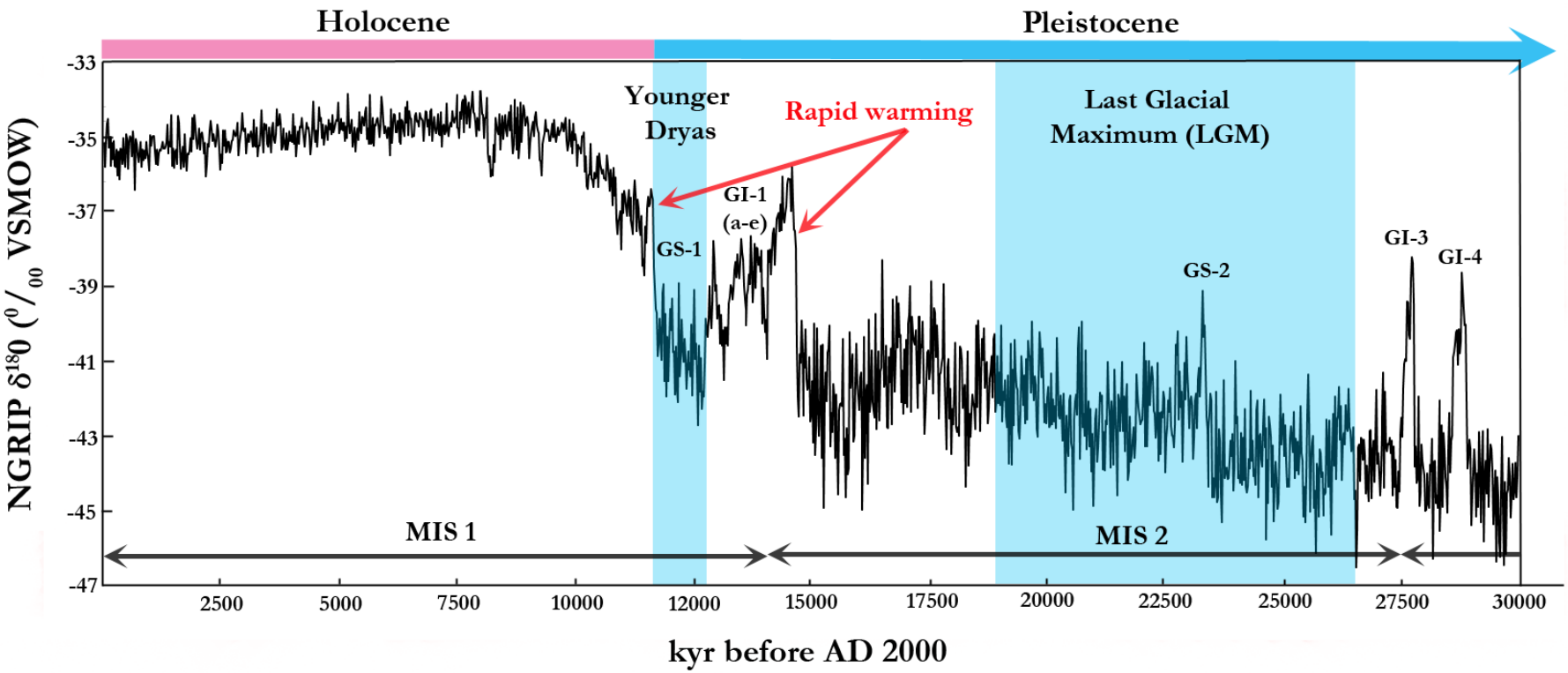


Figure 1.3. Chronology of climatic oscillations during the last 30 kya of the Late Quaternary period. Marine Oxygen isotope data extracted from the North Greenland Ice Core Project (NorthGRIP) using the Greenland Ice Core Chronology 2005 time scale (Svensson *et al.* 2006, 2008). MIS = Marine Isotope Stage. GI = Greenland Interstadial; GS – Greenland Stadial. Cold phases - the Last Glacial Maximum (LGM) and the Younger Dryas - highlighted in blue.

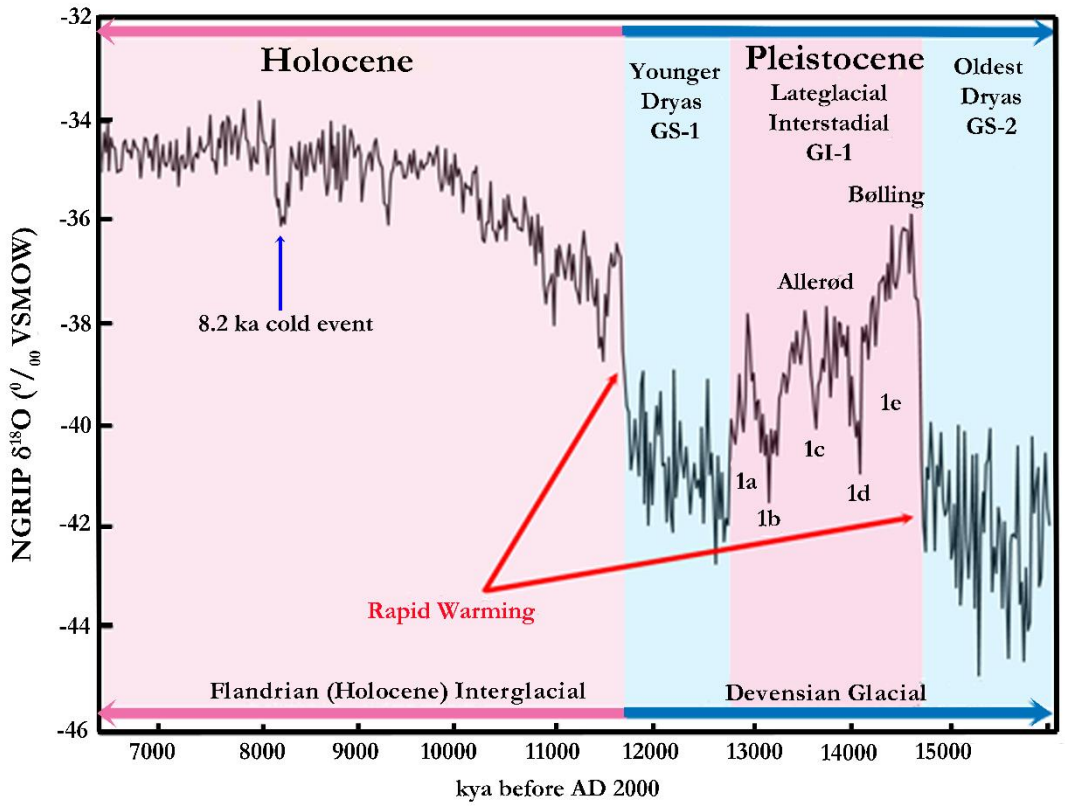


Figure 1.4. Chronology of climatic oscillations during the study period. Marine Oxygen Isotope data extracted from the North Greenland Ice Core Project (NGRIP) using the Greenland Ice Core Chronology 2005 time scale (Svensson *et al.* 2006, 2008). GI = Greenland Interstadial. GS = Greenland Stadial. Broad divisions into cold and warm periods are highlighted blue and pink respectively. The four main climatic periods of interest - the end of the Oldest Dryas, the Lateglacial Interstadial, the Younger Dryas Stadial and the beginning of the Holocene Interglacial are all shown. Note that the Lateglacial is not a period of uniformly warm temperatures. The Bølling (GI-1e) is separated from the Allerød (GI-1c) by the Oldest Dryas brief cold phase (GI-1d). The Allerød is also followed by another brief cooling phase (GI-1b). Also highlighted is the 8.2 ka Holocene cold event.

expose vast areas of the sea bed in the North Sea and English Channel, reconnecting Britain as a peninsula of north-west Europe. At around 20,000 – 19,000 cal BP, increasing levels of solar insolation in the Northern hemisphere triggered global deglaciation (Clark *et al.* 2012). As a small ice sheet, the BIIS was highly dynamic and its position in NW Europe (adjacent to the North Atlantic Gulf Stream where large climatic oscillations are known to have occurred) meant that it was highly vulnerable to changes in climate and melting of other, larger, ice sheets. Therefore, all marine sectors of the BIIS were fully collapsed by *c.* 17,000 cal BP and most terrestrial regions (bar some areas of the Scottish Highlands and Islands) were free of ice by *c.* 15,000 cal BP, just prior to the abrupt warming of the Lateglacial Interstadial (Figure 1.6; Ballantyne 2010; Clark *et al.* 2012). Sea levels remained sufficiently low as to allow mainland Britain to retain a terrestrial connection to NW Europe well into the

Holocene, with Britain only regaining its island status *c.* 8, 200 cal BP when the land-bridge was fully drowned (Weninger *et al.* 2008).

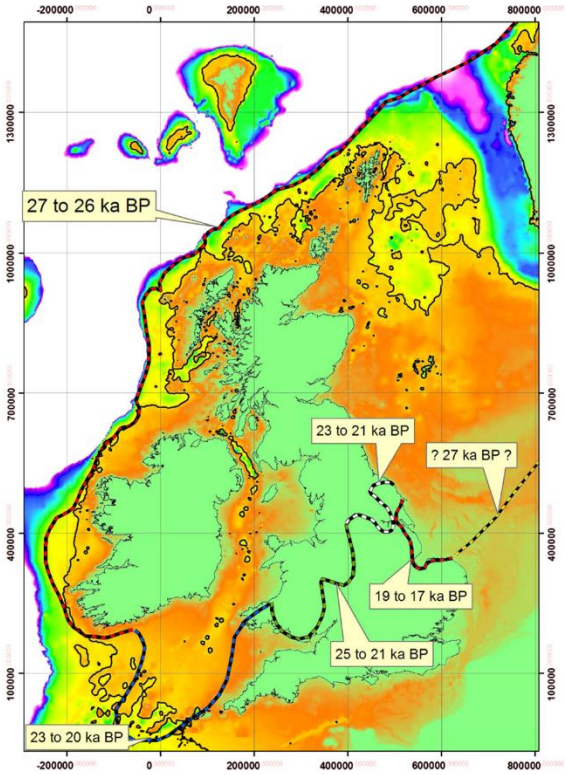


Figure 1.5. Britain and Ireland showing the maximum limits of the British-Irish Ice Sheet (BIIS) and approximate dates on when each section reached its maximum extent. Colours indicate shelf bathymetry from blue (-300m) to brown (0m; modified from Clark *et al* 2012).

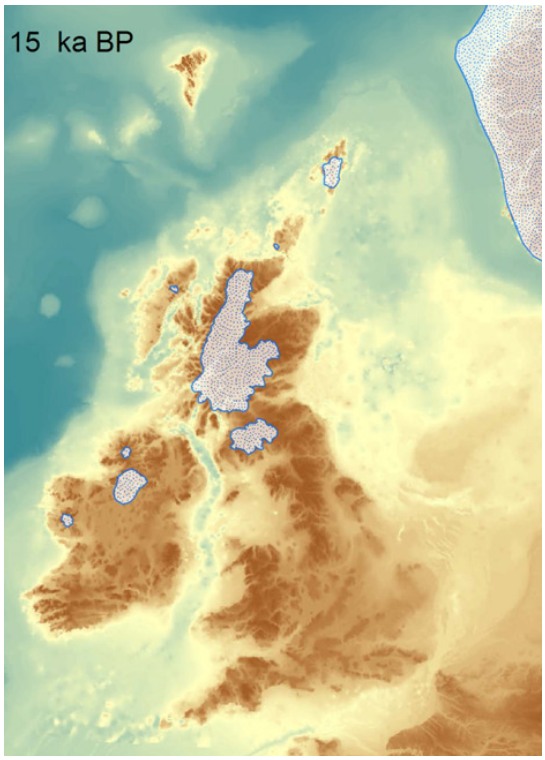


Figure 1.6. Extent of glaciation and coastline of Britain at the start of the study period. Showing glaciated areas restricted to NW and south Scotland (taken from Clark *et al* 2012).

1.1.6 The British-European landbridge – Doggerland

The contiguous area of land that connected Britain to NW Europe throughout the last glacial period and into the Holocene is termed Doggerland, after the Dogger Bank – an area of raised sand bank in the North Sea off the east coast of England. Doggerland would have represented a habitable area capable of supporting mammalian fauna in its own right but crucially, when considering the recent history of the British mammalian fauna, it would have functioned as a vital recolonisation route for species from mainland Europe (Coles 1998). The topography of Doggerland has proved difficult to establish due to the combined effects of erosion and siltation. However, geological surveys and commercial exploration of the North Sea bed have provided sufficient data to enable terrain mapping of the area, most interestingly via

3D seismic data. At its maximum extent in the late Devensian, Doggerland was a vast plain extending from the east coast of Britain to the present coasts of the Netherlands, Denmark and north Germany (Fig 1.7: Clark *et al.* 2012). Quaternary glaciations have carved out many of the geological features of Doggerland. Sub-glacial drainage channels were exposed following glaciation, as long, narrow and steep-sided tunnel valleys, which were occasionally filled with fresh water. Between NE Scotland and the Dogger Bank, some of these tunnel valleys were 1-3km wide, 100m deep and 25-60m long (Coles 1998). Three major river networks were present, two of which drained to the north and one of which drained to the south. As sea levels began to rise, *c.* 12 000 cal BP, the coastline of Doggerland began to shift in response, becoming longer as large areas of the land became submerged under sea water. Around 8, 200 cal BP, the Storegga Slide Tsunami is hypothesized to have caused a catastrophic flood that completely severed the remaining land links between Britain and mainland Europe (Weninger *et al.* 2008), ending mammalian immigration from the continent. The major geographical features of Doggerland would have both facilitated and acted as barriers to movement to mammals recolonizing Britain from mainland Europe, depending on the specific ecology and dispersal abilities of individual species. For example, major riverine systems may have acted as corridors and facilitated dispersal for semi-aquatic species such as the Eurasian beaver *Castor fiber* but would have proved a major biogeographical barrier to terrestrial small mammals.

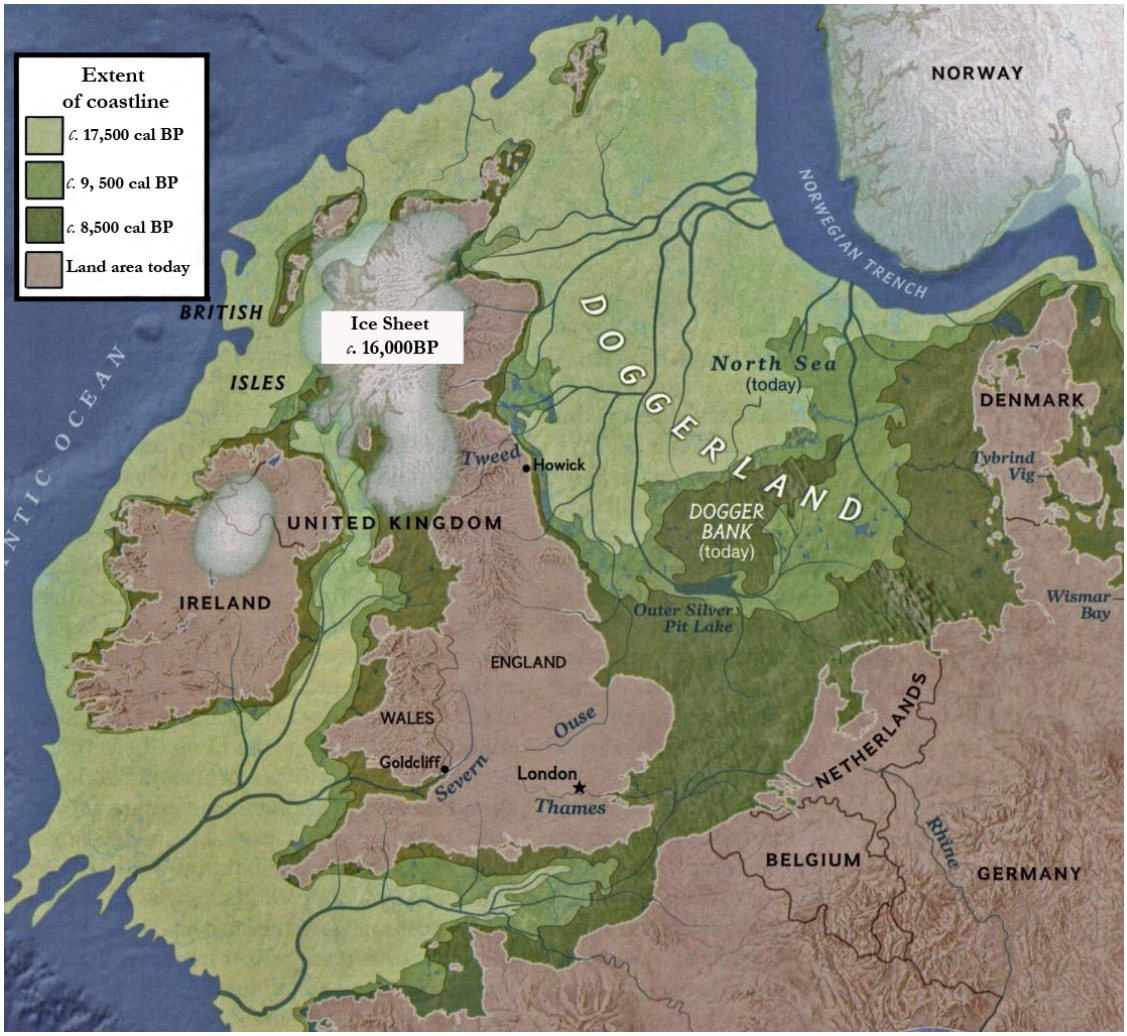


Figure 1.7. Approximate extent of the British coastline from the Lateglacial to the Holocene. Modified from McNulty and Cookson (2012).

1.1.7 Terminal Pleistocene and Holocene Britain – climate and environment

External orbital forces triggered a period of global warming beginning *c.* 21 000 cal BP (Peltier and Fairbanks 2006), which ultimately resulted in a global deglaciation and a rise in sea levels. At the start of the study period (*c.* 15,000 cal BP) deglaciation of the BIIS sheet was almost fully complete, with only some glaciated areas remaining in the NW of Scotland (Ballantyne 2010). At this time, the end of the Older Dryas Stadial, the climate in Britain was severely cold with an open steppe-tundra habitat (Lowe and Walker 1997). The subsequent periods of rapid and abrupt Lateglacial/Early Holocene climatic fluctuations are known to have caused considerable modification of vegetational communities with transitions from open steppe-tundra to woodland floras associated with climatic transitions from cold to

warm periods (Lowe and Walker 1997). Multiple lines of evidence provide a basis for interpreting patterns of vegetational succession and environmental change, with data gathered from pollen (Table 1.1), plant macrofossil, coleopteran, mollusc, vertebrate (for mammals see *e.g* Currant & Jacobi 2011), geological and isotopic sources (Walker *et al.* 1994; Lowe and Walker 1997). Lowe and Walker (1997) recognise four climatic episodes within the Lateglacial/Early Holocene period that can be associated with discrete vegetational communities: the end of the Oldest Dryas (or Dimlington Stadial) prior to *c.* 14,700 cal BP - a cold period, at the end of which the first signs of warming are apparent, the Lateglacial Interstadial *c.* 14,700- 12,900 cal BP - a warm phase, the Younger Dryas Stadial *c.* 12,900 – 11,600 cal BP – a brief cold phase, and the very early Holocene (Flandrian) Interglacial *c.* 11,600 cal BP – a period of climatic amelioration (Table 1.2).

Table 1.1. British pollen zone subdivisions of the Lateglacial/Early Postglacial

Years BP	cal years BP	Pollen Zone	Zone Name	Archaeological Period
8000	8853 (± 142)	VI	Boreal	Mesolithic
9000	10091 (± 142)	VI		
10000	11549 (± 205)	IV	Pre-Boreal	
11000	12923 (± 117)	III	Younger Dryas/Loch Lomond Stadial	Late Palaeolithic
12000	13989 (± 232)	III	Windermere Interstadial	
13000	15860 (± 428)	VI	Older Dryas	

Table 1.2. Vegetation and climate of Britain during the major climatic episodes recognised by Lowe and Walker (1997)

Vegetation Type	Climate	Years BP	cal BP	Chronostratgraphic Units
Woodland (some open)	Warming	9,500	10, 850 (± 189)	Flandrian Interglacial
Transition to woodland	Warming	10,000	11, 549 (± 205)	
Tundra and low alpine scrub	Arctic	10,500	12,406 (± 205)	Younger Dryas/Loch Lomond Stadial
		11,000	12,923 (± 117)	
Birch woodland	Milder	11,500	13,399 (± 140)	Lateglacial/Windermere Interstadial
Birch woodland (sporadic in north)	Cooling	12,000	13,989 (± 232)	
Open wood and scrub		12,500	14,806 (± 329)	
Open ground communities	Warm	13,000	15,860 (± 428)	
Open tundra and/or steppe	Arctic	pre-14,000	17,252 (± 227)	Oldest Dryas

At the end of the Oldest Dryas Stadial, just prior to the beginning of the Windermere Interstadial *c.* 14,700 cal BP, the dominant habitat type in Britain was steppe-tundra – a vegetation type consisting of herbaceous floral communities adapted to dry and cold climatic conditions with few or no woody species (Lowe and Walker 1997) - and a coleopteran fauna indicative of arctic conditions (Coope 1977). The commencement of the Lateglacial (Windermere) Interstadial was marked by a rapid climatic amelioration and the thermal maximum for this period was reached almost immediately (within GI1-e) with mean July temperatures reaching up to *c.* 20°C in southern areas (Atkinson *et al.* 1987; Walker *et al.* 2003). Maximum July temperatures (TMAX) estimated from coleopteran fossil assemblages in Gransmoor site (east Yorkshire) using the Mutual Climatic Range (MCR) method give a range of *c.* 12.5 – 25.5°C for the very early part of the Lateglacial (Elias & Matthews 2014). During the early (Bølling) phase of the Windermere Interstadial, the dominant vegetation communities were composed of grasses (Poaceae), sedges (Cyperaceae), and herbaceous plants such as docks and sorrels (*Rumex* spp.) and *Artemisia* spp., which thrive in open ground (Warne 1981). Succession to a vegetation type more typical of temperate climates then occurred, with a transition to open woodland/scrub environments composed of juniper (*Juniperus*) and willow (*Salix*). In central and southern England, open birch (*Betula*) and pine (*Pinus*) woodland is known to have developed but was initially largely absent in north-east England and eastern Scotland, suggesting that Britain had a mosaic of habitats at this time with regional variation in vegetational communities (Lowe and Walker 1997).

Two periods of cooling, which are reflected in the vegetation, can be detected within the Lateglacial Interstadial at *c.* 14,100 cal BP (GI1-d) and *c.* 13,100 cal BP (GI1-b). During these cool phases, temperatures fell from the thermal maximum by *c.* 5°C in the former and a further 4-5°C in the latter (Walker *et al.* 2003). A reduction in *Juniperus* scrub type vegetation is seen in the first of these cooling periods and a significant contraction of *Betula* woodland is seen in GI1-b. These cooling events are separated by the Allerød, a period of warming that again saw transition to woodland ecotypes (Birks & Birks 2014). Notably, the spread and development of closed canopy

woodland during the Allerød is associated with major declines and local extirpations of wild horse populations in the south of England (Kaagan 2000).

The Lateglacial Interstadial closed with a short warming period (GI1-a), which enabled the re-expansion of birch woodland (Lowe and Walker 1997). The climate deteriorated *c.* 12,900 cal BP initiating a brief cold phase, the Younger Dryas/Loch Lomond Stadial *c.* 12,900 – 11,600 cal BP. This abrupt transition to colder temperatures is reflected in the beetle TMAX MCR data with a maximum July temperature range of 11-13^{o}C in the terrestrial record and 12-14^{o}C in the aquatic record (Gransmoor, east Yorkshire; Elias & Matthews 2014). In the early stages of the Younger Dryas, southern Britain supported a *Betula* and *Populus*-dominated woodland while wooded habitat was present, but rare, in the north of country (Walker *et al.* 1994). At the coldest point of the Younger Dryas, mean January temperatures were as low as -20^{o}C and mean July temperatures were around 12^{o}C (Isarin & Bohncke 1999). Glaciation occurred in the west Scottish Highlands (Hubbard 1999) and at this point, arctic conditions had returned to Britain, with a landscape once again dominated by grasses and sedges associated with tundra and low-alpine scrub.

Climatic amelioration began during the latter part of the Younger Dryas with stadial conditions ending *c.* 11, 700 cal BP when Holocene warming began. Within five decades, temperatures had reached a mean July average of *c.* 17^{o}C (Atkinson *et al.* 1987) and the climatic maximum for Europe was reached *c.* 9000 cal BP and lasted until *c.* 6300 cal BP (Kalis *et al.* 2003) from where temperatures have gradually declined. The Lateglacial/Early Holocene transition saw the open ground/heathland vegetation communities replaced by more temperate-adapted deciduous woodland species with full climatic amelioration and re-afforestation completed by *c.* 10 500 - 8900 cal BP, although some species did not reach their northern limits until *c.* 6000 cal BP (Fyfe *et al.* 2013).

1.1.8 The effects of Pleistocene climate change on mammals

Species can respond in three ways to environmental change depending on the tempo and mode of modification, *i*) extirpation or extinction *ii*) migration (habitat tracking) and/or *iii*) adaptation (genetic, phenotypic and/or behavioural). The classic refugium concept states that (in Europe) during the last ice sheet advances, temperate mammal species were forced to retreat to the multiple Mediterranean peninsulae where they became isolated until the climate ameliorated and they could re-colonize mainland Europe (Fig 1.8 – 1.9; Hewitt 1999, 2004). During cold stages, cold-adapted species are known to have to have experienced south and southwestern range expansion out of their Northern 'refugia' into previously unsuitable habitat, and range contraction or extinctions during warm periods (Hewitt 1999, 2004; Řičánková *et al.* 2015). This standard model of refugium theory has been supported by a number of studies on temperate mammals that show a characteristic genetic signature: broad geographical subdivision into (usually mitochondrial DNA) clades derived from the Iberian Peninsula (western group), the Italian Peninsula and/or the Balkan Peninsula (southern group) and the Caucasus (eastern group; Hewitt 2000). However, recent research has revealed a greater level of complexity than previously hypothesised in the classic refugium expansion-contraction paradigm, with many species showing individualistic responses to Quaternary climate fluctuations and potential cryptic northern refugia detected for some temperate-adapted species (Stewart & Lister 2001; Hofreiter & Stewart 2009; Stewart *et al.* 2010).

The individualistic response of species to environmental change during the Quaternary period has been suggested as a major factor in the creation of disharmonious, or non-analogue, faunas (Williams *et al.* 2007). These are communities composed of combinations of species that today have allopatric distributions. Stratified mammal remains have long been used to reconstruct patterns of palaeoenvironmental change under the assumption that species had the same climatic envelope as they do today and that this determines their geographical distributions (Polly & Eronen 2010). Non-analogous faunas thus present a conflicting picture of the palaeoenvironment as they imply that environmental overlap occurred in the past,

allowing the co-existence of species that today do not share the same geographic range. However, an individual response by species to environmental change implies that the geographical definition of refugia will vary among species, that species will spend differing amounts of time within those refugia and respond in differing ways to environmental change (Stewart *et al.* 2010). Moreover, range expansions/contractions and extinction events are unlikely to be synchronous across species (Stewart 2008). It has even been suggested that individualistic responses may occur at the population level in situations where adaptive variation to local conditions exists over a species' range (Stewart 2008, 2009). This phenomenon and the existence

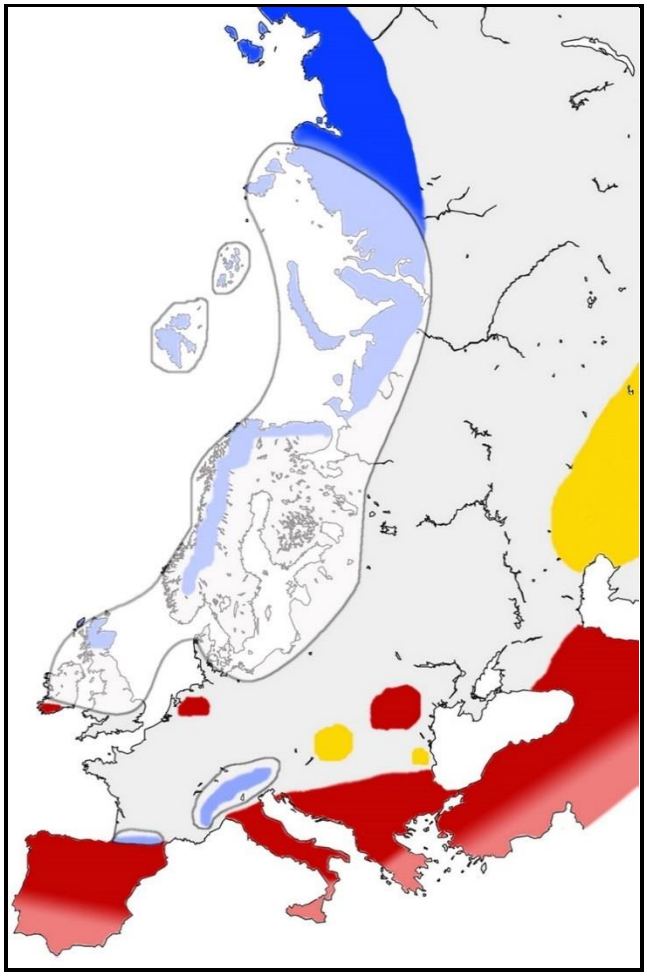


Figure 1.8. Glacial refugial regions for Europe and western Asia. Blue areas represent interglacial refugia for cold-adapted taxa; Red areas show glacial refugia for temperate taxa (including cryptic northern refugia); Yellow areas indicate interglacial refugia along the oceanic/continental gradient (including cryptic refugia in the west). Modified from Stewart *et al.* 2009.

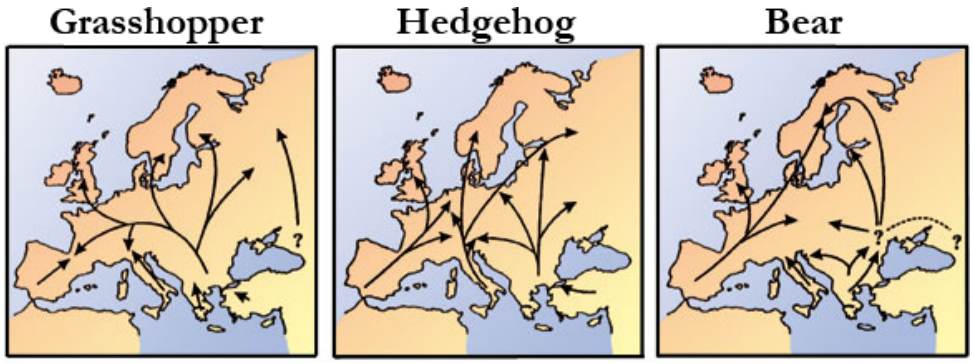


Figure 1.9. Differential contribution of Mediterranean refugia for different species under the traditional glacial refugium model for the Last Glacial Maximum. From Hewitt (1999).

of cryptic northern refugia could lead to the co-existence of taxa that are not analogous to current, Holocene, communities (Stewart & Lister 2001; Williams & Jackson 2007; Stewart 2008, 2009).

1.1.9 Late and Postglacial recolonisation of Britain

Non-analogue communities were typical, rather than uncharacteristic, of the British mammalian fauna during the Quaternary period (Stewart 2008, Polly & Eronen 2010), an observation that seems to be supported by ^{14}C dating of key co-existing taxa. However, it has to be taken into consideration that erroneous historical ^{14}C dates (prior to current ultrafiltration methods) may obscure true patterns. Faunal assemblages can also be misinterpreted due to a variety of other factors such as post-depositional stratigraphic mixing, misidentification and the habitation of the same site by seasonal migrants which used the site at different times of the year.

During the Lateglacial Interstadial, apparently mixed assemblages of cold, open habitat species such as *Rangifer tarandus* (reindeer), *Vulpes lagopus* (Arctic fox), *Mammuthus primigenius* (mammoth), and *Saiga tatarica* (saiga antelope) are seen in association with species more typical of warmer, woodland habitats such as *Cervus elaphus* (red deer), *Alces alces* (elk), *Lynx lynx* (lynx) and *Bos primigenius* (aurochs). While regional differences exist in the distribution of species across Britain in the Late Glacial/Early Holocene, Currant and Jacobi (2011) have suggested a typical faunal assemblage from Gough's Cave, Somerset (the Gough's Cave mammal assemblage

zone - MAZ), which serves as a type locality, broadly representing the Lateglacial from 13,000 cal BP (Table 1.3). No single site faunal assemblage from the Younger Dryas Stadial yet exists with most taxa of the period identified via ^{14}C dating of individual bones/teeth over a range of sites (Currant & Jacobi 2011). However, there is a clear transition to cold adapted taxa during the Younger Dryas and associated reduction in species richness. Mammal assemblages were dominated by open habitat species such as *E. ferus*, *R. tarandus*, Norway lemming (*Lemmus lemmus*) and collared lemming (*Dicrostonyx torquatus*). Also present are species associated with steppic environments such as *Ochotona pusilla* (steppe pika; Currant & Jacobi 2011). Temperate, woodland species re-colonize around the early Holocene, with *Bos primigenius*, *Castor fiber* and *Capreolus capreolus* (roe deer) all described from this period. The co-existence of temperate and cold adapted species in many Lateglacial - Early Holocene mammal assemblages implies a high level of complexity within mammalian communities at this time, raising questions regarding how these faunal elements came together (Yalden 1999; Current & Jacobi 2011; Polly 2011).

Table 1.3. Gough's Cave mammal assemblage zone (MAZ) fauna. From Currant and Jacobi (2011)

Genus	Species	Common name
Homo	*sapiens*	Modern human
Lepus	*timidus*	Arctic hare
Dicrostonyx	*torquatus*	Collared lemming
Lemmus	*lemmus*	Norway lemming
Arvicola	*terrestris*	Water vole
Microtus	*oeconomus*	Northern vole
Microtus	*gregalis*	Narrow-skulled vole
Canis	*lupus*	Wolf
Vulpes	*vulpes*	Red fox
Vulpes	*lagopus*	Arctix fox
Ursus	*arctos*	Brown bear
Lynx	*lynx*	Lynx
Mammuthus	*primigenius*	Mammoth
Equus	*ferus*	Horse
Cervus	*elaphus*	Red Deer
Rangifer	*tarandus*	Reindeer
Bos	*primigenius*	Aurochs
Saiga	*tatarica*	Saiga antelope

Little is known about the process of recolonisation of the British Isles by mammal taxa. The largest body of work accumulated thus far is for small mammals. Studies using mitochondrial DNA and/or chromosomal characteristics have observed strikingly similar patterns of recolonization between small rodents and the hypothesised human recolonisation scenario for Britain, termed the 'Celtic Fringe' (Searle & Wilkinson 1987; Piertney *et al.* 2005; Searle *et al.* 2009; Brace *et al.* 2016). In peripheral geographic areas of Britain and north-west Europe (Scotland, Ireland, Wales, Brittany and north-west France), human occupants have a higher frequency of the mtDNA marker J/16192 and belong largely to the Y-chromosomal gene cluster R1b-14. In central Britain (England), higher frequencies of mtDNA maker J/16232 and Y-chromosomal cluster J/16192 are detected (Forster *et al.* 2004; Oppenheimer 2006). This has been explained as a population replacement event of the earlier Celts by later Anglo-Saxons (but remains contentious; Searle 2009). A number of small mammal species also show the same pattern of mtDNA population replacement and today have mtDNA haplotype distributions that mirror this; common shrew *Sorex araneus* Searle & Wilkinson 1987; bank vole *Myodes glareolus,* field vole *Microtus agrestis,* pygmy shrew *Sorex minutus* (Searle *et al.* 2009) and water vole *Arvicola amphibius* (Brace *et al.* 2016). However, little is known of whether this 'two-wave' pattern of recolonisation applies to larger mammals and no study as yet has directly investigated the impact of the abrupt fluctuations of the Lateglacial Interstadial– Younger Dryas – Holocene climatic changes on mammal populations.

1.1.10 Ancient DNA – history and overview

Until only a few decades ago, studies into micro-evolutionary patterns and processes within and between ancient Quaternary mammal populations relied mostly on what could be inferred from presence/absence data, morphological character assessment of stratified mammal remains and indirect inferences from the DNA of modern taxa. The initiation and subsequent development of techniques to extract and rapidly amplify ancient DNA (aDNA) has opened up an entirely new area of research into the evolutionary biology and population genetics of past mammalian fauna. Such studies are now becoming commonplace (*e.g.,* Hofreiter *et al.* 2001; Pääbo *et al.* 2004).

Any DNA that is of sufficient age and has been recovered from archaeological and/or paleontological sources can be considered aDNA. In the study of Quaternary mammals, aDNA is usually extracted from skeletal remains (bones and teeth, antler, ivory) and, less often, hair and coprolites (*e.g* Gilbert *et al.* 2008; Adler *et al.* 2011). *Post mortem* decay of DNA presents many methodological problems, particularly if the preservation environment has been poor. However, the relatively 'young' geological age of Pleistocene/Holocene material and the favourable preservation conditions in the cold regions of northern Europe (and further afield in the northern Holarctic) have meant that evolutionary and population genetics studies of Late Quaternary mammals from the northern hemisphere are some of the most abundant and best-resolved in the field of ancient DNA.

Ancient DNA is a relatively young science, with the first research article to report successful extraction and cloning of mammal aDNA published in 1984, having sequenced a 229bp fragment from the soft tissue of an extinct subspecies of *Equus quagga* (plains zebra; Higuchi *et al.* 1984). A year on, a 3,000bp long fragment was obtained from a 2,430-year-old Egyptian mummy (Pääbo *et al.* 1985). The following decades saw a flurry of high impact articles claiming to have extracted and amplified aDNA from specimens millions of years old (*e.g* dinosaurs, Woodward *et al.* 1994; amber inclusions, Cano *et al.* 1992*ab*, 1995; and humans, Zischler *et al.* 2003). The subsequent discovery that most of these results were either the product of contamination from modern DNA, or that they could not be reliably replicated, led to the field of aDNA research developing a reputation for unreliability. However, during the 1990s, less high-profile but more trustworthy studies were also published, which used aDNA to begin resolving the phylogenetic relationships of extinct Pleistocene fauna such as moas, mammoths and cave bears (Cooper *et al.* 1992; Hagelberg *et al.* 1994; Hånni *et al.* 1994). These studies essentially kickstarted the use of aDNA in molecular phylogenetics research but it was not until the 2000s that the first population level studies were published (*e.g,* brown bears Leonard *et al.* 2000; Barnes *et al.* 2002 and cave bears Orlando *et al.* 2002; Hofreiter *et al.* 2002).

Until the advent of aDNA, inferences of molecular phylogenetic affinities and population histories among Late Pleistocene and Early Holocene mammals were based exclusively on what could be extrapolated from modern DNA derived from extant taxa. The ability to answer phylo-chronological and population-scale questions directly through the analyses of ancient DNA has revealed a high level of previously unappreciated complexity in Pleistocene and Holocene mammals (see Ramakrishnan & Hadly 2009 for a review). Research has predominantly been directed towards the charismatic Palaearctic megafauna; the woolly mammoth (*Mammuthus primigenius*) and cave bear (*Ursus spelaeus*) being two of the most high-profile extinct species to be studied. Evidence from partial and whole mitochondrial genome sequencing has revealed major insights into the phylogenetic and demographic history of *M. primigenius,* including the existence of several previously unknown highly divergent clades, Late Pleistocene range expansions and relict island populations (Barnes *et al.* 2007; Gilbert *et al.* 2008; Graham *et al.* 2016). Cave bears, which went extinct *c.* 15 000 cal BP, demonstrated strong phylogeographic patterning with three reproductively isolated Eurasian phylogroups detected – evidence that supports pre-existing morphological research suggesting that these clades in fact represent distinct species (Hofreiter *et al.* 2004). To a lesser extent ancient meso and micro mammal evolution and population dynamics have also been investigated through the use of aDNA. For example, during climatic warming at the end of the last glacial, mid-latitude European populations of Arctic fox (*Vulpes lagopus*) have been shown to have become extinct as a result of an inability to 'track' their habitat north as it declined (Dalén *et al.* 2007).

The accumulation of ancient DNA data from a wide range of species has begun to allow the transition from research questions focusing on broad-scale phylogeography to more specific questions relating to fine-scale temporal and spatial demographic processes to be investigated. For example, Brace *et al.* (2012) uncovered evidence of Late Pleistocene instability in populations of the cold-adapted collared lemming (*Dicrostonyx torquatus)*, where a severe reduction in genetic diversity occurred over the last 50,000 years, during which time a number of local extinction-recolonisation events were discovered in NW Europe.

The advent of technological, methodological and bioinformatics advances such as high-throughput sequencing and in-solution hybridization capture have now brought ancient DNA into the genomics era. The multi-fold increase in both the quantity of DNA that can be retrieved with Next Generation Sequencing (NGS) and the time depth of the material from which DNA can be extracted has allowed whole genomes from Pleistocene mammalian fauna to be sequenced (Hofreiter *et al.* 2015). In 2008 a 0.7x coverage genome of a mammoth was completed (Miller *et al.* 2008); by 2014 the Neanderthal genome was sequenced to 50x coverage (Prűfer *et al.* 2014). As of 2015, the ancient genomes of over 30 species had been assembled (Hofreiter *et al.* 2015). This has allowed finer-scale analyses to be undertaken by incorporating nuclear DNA, which can identify not only matrilineal phylogeographic histories but also instances of migration, admixture, whole and partial population replacement, domestication and the impact of selection on genomic evolution (Orlando and Cooper 2014; Hofreiter *et al.* 2015).

1.1.11 Geometric morphometrics - history and overview

Until around the 1980s, the quantification and analyses of biological form was generally undertaken using sets of linear distance measurements, ratios and angles, which were then subjected to uni- and multivariate statistics (*i.e.* traditional morphometrics; Reyment 1996; Adams *et al.* 2013). These methods had limitations, however, as graphical depictions of shape and size changes were often lost in the measurements, restricting the biological interpretations that could be made (Rohlf and Marcus 1993; Adams 2004). The need for a method of quantifying the shape and size of biological forms that *preserved* geometric shape information led to a radical paradigm shift in the 1980s - the 'morphometric revolution' (Rohlf and Marcus 1993; Adams *et al.* 2013). This represented both a sweeping change in the means of acquiring shape information and in the way this data was analysed (Adams *et al.* 2013). The subsequent application of robust mathematical principles and integration with rigorous statistical and conceptual principles led Bookstein (1996) to describe this as the 'Morphometric Synthesis'.

Geometric morphometrics is the statistical analysis of shape variation and its co-variation with other variables. The approach utilises fundamentally different types of input data to traditional methods, using landmark co-ordinate points, outline curves and surfaces contours over ratio-based measurements and can be applied to both two and three-dimensional datasets (Zelditch 2004). All geometric morphometric analyses are based on the 'Procrustes Paradigm' (MacLeod 2009; Adams *et al.* 2013), which is implemented as a series of operations designed to superimpose constellations of landmark data across all specimens to a common co-ordinate system and output a new set of shape co-ordinates for downstream multivariate statistical analyses. This procedure removes the effects of size, rotation and position and allows for complete separation of size and shape variables that can then be analysed as discrete datasets (MacLeod 2009).

Quantification of shape by these means allow for far more diverse and complex testing of biological hypotheses than traditional methods. Often ordination methods such as Principal Components Analysis (PCA) and Canonical Variates Analysis CVA (CVA) are used (as in this study) to investigate patterns of shape variation between individuals and *a priori* defined groups respectively (*e.g* MacLeod 2005, 2007). Covariation between shape variables and other continuous variables can be tested with statistics such as Partial Least Square (PLS) and multivariate regressions while shape partitioning into various symmetric components can be used to test a multitude of hypotheses relating to environmental and genetic influences on shape (Rohlf and Corti 2000; Klingenberg *et al.* 2002). The separation of size and shape is also of huge advantage as it allows the allometric relationship between these variables to be examined (Adams *et al.* 2013). A final benefit to the use of geometric morphometric methods is the ability to use PCA and CVA ordination methods to visualise geometric transformations both between groups and in individuals through techniques such as the thin-plate spline (Bookstein 1991) and strobe models (Macleod 2009). Together, the geometric morphometric toolkit represents a powerful set of data analysis and visualisation techniques that offer far superior statistical power and shape change visualisation over traditional methods.

1.2 Thesis aims

In order to assess individualistic or common recolonisation and population level patterns and processes between species, this study focuses on three species with disparate ecological niches and dispersal abilities;

i) The common vole, *Microtus arvalis* (Pallas, 1778). A herbivorous, abundant and widespread micromammal associated with grassland ecosystems (Jacob 2013).

ii) The Eurasian beaver, *Castor fiber* (Linnaeus, 1958*).* An herbivorous, semi-aquatic, large-bodied rodent associated with forested riverine systems (Wilson and Reeder 2005).

iii) The European wildcat, *Felis silvestris silvestris* (Schreber, 1777). A medium-sized obligate carnivore, associated with woodland habitat (Wilson and Reeder 2005).

The broad aims of this study are:

Aim 1. Can the geographical source area for British populations of *M. arvalis*, *C. fiber* and *F. s. silvestris* be identified and can this be used to infer the most likely glacial refugium?

Aim 2. Can patterns of population continuity or discontinuity for *M. arvalis*, *C. fiber* and *F. s. silvestris* over the Lateglacial Interstadial – Younger Dryas – Holocene climatic periods be detected?

and

Aim 3. Does morphological evidence exist for phenotypic adaptation to abrupt changes in climate and/or population turnover events?

To achieve these aims, a synthesis of ancient DNA, 2D and 3D geometric morphometrics and ^{14}C dating is applied to bones and teeth from a range of British

cave and surface sites of Lateglacial to Holocene age. Where species-specific questions arise (for example conservation implications in the case of *C. fiber* and *F. s. silvestris*) these are addressed in the appropriate chapter. The combination of methods used here allow for a complimentary, multi-proxy approach to the research aims and also an opportunity to evaluate the applicability of these methods, in particular geometric morphometrics, to faunal material of Lateglacial to Holocene age.

Chapter 2. Methodology

2.1 Overview

This study employs a tripartite methodology that uses ^{14}C radiocarbon dating, 2D and 3D geometric morphometrics and ancient DNA applied to zooarchaeological material from a range of British sites. This allows both an analysis of genetic and ecomorphological change/stasis in response to a changing climate, underpinned by direct dating. The following chapter provides a general overview of the theory, background and technical details of methods employed here. Chapter-specific methodologies are discussed in those chapters.

2.2 Accelerator Mass Spectrometry (AMS) ^{14}C Radiocarbon Dating

All Accelerator Mass Spectrometry (AMS) ^{14}C radiocarbon dating analyses undertaken in this study were performed at the Oxford Radiocarbon Accelerator Unit (ORAU) at the University of Oxford (Appendix A, Table AT1, Fig AF1). Dates could only be obtained where material was available and sampling permission was granted. Where possible, material was sought for which existing ^{14}C dates were available. All samples that had been dated by previous authors prior to the introduction of the calibration curve IntCal13 (see below; Raimer *et al.* 2013) were recalibrated. No ^{14}C age estimates were included that had been derived prior to the introduction of ultrafiltration (see below). In some instances, (*e.g.,* for beavers) material that had ^{14}C age estimates pre-ultrafiltration were re-dated for this study. In the case of the common vole, *Microtus arvalis*, each individual was represented by a single molar and ancient DNA extraction resulted in complete destruction of this tooth. Therefore, age determination was carried out indirectly using associated molluscan material, mammalian biostratigraphical evidence and ^{14}C dating of co-occurring vertebrate bone (see Chapters 3 and 4).

2.2.1. Accelerator Mass Spectrometry (AMS) ^{14}C radiocarbon dating - general theory

Carbon-14 and non-radioactive carbon isotopes retain the same equilibrium ratio within the tissues of living organisms as that found in the atmosphere. *Post-mortem*,

living organisms no longer take in ^{14}C and thus cease to partake in the global carbon exchange reservoir (Taylor & Bar-Yosef 2014). The unstable nature of ^{14}C means that its concentration no longer remains constant and it will decay at an exponential rate by beta-decay. The excess energy created by the possession of 8 neutrons by ^{14}C results in slight nuclear instability. This excess energy is emitted as nuclear radiation in the form of beta (β) particles during nuclear decay, as ^{14}C transmutes to its daughter isotope Nitrogen-14 (^{14}N). The immutable and constant rate of ^{14}C radioactive decay along with the fact that this process is entirely unaffected by the surrounding environment, provides (in principle) a method for dating organic material of unknown age (Taylor & Bar-Yosef 2014). This is achieved through comparison of the amount of ^{14}C activity (rate of beta-emissions) with a modern sample of known age or, as in Accelerator Mass Spectrometry (AMS), the direct measurement of ^{14}C atoms and their ratio to carbon-12 (Taylor 2012). The rate at which ^{14}C concentration decreases is determined by the law of radioactive decay:

$$\boldsymbol{A = A_0 \exp(-\frac{t}{8033})}$$

Equation 2.1: Law of radioactive decay

where $\boldsymbol{A}$ is the $^{14}C/^{12}C$ ratio in a sample of age t years, $\boldsymbol{A_0}$ is the $^{14}C/^{12}C$ ratio of A for $t = 0$ (starting activity) and the value of 8033 years is the meanlife. The meanlife (τ) is the average life expectancy of a single ^{14}C atom and can be related to the half-life by:

$$\boldsymbol{T_{\frac{1}{2}} = 0.693\tau}$$

Equation 2.2. Relation of mean life to half-life

where 0.693 is the natural log (Ln) of 2. The half-life of ^{14}C is therefore 5568 years.

2.2.2 Calibration

There are a number of assumptions fundamental to the technique of ^{14}C dating (modified from Taylor, 2012):

- Atmospheric ^{14}C concentration has remained constant over millennia (this implies a constant rate of production and constant and rapid mixing, exchange and transfer rates in addition to constant sizes of reservoirs)
- The biosphere has the same overall concentration as the atmosphere (implies rapid mixing between these two reservoirs)
- The same ^{14}C concentration exists in all parts of the biosphere
- Carbon-14 exchange with the global carbon exchange reservoir ceases upon death
- *Post-mortem*, ^{14}C concentration is affected only by radioactive decay.

Many or all of these theoretical assumptions are not met and if a 'true' or 'calendar' age is to be established, these differences must be corrected for via calibration against a known standard. Calibration curves are created from both dendrochronological records and, beyond the limit of continuous dendrochronological data, from matched Uranium (U) series dating and ^{14}C dating of marine corals and foraminifera (Walker 2005). The most recent internationally agreed calibration curve (and which is applied in this study) for terrestrial material is IntCal13 (Reimer *et al.* 2013), a revision on IntCal09 and its predecessor IntCal04. This calibration curve uses tree-ring data (updated from IntCal04 by the addition of additional datasets from 0-12,550 cal BP) from the period covering 0-13.900 cal BP and supplemented by Lake Suigetsu plant macrofossil data from 1300 to the end of the dating range, 50 000 cal BP (Raimer *et al.* 2013; Fig 2.1).

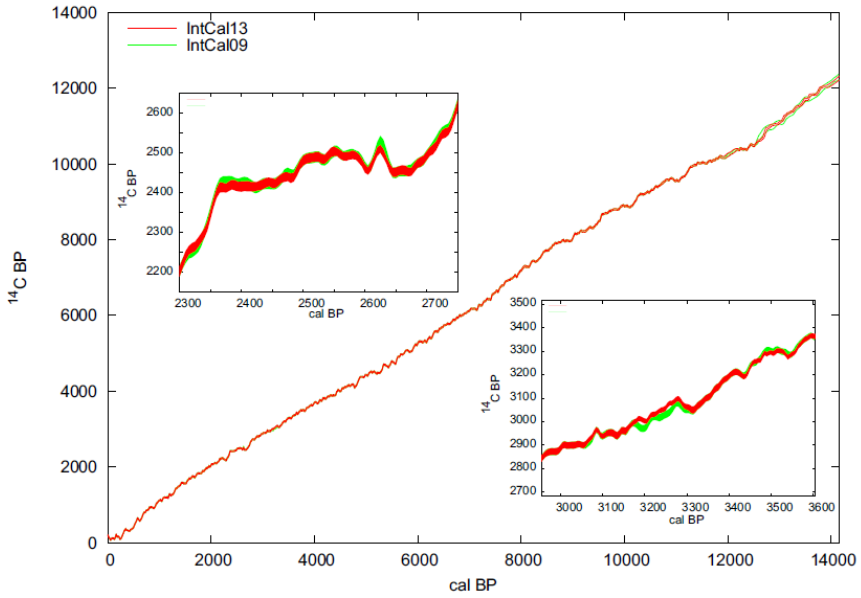


Figure 2.1. Comparison of the IntCal13 and IntCal09 calibration curves (Raimer *et al.* 2013)

2.2.3 Chemical composition of bone

Mammalian bone is a dense connective tissue largely composed of collagen (an organic protein) and inorganic minerals, which combine to provide structural support and facilitate mechanical function. Organic compounds account for *c.* 25 percent of the dry weight of bone and of this, around 90-95 percent is collagen (Jacobi *et al.* 2006). Collagen molecules are combined into triple helical structures (tropocollagen) composed of three polypeptides each comprising *c.* 1000 amino acid per chain ("alpha chains"). These form collagen fibrils and are embedded in the bioapatite matrix. The pattern and orientation of collagen fibrils distinguishes two types of bone;

- Cortical (or lamellar) bone with a regular, parallel alignment of collagen fibres organised into mechanically strong sheets (lamellae). This type of bone is compact with low porosity (*c.* 5 - 30%) and accounts for *c.* 80 percent of total bone mass in an adult skeleton.

- Cancellous (non-lamellar) bone with a haphazard organisation of collagen fibres and an open network of rod and plate-like elements. This bone type is

mechanically weak with a high level of porosity (*c.* 30-90%) and accounts for the remaining 20 percent of total bone mass (Wang *et al.* 2010).

Bone collagen is primarily targeted for ^{14}C dating as it is the dominant organic protein and significant methodological problems exist with other materials (*i.e* separation of indigenous bone carbonate from diagenetic carbonate can be problematic; Brown & Brown 2011). However, poor preservation and contamination of collagen in many depositional contexts poses significant difficulties in the ^{14}C dating process. The rate and nature of decomposition of bone tissue are dependent on the preservation environment after deposition, with a host of interrelated biotic and abiotic factors contributing to collagen diagenesis. The hydrological characteristics, temperature, pH, oxygenation and micro-organismal activity of the matrix in which the specimen is preserved will all have an effect on the pattern and process of diagenesis (Grupe 2007). Moreover, these will vary significantly between sites, over the period of preservation and pre- and post-excavation; generally, bone material preserved in colder and dry climates will present in a better state of preservation than material from warm and humid environments (Higham *et al.* 2006).

Immediately *post-mortem*, fungal and bacterial microorganisms begin to act on both the inorganic crystalline matrix and the organic fraction, with bone collagen being broken down and parts of the collagen fibrils being freed (Brown & Brown 2011). Hydrolytic damage causes collagen protein to break down first into peptides and subsequently into amino acids, while spontaneous re-arrangement of the inorganic crystalline matrix weakens the protein-mineral bond and increases bone susceptibility to dissolution (leading to loss of some amino acids) by internal and external agents (Grupe 2007). In addition, random cross-linking, humification of parts of the collagen molecule and attachment of exogenous humic materials all contribute to the diagenetic processes acting on collagen (Brown & Brown 2011).

In addition to diagenesis, post-depositional contamination is a major concern in ^{14}C dating of ancient bone samples. The extent and type of contamination is sample-specific and will vary considerably with the age and depositional environment of the

specimen. Natural contaminants of concern include rootlet intrusion and humic acids circulating in the soil/substrate matrix, which can adhere to samples or partake in carbon exchange, resulting in a date that is erroneously young (Brown & Brown 2011). Post-excavation, a huge variety of modern artificial contaminants (such as preservatives, tobacco ash, paper, hair and fabric fibres, etc.) can produce the same effect. Contamination is particularly problematic for ancient material that may have depleted levels of ^{14}C. Therefore, determining the nature and size of contamination and the relationship between sample, contaminant and magnitude of error is vital. Gupta and Polach (1985) have developed a methodology for determining the extent of the contamination error under the assumption that the contamination fraction occurred instantaneously (*i.e.* it did not vary temporally; a simplified scenario). This demonstrated a general relationship between the age of a sample and the effect of contamination on ^{14}C age in older samples - showing greater error even if the percentage of contaminant is small (Table 2.1; Gupta and Polach 1985)

Table 2.1. Errors associated with contamination of samples of known age with older and younger carbon. 'True' age is 900 BP and brackets show percent age error. From Gupta and Polach (1985)

	% Contamination by modern carbon					
	0%	1%	5%	10%	25%	50%
At 900 BP	900 (0)	890 (5)	850 (5)	810 (11)	670 (26)	440 (51)
	% Contamination by old carbon					
	0%	1%	5%	10%	25%	50%
At 900 BP	900 (0)	980 (9)	1320 (47)	1770 (47)	3280 (264)	6630 (637)

A variety of bio-indices are employed by radiocarbon dating laboratories to give proxy indications of the state of collagen preservation in ancient bone samples[1]. These include (but are not limited to) *i*) determining total collagen content via estimation of nitrogen (N) in the sample *ii*) analyses of stable carbon (C) and nitrogen (N) isotopes and iii) determination of C:N ratios. Pre-treatment of the collagen via the ultrafiltration method has proved to be one of the most successful methods for

[1] Collagen follows the definition of van Klinken (1999), referring to the insoluble residue than remains after the bone tissue has been decalcified and the triple-helical collagen molecule has been untwisted by denaturation, 'the Longin collagen method' after Longin (1971).

collagen purification and has been shown to be extremely effective in removing low molecular weight contaminants from archaeological bone samples (Higham *et al.* 2006). Collagen fibrils typically weigh between 95 and 102 kD (kilo Daltons). Ultrafiltration enables the separation of materials > 30 kD (including un-degraded alpha chains) from materials < 30kD (which will include degraded and broken down collagen fragments, salts and non-indigenous amino acids), thereby producing a purified gelatine fraction of high quality as shown by improved C:N ratios compared to other filtration methods (Higham *et al.* 2006). While problems do exist with this method, it has resulted in improved accuracy in chronological dating of many Late Pleistocene bone samples (Jacobi *et al.* 2006, with caveats) and is employed on all radiocarbon dating procedures in this project via the Oxford Radiocarbon Accelerator Unit (ORAU; Appendix A; Table AT1, Fig AT1).

2.3 Ancient DNA

Ancient DNA analysis is a major component of this study and has been applied to specimens ranging in age from the very early Lateglacial (*c.* 15,000 cal BP; common vole *M. arvalis*) to the Romano-British late Holocene period (1,814 – 1618 cal BP: wildcat *F. s. silvestris*). Analyses primarily target whole mitogenomes or the non-coding Control Region (CR) and the protein coding cytochrome*b* (cyt*b*) genes from the mitochondrial genome. In the case of the wildcat, whole mitogenomes are also incorporated into the analyses.

2.3.1 Ancient DNA – degradation and contamination

The major complications associated with ancient DNA are *i*) degradation *ii*) fragmentation and *iii*) contamination. *Post mortem*, autolysis occurs predominantly via the enzymatic action of endogenous nucleases, which ultimately results in the fragmentation of DNA polymers, with many ancient specimens possessing DNA fragments in the range of only 50-500 base pairs (bp) in size, and with a mean size that may be as low as 50bp (Pääbo 1989; Handt *et al.* 1994; Höss *et al.* 1996; Dabney *et al.* 2013). Hydrolytic damage can cause direct cleavage of the phosphodiester backbone, creating breaks in the DNA strand (the prime causal factor of both DNA loss and short template length in ancient material), depurination and breakage of the

sugar backbone creating baseless sites (Dabney *et al.* 2013). Hydrolytic deamination of bases can result in miscoding lesions that potentially lead to problems during the polymerase chain reaction (PCR), where incorporation of erroneous bases can occur (Fig 2.2).

Transitions tend to be the most frequently observed miscoding lesions: adenine to guanine and *vice versa* (A→G/G→A) and cytosine to thymine and *vice versa* (C→T/T→C). These have been classified into Type I (A→G/T→C: deamination of adenine to hypoxanthine) and Type II (C→T/G→A: deamination of cytosine to uracil) with Type II miscoding lesion being the most commonly observed (Hofreiter *et al.* 2001; Gilbert *et al.* 2003; Dabney *et al.* 2013). Oxidative processes can also modify bases and lead to 'blocking lesions', which block *Taq* polymerase and promote 'chimeric sequences', amplicons where the DNA sequence originates from two or more transcripts (Willerslev & Cooper 2005). The final major contributor to DNA degradation is the formation of intermolecular or inter-strand crosslinks by either Maillard reaction or alkylation. These can increase the risk of contamination by preventing the amplification of endogenous template molecules (Hofreiter *et al.* 2001; Willerslev & Cooper 2005; Kefi 2011). While some measures can be taken to undo some of the decay to DNA molecules *e.g.,* treatment with uracil-N-glycosylase to remove erroneous uracil base substitutions and the use of ligases and N-phenacylthiazolium bromide to reverse cross-linkage (Reiss 2006), most damage to aDNA molecules is presently irreversible. The degree of degradation and chemical alteration will depend almost exclusively on the age of the specimen and the environment in which it was preserved; low temperatures, rapid desiccation and a high salt content in the surrounding medium will yield the best quality aDNA (Hofreiter *et al.* 2001; Dabney *et al.* 2013). Unsurprisingly, permafrost environments have generally provided authenticated aDNA of the greatest age with plant and invertebrate DNA from Arctic ice cores of around 800,000 years old successfully amplified via PCR (Willerslev *et al.* 2003; Willerslev *et al.* 2007) and the genome of a *c.* 700,000-year-old Middle Pleistocene equid sequenced via Next Generation Sequencing (Orlando *et al.* 2013).

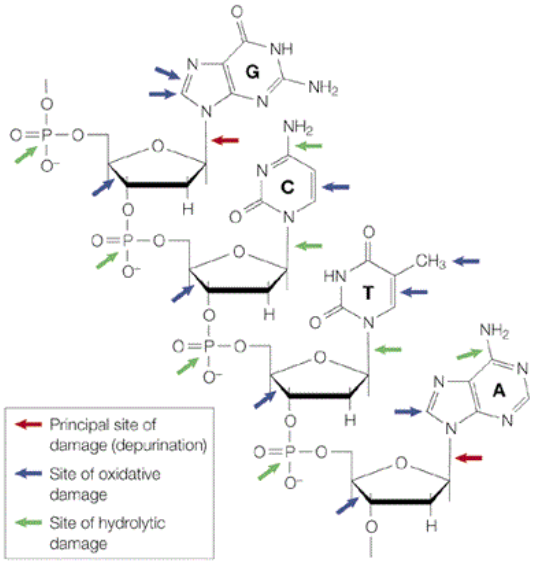


Figure 2.2 Principal sites of depurination, oxidative and hydrolytic damage in ancient DNA. Sites vulnerable to depurination indicated with red arrows, oxidative damage with blue arrows and hydrolytic damage with green arrows. From Hofreiter *et al.* (2001)

Contamination is a chief concern in aDNA analyses and can occur through a multitude of pathways, from the site of preservation to the laboratory. The small and fragmented nature of most aDNA starting templates means that if the sample becomes contaminated with modern DNA, this non-target modern DNA can then be preferentially amplified during PCR. Contamination prior to analysis can be a major problem in palaeo-biological and archaeological studies where specimens have typically been exposed to a host of DNA sources beginning at the site of preservation and continuing through the excavation process to handling, cleaning and preservation. In addition, there is huge potential for contamination in the laboratory from previously amplified PCR products. A successful 50µl PCR reaction can contain 10^{12} - 10^{15} amplified DNA molecules and a single microscopic water droplet can contain over a million copies per 0.0005µl (Willerslev & Cooper 2005). It is therefore essential that aDNA labs are isolated from other molecular facilities (particularly post-PCR labs) and that users only move from an aDNA laboratory to a modern laboratory (up the concentration gradient). False-positive results within aDNA studies are still a major problem but adherence to strict protocols (below) should reduce the likelihood of contamination and help authenticate results (Cooper & Poinar 2000; Willerslev & Cooper 2005).

Two further considerations when working with mitochondrial DNA are, *i*) the insertion of mitochondrial DNA (mtDNA) sequences into the nuclear genome (often referred to as nuclear mitochondrial DNA or, simply, 'NuMTs; Lopez *et al.* 1994) and, *ii*) heteroplasmy. Mitochondrial DNA can invade the nuclear genome during genome evolution, after duplication (Cooper & Poinar 2000), and can seriously mislead phylogenetic and population genetic analyses if mistaken for genuine mtDNA sequences. Detection of NuMTs can be accomplished by amplifying overlapping DNA fragments and/or by cloning PCR products (Willerslev & Cooper 2005), but the latter can prove expensive. Heteroplasmy refers to the presence of multiple mitochondrial sequences within a cell or individual (Wallace & Chalkia 2013). This can appear in the data as double peaks in chromatograms and as multiple base calls per site in high-throughput sequencing reads. Heteroplasmy can be present in a variety of forms but it is often observed as either single heteroplasmic mutations or as strings of tandem repeats or minisatellites, particularly in the Control Region (Lightowlers *et al.* 1997; Wallace & Chalkia 2013). It can be particularly problematic to identify in ancient DNA due to the confounding effects of highly degraded DNA, high-throughput sequencing error and the potential presence of NuMTs (Edwards *et al.* 2010; Li *et al.* 2012). While few ancient DNA studies have specifically addressed this problem, heteroplasmic sequences have been identified in the extinct auroch (*Bos primigenius*; Edwards *et al.* 2010) and the mammoth (*Mammuthus primigenius*; Rogaev *et al.* 2006).

2.3.2 Ancient DNA sampling, extraction, amplification and sequencing

Drilling was carried out with a hand-held Dremel Drill (Model 398) and 2-4mm drill bits. Drill-induced heat damage has been shown to substantially decrease mtDNA yields (Adler *et al.* 2011) so drill speeds were kept below 1000 RPM. Where teeth were available, the cementum-rich tooth root tip was sampled as this area has been shown to yield higher quantities of DNA than dentine (Adler *et al.* 2011, Higgins *et al.* 2013). Sample sites in tooth and cranial and post-cranial bone material depended largely on what skeletal remains were available and also on whether permission to sample particular parts of the bone was granted from curators. In general, dense and compact parts of the bone - such as the diaphyses of long bones - were targeted as

these have generally been shown to yield the highest concentrations of aDNA (Pinhasi *et al.* 2015).

This study employed an aDNA extraction protocol optimised to enable retrieval of phylogenetically-informative sequences from material where surviving aDNA length may be as short as < 50bp (Dabney *et al.* 2013; full protocol in Appendix A, Protocol A1). Bone or tooth powder is first subjected to an 18 hour overnight initial digestion stage in ethylenediaminetetraacetic acid (EDTA) and proteinase K to decalcify and lyse cells. The resultant supernatant containing DNA is then added to a binding buffer solution containing sodium acetate, sodium chloride and guanidine thiocyanate and bound to a silica membrane (Rohland & Hofreiter 2007). Silica particles and conjugated DNA are purified and de-salted by repeated ethanol washes prior to elution in a low-salt TET buffer (Tris-HCL, EDTA and Tween-20; Dabney *et al.* 2013).

From that point, DNA was either amplified and sequenced via PCR and Sanger sequencing or subjected to DNA library builds and/or in-solution hybridisation capture enrichment before shotgun sequencing on an Illumina NextSeq500 platform (Figure 2.3). PCR was used only when small regions of the mitochondrial genome (*c.* 500bp or less) were required to be amplified. As ancient DNA typically survives in small fragment lengths, multiple and overlapping nested primer pairs were designed to target regions of *c.* 80-150bp to increase the likelihood of successful amplification. Overlapping fragments ensure miscoding lesions and contaminants are more easily identified and, as a further precaution, PCR reactions and Sanger sequencing were repeated at least twice for each target region to ensure accurate base calls (PCR protocol in Appendix A, Protocol A2).

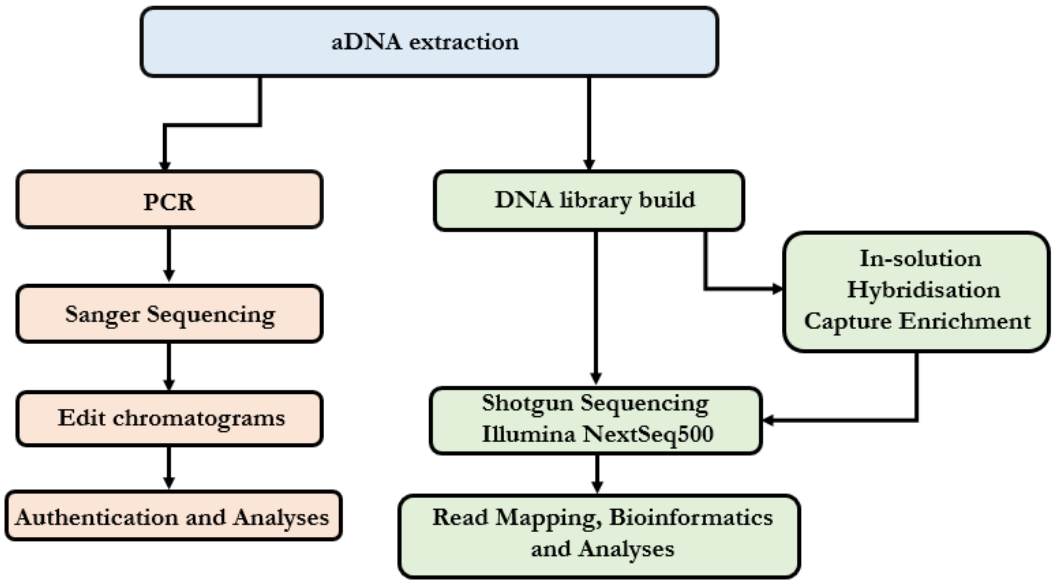


Figure 2.3. Generalised workflow used in ancient DNA component of this study.

Next Generation Sequencing (NGS) technologies were performed to *i*) screen aDNA extracts for endogenous (target) DNA content *ii*) sequence products of in-solution target enrichment and *iii*) sequence genome-wide reads from samples that had not undergone the enrichment procedure. Prior to shotgun sequencing and capture enrichment, all extracts were subjected to single-index, double-stranded DNA library builds following the protocol of Meyer & Kircher (2010; Fig 2.4; Appendix A, Protocol A3). The library procedure included a number of steps which *i*) convert isolated DNA fragments to blunt end DNA *ii*) ligate universal adapter sequences to the DNA fragment and *iii*) attach an index (barcode) sequence to the DNA-adapter fragment prior to target enrichment and/or parallel sequencing on the Illumina NextSeq500 platform. Library enrichments were performed when whole mitogenomes and/or nuclear exomes were required but could not be retrieved by random shotgun sequencing alone.

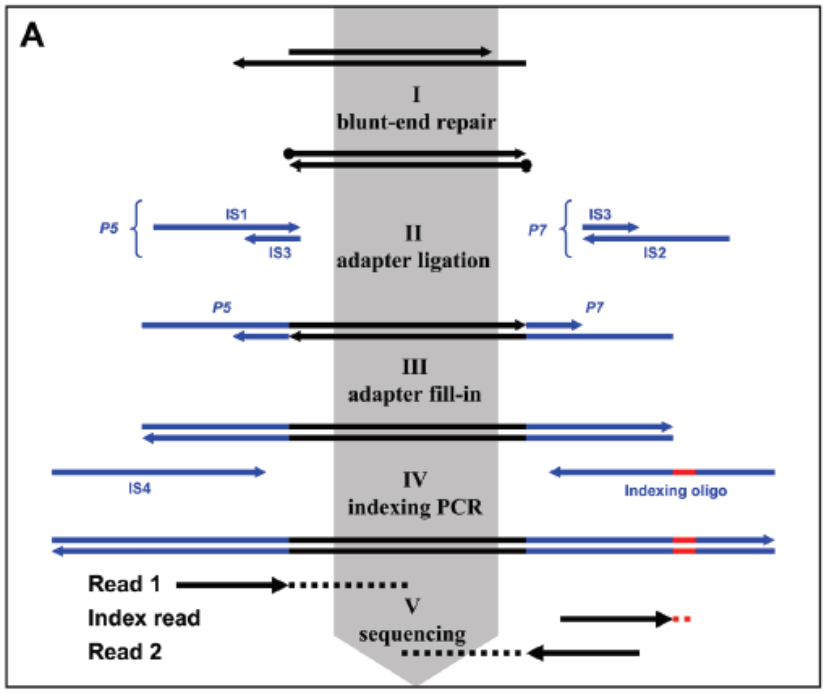


Figure 2.4. Schematic of the DNA library build protocol used in this study. From Meyer & Kircher (2010). **I.** Blunt-end repair - overhanging 5' and 3' ends of small fragments of aDNA are filled in or removed by T4 DNA polymerase. **II.** Adapter ligation – 5' phosphates are attached using T4 polynucleotide kinase. Two different adapters, P5 and P7, are ligated to both ends of the molecules using T4 DNA ligase. **III**. Adapter fill-in – Bst polymerase, which possesses strand-displacement activity, removes nicks in the fill-in process. **IV**. Index PCR. Indexes and full-length adapters are added by amplification with 5'-tailed primers.

This procedure utilised synthesised RNA 'bait' molecules to hybridise with the target DNA region before removal with streptavidin magnetic beads prior to amplification and shotgun sequencing (Figure 2.5; Appendix A, Protocol A4). The number of rounds of PCR amplification, both in library builds and target enrichment were kept to a minimum to avoid loss of library complexity via the amplification of excess duplicate reads. Sequencing was performed on a single-lane on an Illumina NextSeq500 within the sequencing facilities at The Natural History Museum, London, using a 2 x 75bp mid-output kit.

Read mapping of paired-end Illumina reads to a reference sequences was performed either on the CLC Workbench software v.8 (CLC Bio-Qiagen, Aarhus, Denmark) in cases where single mitochondrial genes were of interest, or via the Burrows Wheeler Aligner (BWA: Li & Durbin 2009) where whole nuclear or mitogenome mapping was required. In CLC, paired Illumina reads are imported into the workbench where sequences are trimmed using quality scores, ambiguity criteria and read length indices.

CLC (and BWA) use Phred quality scores (Q) based on a modified-Mott trimming algorithm calculated as:

$$Q = 10log10(P)$$

Equation 2.3. Calculation of Phred quality scores (Q)

where *P* is the base-calling error probability used to set the limit for which bases should be trimmed. The default value of 0.05 was used for this parameter. The maximum number of allowed ambiguities per read was set at 2 and reads below lengths of 25bp were discarded. Reads were merged using CLC default parameters of 2 for mismatch cost and 3 for gap cost. Trimmed and merged reads are then mapped to a genome reference sequence. Alignment parameters used default insertion and deletion costs of 3 and 3 respectively. Length fraction (minimum percentage of total read alignment length that must match the reference sequence) and similarity fraction (minimum percentage identity between aligned region of the read and the reference sequence) were both set to 0.94. Thus, at least 94 percent of the alignment must have at least 94 percent identity to the reference sequence. To map paired-end Illumina reads with BWA (Li & Durbin 2009), parameters were optimised for both ancient DNA and Illumina platform specific sequencing error as in Schubert *et al.* (2012). As BWA uses the first 32 nucleotides as a 'seed region' and aDNA typically exhibits high rates of mismatches at the 5' end, this option was disabled. Mismatch penalty (*-n*) was set to allow a 2 percent uniform base error rate (0.04) and all other parameters set so that this mapping procedure was analogous to those used in CLC.

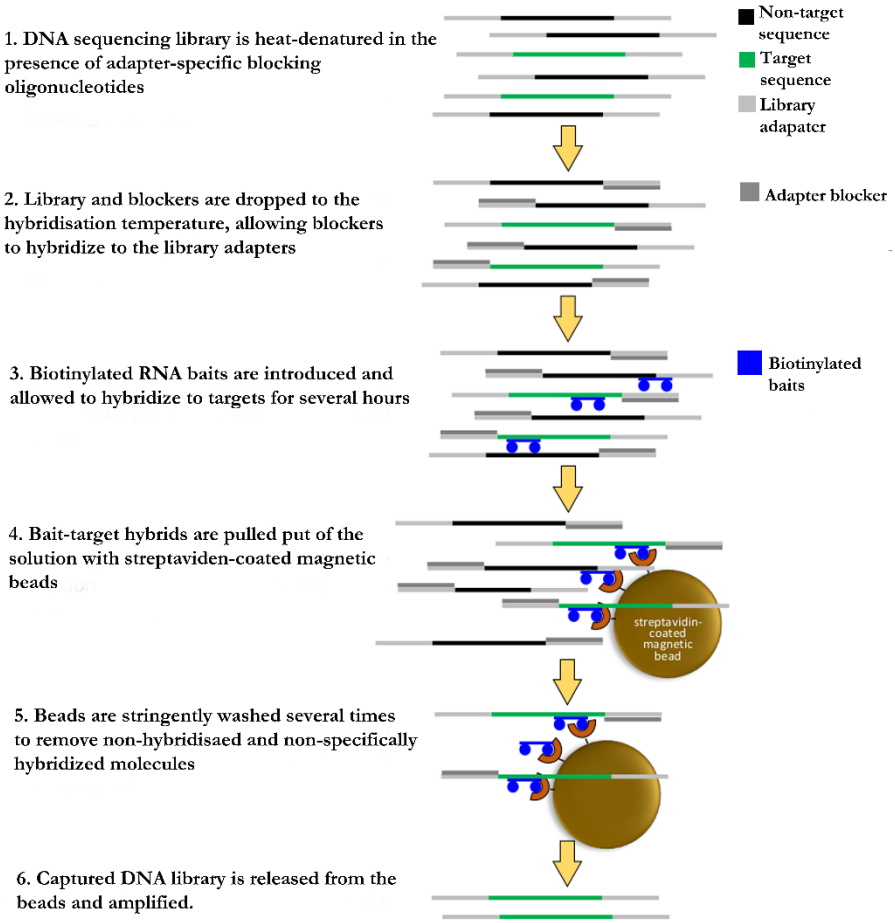


Figure 2.5 Schematic of the MYBaits in-solution target enrichment protocol.

2.3.3 Precautions and authentication

Drilling for bone and tooth powder was carried out on bleach-sterilised workbenches with metalwork, drills, plastics and tinfoil that had been thoroughly UV-irradiated prior to use. Bone and tooth surfaces were gently scraped with a 2mm drill bit to remove surface contaminants before drilling for internal bone/tooth powder. All DNA extractions and PCR recipes were performed within a fume hood in the dedicated ancient DNA laboratory at The Natural History Museum, London, which is physically isolated from all post-PCR laboratories. Surfaces, plastics and re-agents (where appropriate) in the lab were either bleach-sterilized or UV irradiated. Negative controls to check for exogenous contamination were run with all extraction, PCR and library build procedures. Positive controls were not carried out due to the risk of contamination. Repeat PCR and Sanger sequencing were carried out on all target regions to ensure correct base calls. PCR-generated DNA sequence contigs were subjected to the Basic Local Alignment Search Tool (BLAST; Altschul *et al.* 1990) to

confirm correct taxonomic identification and were translated to ensure no nonsense base calls. Authentication checks for Illumina reads included BLAST searches of *de-novo* assembled contigs and quantification of diagnostic patterns of damage associated with aDNA using MapDamage2 (Jónsson *et al.* 2013).

2.3.4 Phylogenetic analyses

Molecular phylogenetics allows the study of the evolutionary relationships between biological entities (species, individuals and genes) from DNA sequence data based on the observation that closely-related species, individuals and genes share a common ancestor more recently than distantly-related species, individuals and genes. In general, genomes evolve by the gradual accumulation of point mutations over time. Therefore, the amount of difference between nucleotide sequences in a multi-gene/genome comparison can indicate how recently those genomes shared a common ancestor. This study was solely concerned with phylogenetic relationships below the level of the species (*i.e* population level) and employed both phylogenetic tree-building methods and haplotype networks to reconstruct the phylogeography and population histories of the species of interest.

Nucleotide substitution rates are heterogeneous across the mitochondrial and nuclear genomes, between species and among sites within a genome (Rodriguez *et al.* 1990; Yang 1996). The rate of substitution is affected by a variety of factors including *i*) selection pressures and genetic drift, *ii*) functional constraints *iii*) base position (with third codon positions generally showing a higher mutation rate than first and second codon positions), *iv*) the relative impact of mutagens and *v*) the rate of repair (Yang 1996; Comeron *et al.* 2008; Bromham 2009). There is also a strong correlation between clock rate variation and systematic components of mammalian life history traits, with the molecular clock rate varying consistently with parameters associated with body size, population dynamics, lifestyle and location (Bromham 2009 and references therein). Small mammals with short generation times typically have higher rates of germline cellular copying/recombination events, which will provide more opportunity for mutations to arise than in larger mammals with longer generational turnovers (Bronham *et al.* 1996, Bromham *et al.* 2009). As the second largest rodent

in the world, the Eurasian beaver *C. fiber*, for example, shows a much slower mutation rate across its mitochondrial genome than smaller-bodied rodents (Horn *et al.* 2014).

DNA sequences can gather point substitutions via a number of pathways; *i*) multiple substitutions *ii*) single substitutions *iii*) parallel substitutions *v*) convergent substitutions and iv) back-substitutions (Yang *et al.* 2006; Foster 2009; summarised in Figure 2.6). Model-based phylogenetic methods used in this study, such as the Bayesian Markov chain Monte Carlo (MCMC) approach, estimate the probability of the hypothesis (or model) given the data and, as such, require an appropriate model of nucleotide substitution (Foster 2009). Models of nucleotide evolution are based on the probability of change from one nucleotide state to another, incorporating the amount of time that has passed and the substitution rate. Point mutations can be classified into transitions (purine to purine or pyrimidine to pyrimidine) and transversions (purine to pyrimidine or pyrimidine to purine; Figure 2.7) with transitions being more often observed (Kimura 1980). This study calculates all appropriate nucleotide evolutionary models using JModeltest2 (Darriba *et al.* 2012) which, in addition to calculating nucleotide substitution models, estimates two further parameters – the gamma shape distribution (α), which is proportional to the amount of among-site rate difference and the proportion of invariable sites (I), which is an estimate of the proportion of sites in the dataset that are unchanging. It has been argued that the proportion of invariable sites parameter (I) should not be incorporated into models of rate heterogeneity within population-level datasets (Jia *et al.* 2014). This parameter assumes that a proportion of sites in a sequence are completely resistant to change and this assumption is commonly violated in DNA sequences below the level of the species as 'invariant' sites are difficult to distinguish from those that simply have not changed yet. Therefore, this parameter was not estimated in any of the intraspecific datasets used in this study.

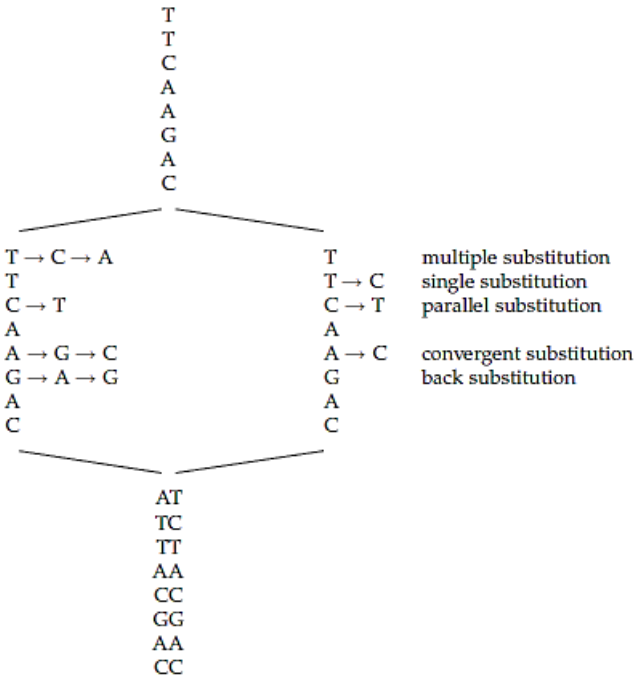


Figure 2.6. DNA substitution types during speciation. Modified from Yang (2006).

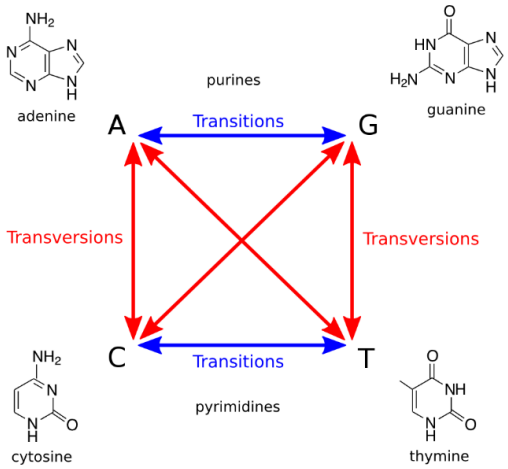


Figure 2.7. Transition and transversion point mutation types.

2.3.5 Mitochondrial and nuclear markers

This project targeted almost exclusively mitochondrial markers for phylogenetic inference. Partial or whole mitochondrial genomes are widely used in studies of ancient population genetics and evolution, primarily due to their high cellular copy

number, maternal pattern of inheritance, well-defined set of contiguous genes and increased mutation rate in comparison to nuclear genomes (Hofreiter 2001). The mammalian mitogenome is generally around 16-17kb in length with a structure, genetic content and organisation that is highly conserved among mammalian species (Saccone *et al.* 1991). It exists as closed circular strands with a set of 13 protein-coding genes, two ribosomal RNA (rRNA) genes and a full set of 22 transfer RNA (tRNA) genes that are asymmetrically distributed on two strands which can be distinguished on the basis of their G+T content (Fig 2.8). The Heavy-(H-) strand encodes the majority of genetic information; the 2 rRNAs, 14 tRNAs and 12 out of 13 messenger RNAs (mRNAs) while the Light-(L-) strand encodes the remaining 8 tRNAs and one mRNA. The mitogenome contains two non-coding regions, which contain most of the known regulatory functions. The larger of these is the Displacement Loop (D-Loop), situated between the genes for $tRNA^{PHE}$ and $tRNA^{PRO}$.

The regions of the mitogenome used for phylogenetic reconstruction in this study are cytochrome*b* oxidase and the Hypervariable Region I (HV-I) of the Control Region (CR). Whole mitogenomes are analysed for the European wildcat (*F. s. silvestris*) only. Cytochrome*b* is a well-conserved region showing slow evolution of non-synonymous substitutions and relatively fast evolution of silent positions (Irwin *et al.* 1991). The CR is particularly suited to population level analyses due to its hypervariable flanking regions, which accumulate rapid mutations relative to other regions of the mitochondrial genome. However, the CR can often exhibit insufficient resolution to reconstruct deeper levels of a phylogeny due to the accumulation of back mutations (site saturation; Avise 2000). Both of these regions have been shown to be informative in population-level analyses of ancient mammalian DNA (*e.g.,* collared lemming Brace *et al.* 2012; Eurasian beaver Horn *et al.* 2014; water vole Brace *et al.* 2016).

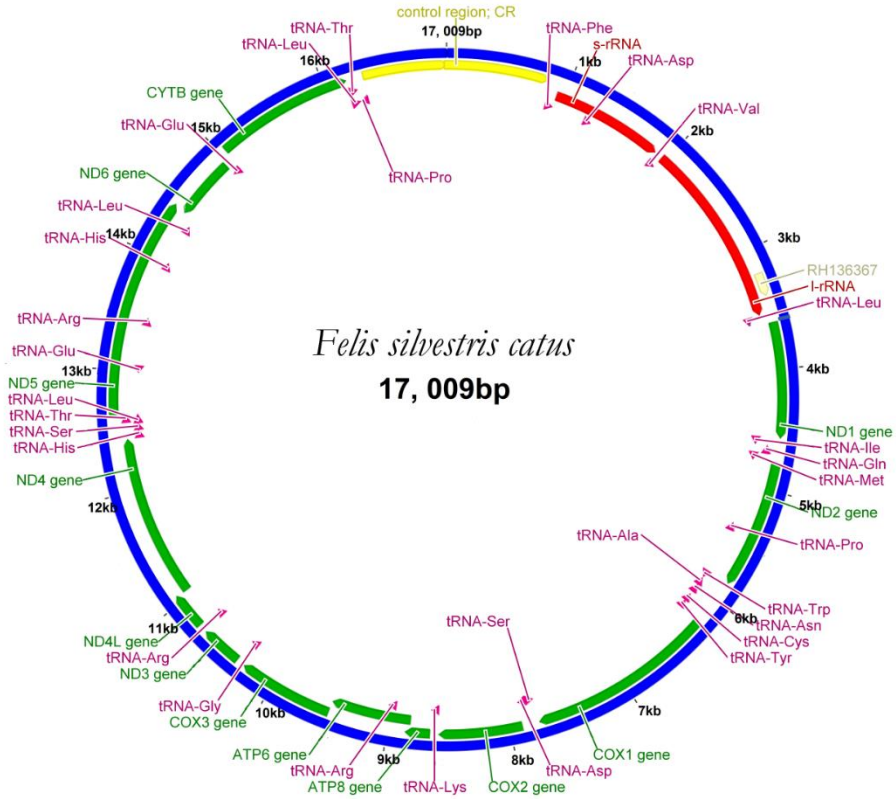


Figure 2.8. Annotated illustration of the mammalian mitochondrial genome. Example from domestic cat *Felis s. catus*. H- and L-strand not indicated but can be inferred from transcription direction. Annotations from Genbank *Felis s. catus* accession # NC_001700. Annotated in Geneious v. 7.1.7 (http://www.geneious.com, Kearse *et. al.*, 2012).

Attempts to recover ancient nuclear genes were made only for *F. s. silvestris* via in-solution hybridisation target enrichment. Use of single, non-recombining, mitochondrial genes in reconstructing phylogenies can suffer from the effects of incomplete lineage sorting (*e.g.,* where the coalescence history of the gene and that of the species show incongruence), site saturation and a lack of informative characters (Hurst and Jiggins 2005). The addition of phylogenetically informative nuclear genes, which have a slower mutation rate and can recover deep splits phylogeny, are often used in addition to mtDNA genes (Springer *et al.* 2001). However, population-level studies are increasingly moving away from tree-based phylogenetics and towards genotypic analyses of population histories using non-linked, biparentally inherited, neutral nuclear markers (*e.g.,* Brumfield *et al.* 2003). These are based on coalescent theory, allow the estimation of population parameters such as genetic diversity, divergence times, growth rates, gene flow and migration events that are not easily

detectable by single-locus phylogeographic methods. The choice marker in ancient DNA studies are Single Nucleotide Polymorphisms (SNPs; Sanchez *et al.* 2006), typically bi-allelic, single base mutations widely dispersed throughout the genome. SNPs exhibit a slower mutation rate than microsatellites (*c.* 10^{-8} – 10^{-9} compared to 10^{-4}; Brumfield *et al.* 2003) and it is estimated that three times as many SNPs than microsatellites are needed to gain comparable statistical power (Brumfield *et al.* 2003). The fragmented nature of ancient DNA is not well suited to recovering long stretches of tandem repeats that characterise microsatellites. Moreover, the slower mutation rate in SNPs means that these markers do not usually show multiple mutations at a single site, thereby being particularly suitable for high-throughput sequencing. Genome-wide SNP analyses of population history is beyond the scope of this study due to the lack of existing modern and ancient data with which to identify SNP polymorphisms.

Two methods of tree-building were used in this study – a distance based method, Neighbour Joining (NJ) and a model-based method, Bayesian MCMC. The NJ algorithm applies an agglomerative process to an arbitrary distance matrix in order to construct a bifurcating phylogenetic tree. A star-shaped starting tree is used alongside evolutionary distance data to find pairs of operational taxonomic units (OTUs), which minimise the total branch length across the tree topology at each stage of the clustering process (Saitou & Nei 1987). The benefits of this method is that it can generate a rapid and accurate phylogenetic tree. However, when phylogenetic information in a dataset is poor, NJ often does not have the sensitivity to produce a fully resolved phylogeny. The Bayesian MCMC method implements a likelihood function to estimate the likelihood of the tree topology + model of sequence evolution, given the sequence data, with the aim of producing a probability distribution over the full range of possible tree topologies and model parameter values. The optimal tree is the one that optimises the posterior probability, which is proportional to the likelihood multiplied by the prior probability (Holder & Lewis 2003). This is achieved by use of the MCMC algorithm, where several chains are run in parallel in a 'random walk' through tree and parameter space sampling values in proportion to their probability distribution. The Bayesian MCMC method does not

attempt to find the highest point in the space of all parameters but rather rejects or accepts random trees and parameters based on the value of the posterior probability (Holder & Lewis 2003). After a burn-in period, likelihood values settle on a convergence plateau and samples from this space can be used to construct a consensus tree. This method allows the incorporation of complex models of evolution and has more power to produce fully resolved trees from information-poor, single-gene datasets, but is considerably more computationally intensive. A Bayesian statistical framework is also applied to carry out genealogy sampling in Bayesian Evolutionary Analysis Sampling Trees (BEAST; Huelsenback *et al.* 2001). This uses the Bayesian MCMC algorithm and weighted posterior probabilities while also incorporating explicit molecular clock information and demographic models into the analysis.

Evolutionary relationships below the species level differ fundamentally from those between species. Relationships between genes sampled from species are hierarchical in nature as they have been shaped by reproductive isolation and fission of populations over long time scales. Conversely, relationships between genes sampled from within a species are the result of sexual reproduction, small numbers of recent mutations and (in nuclear genes) recombination. Tree-building methods make assumptions about evolutionary processes that may be violated when applied to intraspecific data and this can result in poor resolution or inaccurate portrayal of genealogical relationships (Posada and Crandall 2001). Trees assume a bifurcating pattern of relatedness, whereas population geneologies often show multifurcating and/or reticulate patterns that reflect interconnecting genealogical pathways. Furthermore, the lower rate of divergence between conspecifics means that DNA sequence data typically have fewer informative characters, which lowers the statistical power of tree-building methods (Posada and Crandall 2001).

Taking these factors into account, this study used two network based approaches Median-Joining (MJ) and Neighbour-Net (NN), in addition to the Bayesian model-based tree-building approach. The MJ method is specifically targeted at recombination-free population level data sets and is thus particularly useful for constructing intraspecific networks from mtDNA haplotypes (Bandelt *et al.* 1999).

This algorithm combines both the minimum spanning network (MSN) and the quasi-median network with distances between sequences calculated using the 'Hamming distance' technique (Bandelt *et al.* 1999). As a first-step, the user defines a parameter Epsilon (ε) that determines the search size for new median vectors. For example, an ε value of zero considers minimal length connections only. Subsequently, median vectors are generated by adding links between triplets of sequence-types with a minimum of two feasible links. The result is a network with a scaffold of median vectors and nodes showing all feasible links in minimal length connections.

Neighbour-Net (NN) is a distance based approach, similar to split decomposition, based on the original Neighbour-Joining (NJ) tree-building method of phylogenetic reconstruction (Bryant & Moulton 2004). Neighbour-Net uses a similar process to that used in NJ tree construction but instead constructs a multifurcating network. The distance matrix is decomposed into weighted partitions and represented using a splits graph, a graphical representation of a collection of the weighted splits in the data. Each split represents a branch connecting partitions delimited by the split, with parallel boxes defining contradictory groupings (incompatible splits) and indicating alternative pathways between sets of haplotype data. Each edge has a length equal to the weight of the split. In contrast to MJ, this method takes all characters into account, including sites in the alignment with gaps, ambiguities and missing data, and does not collapse unique haplotypes (Bryant & Moulton 2004).

2.4 Geometric Morphometrics

The following methods are described in technical detail for two-dimensional (2D) data (*x* and *y* co-ordinates) but can be simply extended to three dimensional (3D) data by the addition of a *z* co-ordinate. Geometric morphometrics were applied only to the vole dataset in this study as material of sufficient quantity and condition was problematic to source for other species within the timescale of this project.

2.4.1 Data acquisition and landmark placement

Two and 3D representations of the lower first molar of the field vole *Microtus agrestis* and the common vole *Microtus arvalis* were acquired by stereomicroscope and micro-computed tomography (microCT) respectively (see Chapters 3 and 4 for full details of these methods). The resultant 2D and 3D images/scans were used as the basis for acquiring landmark point data in the form of raw Cartesian co-ordinates. Landmarks should represent discrete points of correspondence between each specimen in a dataset, which are topologically[2] homologous anatomical loci that can be reliably located on each object (Bookstein 1991; Dryden & Mardia 1998). This results in a configuration of landmarks that serves as a shape proxy for the object and can be described by *km* variables (where *k* is the number of landmarks per configuration and *m* is the number of dimensions, either 2 or 3; Webster 2010).

Landmark selection is a crucial part of the geometric morphometric process and should be chosen so as to provide an adequate proxy of the morphological features of interest. As these are the only points that are considered in downstream analyses, shape variation in regions between these points will not be detected. Landmarks should also be consistently replicable, with a high degree of accuracy across the dataset using the chosen digitisation methods. Three types of landmarks are defined by Bookstein (1991):

Type I – discrete juxtapositions of tissues (such as the intersection of two suture lines).

Type II – maxima and minima of curvature (such as the deepest point in a depression or the most projecting point of a process). In this scenario claimed homology is supported only by geometric evidence.

[2] The datasets used in this study are all either below the level of the species or between species belonging to the same genus. Therefore, the structures standard landmark points are based on can be considered 'biologically homologous' in the sense that they share similarity due to descent from a common ancestor (Smith 1990). However, the definition 'topologically homologous' is preferred here as it is regarded as a more accurate description of mathematically defined landmark points (MacLeod 2001).

Type III – extremal points or points which are defined by virtue of information at other locations on that object (such as the endpoint of a curve or feature; Cooke & Terhune 2015). These generally have at least one deficient co-ordinate and may be reliant on the locations of other landmark points.

While Type I landmarks are the preferred landmarks to use, the strict criteria surrounding their definition means that these are difficult to locate on most biological structures and particularly on molar teeth. This study generally focuses on Type II and Type III landmarks, the last of which generally encompass semi-landmarks. These are a series of points located relative to one another by a consistent rule (Bookstein 1997). They allow the representation of shape information along outlines, facilitating the capture of complex shape information in areas of an object where standard landmark points cannot be located (Green 1996; Bookstein 1997; Perez *et al.* 2006). In this study they are located along tooth cusp outlines and anchored by the placement of Type II landmarks (see Chapter 3).

2.4.2 Centroid size and Procrustes GLS superimposition

Centroid size (CS) is the standard size proxy used in geometric morphometrics. It is calculated as the square root of the sum of squared distances of a set of landmarks and the centroid of the form (Bookstein 1991). Centroid Size is calculated as:

$$\sqrt{\sum_{i=1}^{N}[(x_i - \bar{x})^2 + (y_i - \bar{y})^2]}$$

Equation 2.4. Calculation of Centroid Size (CS)

Where *N* = number of landmarks. As CS is calculated from landmarks and not direct distance measurements as in traditional morphometrics, its value will change dependent on the number, location and arrangement of landmarks used to summarise the shape of a given object (Rohlf & Slice 1990). However, it is a useful proxy metric of overall size of any particular landmark configuration because in the absence of allometric effects, it is mathematically uncorrelated with Kendall's (1977) definition

of shape *i.e* when landmark errors are normal and circular, there is no correlation between size and shape (Webster 2010).

In order to remove the effects of position, size and rotation across landmark configurations in a dataset, this study uses a variant of Procrustes Superimposition – Generalised Least Squares (GLS) - based on the methods of Rohlf & Slice (1990). This is a three-step process that removes the effect of position, size and rotation by minimising the sum of the squared distance between corresponding landmark points, resulting in a new dataset for each structure of aligned shape coordinates. As a first step (translation), this calculates the centroid position (mean *x* and mean *y*) of the dataset:

$$\bar{x} = \sum_{i=1}^{m} \frac{x_i}{m}, \quad \bar{y} = \sum_{i=1}^{m} \frac{y_i}{m}$$

Equation 2.5. Calculation of centroid

Where *m* is the number of landmarks in the dataset. By subtracting the mean values of *x* and *y* from each landmark co-ordinate, the configuration can be translated such that the centroid occupies a common origin between configurations (MacLeod 2009a):

$$xtrans_i = x_i - \bar{x}$$
$$ytrans_i = y_i - \bar{y}$$

Equation 2.6. Calculation of common origin

Post-translation scaling (removal of size effects) is accomplished by finding CS (see equation 2.4). Each *x* and *y* co-ordinate value is then divided by the CS value for the configuration as a whole:

$$xscaled_i = \frac{xtrans_i}{CS}$$
$$yscaled_i = \frac{ytrans_i}{CS}$$

Equation 2.7. Scaling

The final step in Procrustes GLS superimposition is alignment of orientation (rotation). Initially, this procedure selects one target configuration (**T**), and one configuration to be rotated (**R**) into an orientation of maximum shape correspondence with respect to (**T**). The optimal angle through which to rotate **R** with respect to **T** is calculated as:

$$\theta = arctan\left(\frac{\sum_{i=1}^{m} y_{Ti}x_{Ri} - x_{Ti}y_{Ti}}{\sum_{i=1}^{m} x_{Ti}x_{Ri} + y_{Ti}y_{Ti}}\right)$$

Equation 2.8. Calculation of rotation angle

Where *m* is the number of landmarks (Sneath 1967, Macleod 2009a). The target (**T**) configuration is initially defined as the first object in the dataset with which to perform the iterative search for the optimal rotational alignment over all dataset configuration. A mean configuration is then calculated for all landmarks and used as **T** to which all landmark configurations are rotationally aligned. This procedure continues until the optimal rotational fit is achieved (Macleod 2009a). The Procrustes distance between two specimens is the sum of the distances between all corresponding landmarks in each configuration.

Procrustes GLS superimposition results in the loss of four degrees of freedom from 2D data (2*m*-4, where *m* = number of landmarks) and seven from 3D data (3*m*-7; Rolf & Slice 1990). The shape co-ordinates produced via Procrustes superimposition are collinear and non-independent, which makes them inappropriate for many statistical analyses that assume no correlation between variables (Polly *et al.* 2013). This may be corrected for by the use of ordination methods such as Principal

Components Analysis (PCA, see below). In addition, the complete separation of size and shape during this process allows these variables to be analysed independently from one another.

2.4.3 Kendall's shape space

The mathematical framework within which geometric transformations and metrics are applied in geometric morphometrics is termed Kendall's shape space (Kendall 1977, 1984). This is a representation of all possible shapes given a landmark configuration and a dimensionality (either 2D or 3D). In this space, the distance between points representing any pair of landmark-defined shapes corresponds to the Procrustes distance between the respective shapes. Kendall's shape space possesses $2k$ - 4 and $3k$ - 7 degrees of freedom for 2D and 3D configurations respectively where k is equal to the number of landmarks. As complex, mutli-dimensional analogues of (non-Euclidean) curved surfaces, the number of landmarks in most configurations will create a shape-space that is extremely difficult to visualize. Therefore, Kendall's shape space is normally visualized with the use of 2D planar triangles that can be represented by the surface of a sphere (Fig 2.9). When a Procrustes GLS superimposition has been applied, the shape space is collapsed into a hemi-sphere as a result of the vertices being rotated to points of maximum correspondence - Slices space (Kendall 1984; Slice 2001). This is due to differences in the treatment of size in the superimpositions that create the shape space; in Kendall's shape space, all specimens are scaled to unit centroid size while in Slice's space, centroid size can vary, which has the effect of minimising the full Procrustes distances between specimens (Kendall 1984; Cooke & Terhune 2015).

Within shape space, every possible configuration of landmark points has a particular locus. In the illustrative analysis of planar triangles, the arrangement of shapes is typically realized as a shape space whose orientation that has an equilateral triangle at the 'north pole' and its antipode mirror image (also an equilateral triangle) is positioned at the 'south pole' (Kendall 1984; Klingenberg 2016). Collinear triangles will then be located along the 'equator' and isosceles triangle will be located around the 'meridians' (Klingenberg 2016; Fig 2.9). The curvature of shape space can cause

distortions when analysing landmark data with standard statistical techniques designed for Euclidean spaces (Macleod 2009b). However, if variability is appropriately small such that that the data only occupies a small section of all potential landmark configurations, then it is possible to project the data points from Kendall's shape space onto a linear tangent space to perform statistical analyses (Dryden & Mardia 1988). Here, shape approximation can be ascertained in the region of the point where two shapes meet. Procrustes scaling to the sample mean configuration can act as this tangent point and is generally considered the best method to use within Kendall's shape space, since it minimizes the distortion caused via the projection of multi-dimensional data into tangent space (Slice 2001; Fig 2.9).

2.4.4 Principle Components Analysis (PCA), Canonical Variates Analysis (CVA) and visualizing shape change

Principle components analysis (PCA) is an exploratory statistical ordination method with several properties that make it particularly useful in the analyses of morphometric datasets with high dimensionality. The method was founded by Pearson (1901), later developed by Hotelling (1933), and has been one of two preferred methods for producing uncorrelated shape variables in morphometric analyses – the other being partial warp analyses (Rohlf 1993; Zelditch *et al.* 2004; Polly *et al.* 2013). Performing a PCA on partial warp scores is Relative Warp Analyses (RWA; Rohlf 1993) and this is equivalent to ordinary shape PCA when

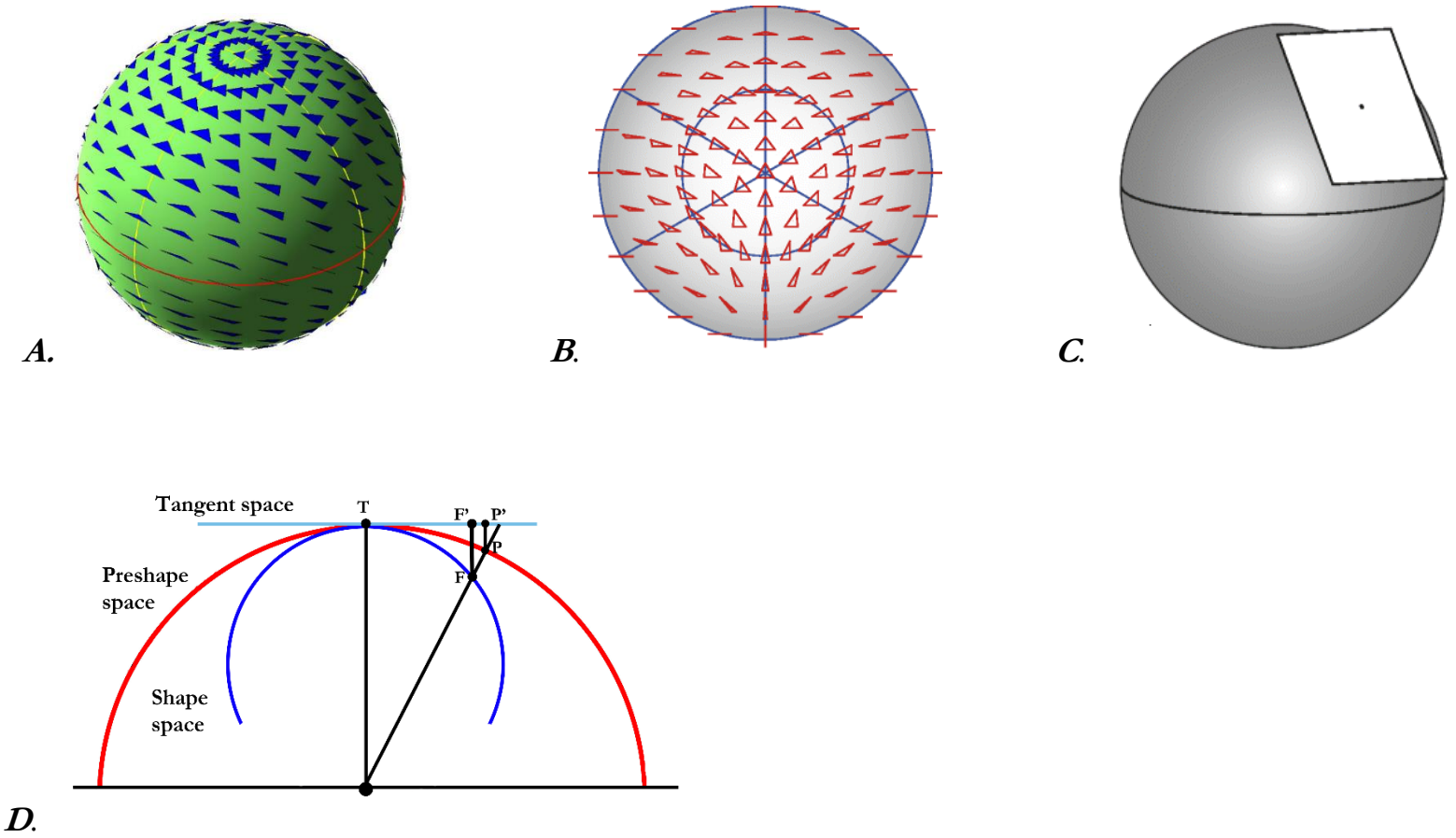


Figure 2.9. Representations of Kendall's Shape Space and Linear Tangent Space. A – Spherical representation of Kendall's shape space using 2D planar triangles. Image retrieved from http://atrouve.perso.math.cnrs.fr/GEF/index.html. **B** – Schematic of one view of triangle shape space showing equilateral triangle at the pole, isosceles triangles at the meridians and 'flat' triangles at the equator. **C** – A plane (tangent space) touching sphere of shape space. **D** – Relationship between Procrustes superimposition and Kendall's shape space. Images B – D from Klingenberg (2016).

principal warps have equal weighting and the uniform component of shape is included. Therefore, PCA is essentially the standard 'shape space' used to visualise shape similarities and differences between objects (Polly 2004; Polly *et al.* 2013). The advantages to using PCA in a geometric morphometric context is that, as the PCs are orthogonal to each other, the initial number of correlated variables (collinear Procrustes coordinates) is reduced to a smaller number of statistically independent variables. Therefore, the original distances between data points and the amount of variation in the dataset as a whole are preserved, while correlations (bar phylogenetic correlations) between variables are removed. This study performs PCA as a singular value decomposition (SVD) of a mean-centered data matrix. Each principal component axis (PC) functions as shape variables, the first of which represents the major axis of variation among objects. Successive PCs, ordinated at right-angles to each other (*e.g.,* an orthogonal geometric relation to one-another) represent successivel axes of variation among objects (Polly *et al.* 2013). Typically, the first three axes will represent the majority of variation in a dataset and therefore show the major similarities and differences between objects.

The following description of PCA calculation is modified from Polly *et al.* (2013) and describes the method used in this study. Jollife (2002) provides a comprehensive review of alternative PCA methods. Calculation of PCA requires four elements: the covariance matrix (**P**), eigenvectors (**U**; which describe the orientation of objects in PCA space), eigenvalues (∑ of **P**; these describe the scaling of objects in PCA space) and the scores (**S**) of the objects on the eigenvalues (Polly *et al.* 2013). The covariance matrix **P** is calculated as:

$$\mathbf{P} = (n-1)^{-1}(\boldsymbol{X} - mean[\boldsymbol{X}])^{T}.(\boldsymbol{X} - mean[\boldsymbol{X}])$$

Equation 2.9. Calculation of a co-variance matrix

Where **X** is a matrix of trait values (the Procrustes superimposed coordinates) and *n* is the number of taxa. This matrix is singular due to the loss of degrees of freedom

during Procrustes alignment (see Section 2.4.2). The eigenvectors of matrix **P** define the PC axes and the eigenvalues give the variance of the objects (individuals represented by constellations of Procrustes superimposed co-ordinates) on each axis. The cosines of the angles of each vector from each of the Procrustes residuals compose the elements of the eigenvectors. Using the SDV algorithm U and ∑ are calculated as:

$$\mathbf{P} = \mathbf{U} \cdot \sum \cdot \mathbf{V^T}$$

Equation 2.10. Calculation of ***U*** *and* $\sum$ *matrices*

Where ∑ is returned as the eigenvalues. The scores, S, are the values of the objects on the PC axes and are calculated by:

$$\mathbf{S} = (\mathbf{X} - mean[\mathbf{X}]) \cdot \mathbf{U}$$

Equation 2.11. Calculation of **S** *matrix*

where *mean*[**X**] is the mean shape calculated by the Procrustes GLS superimposition. This translates the original data to the mean and rotates it to its major axes of variation. Performing PCA on the covariance matrix this way preserves the geometry of shape via maintaining the variance, proportionality of coordinate data and distances between objects.

As this study is principally concerned with between-group discrimination (either between species or climatic periods), the major ordination method employed to investigate alternative grouping hypotheses is canonical variates analysis (CVA). Unlike PCA, this statistical method is directed at identifying the axes along which two or more *a-priori* defined groups can be best discriminated by minimising the within-group variance while maximising between-group variance. Principal components scores representing 95% of the total variation in a dataset were used as the basis for

CVA in all analyses. This allows the major sources of variation to be represented by a small number of composite variables, optimizing the CVA discriminant functions (Macleod 2007a).

Canonical variates analyses (CVA) performs a two-stage rotation and scaling of a subdivided data matrix. The total similarity relations (**T**) within grouped data matrices can be subdivided into 'within-groups' (**W**) and between-groups (**B**). In CVA **T** represents the total sum of squares and cross-products (SSQCP) for all variables (MacLeod 2007a):

$$t_{r,c} = \sum_{j=1}^{k} \sum_{i-1}^{n_j} (x_{i,r,j} - \bar{x}_{i,c,j} - \bar{x}_c)$$

Equation 2.12. Calculation of **T** *matrix*

where *r* and *c* represent the rows and columns of the **T** matrix and *t* is any cell that is occupied by a value *t*. The grand mean is the centre of all pooled samples, therefore **T** is a summary of the dispersion of the total dataset about the grand mean. The 'within-groups' SSQCP matrix **W** has a similar form (MacLeod 2007a):

$$w_{r,c} = \sum_{j=1}^{k} \sum_{i-1}^{n_j} (x_{i,r,j} - \bar{x}_{i,c,j} - \bar{x}_{jc})$$

Equation 2.13. Calculation of **W** *matrix*

where *r* and *c* represent the rows and values of the **W** matrix and *w* is any cell occupied by a value *w*. This matrix represents a summary of the dispersion of each dataset relative to its own group-specific reference. To find the between-groups matrix **B**, which provides a summary of dispersion of the group means from the grand mean, each element of the **W** matrix is subtracted from the corresponding element of the **T** matrix (MacLeod 2007a):

$$\boldsymbol{B = T - W}$$

Equation 2.14. Calculation of **B** *matrix*

The CVA calculations perform a series of integrated data rotations and a scaling that produces a set of axes aligned with the maximum differences between the **B** and **W** matrices. This maximizes the **B:W** ratio by performing a PCA on the $\mathbf{W^{-1}B}$ matrix to calculate the first eigenvector (CV-1). Subsequent eigenvectors represent subdominant modes of variation that each contribute to maximising the **B:W** ratio in decreasing order. These eigenvectors will represent the set of discriminant axes that, together, best achieves maximal separation of group centroid between-group variance and minimal within-group variance (MacLeod 2007a).

To test the equivalence of the group dispersion structure, this study uses a log-likelihood ratio test (logLRT: Manley 1994, Macleod 2007b). The test uses a phi-statistic with a non-parametric chi-square distribution where degrees of freedom are equal to m (k-1) and is calculated as:

$$\phi = [n_t - 1 - 0.5(m + k)]In\left[\frac{|T|}{|W|}\right]$$

Equation 2.15. Calculation of phi-statistic

where n_t is the total number of specimens across all groups ($n^1 + n^2 + \ldots n^i$), m is the number of variables, k is the number of groups, **T** and **W** are the matrices described above and $|\mathbf{T}|$ and $|\mathbf{W}|$ are the determinants of those matrices (which can be thought of as the 'volume' of the matrices; MacLeod (2007a). When $|\mathbf{T}|$ and $|\mathbf{W}|$ are similar, their covariant relationship will likely be similar and, *vice versa,* if their covariance structure is dissimilar, their covariance matrix structures will be likely dissimilar (MacLeod 2007b).

There is a variety of means by which to visualise shape change along PC or CV axes, the most widely-used being the thin-plate spline deformation grid (TPS: Bookstein

1996). Thin plate spline diagrams summarise gross displacements of form in a globally interpolated deformation grid which can make identification of displacement of individual landmarks difficult. This study thus uses strobe models (Lohmann & Schweitzer 1990; MacLeod 2009c) to focus the analysis on specific changes in the positions of the landmarks/semilandmarks themselves. These strobe plots are created with the PC generated eigenvectors, the mean shape for all configurations and model co-ordinates generated at evenly spaced intervals within the ordination space of interest. They represent successive hypothetical models of shape across co-ordinate points in the PCA and/or CVA space. An overlay of these models allows a comparison of the displacement of each particular landmark across the major axes in the ordination space and thus provides a means to examine both shape change in individual landmarks and changes in the configuration as a whole (Macleod 2009c).

Chapter 3: Species differentiation of ancient common vole *Microtus arvalis* and short-tailed field vole *Microtus agrestis* from M_1 morphology and ancient DNA.

3.1. Introduction

3.1.1 Overview

Many mammalian species are recognized as key biostratigraphic indicators in the terrestrial sequences of the Quaternary. They have provided a source of valuable proxy data with which to infer evolutionary, palaeoecological, palaeoclimatic and palaeogeographical change (Lister 1992; Schreve 2001; Currant & Jacobi 2001). In particular, Arvicoline rodents such as voles have become a model group with which to document past environmental conditions, to date archaeological and palaeontological sites and test evolutionary hypotheses (Escudé *et al.* 2013). This is due to *i*) their widespread abundance in faunal assemblages, *ii*) their excellent fossil record and, *iii*) their rapid rate of morphological and genetic change (Chaline *et al.* 1999; Abramson *et al.* 2009). The molar teeth of voles have proved particularly useful in studies addressing patterns of climate-driven evolutionary change and have been utilized in both morphological (*e.g.,* Chaline *et al.* 1999; Roberts & Parfitt 1999; Ledevin *et al.* 2010; Cucchi *et al.* 2014) and also ancient DNA research (Haring *et al.* 2015; Brace *et al.* 2016).

A critical first step in using fossil and sub-fossil material in such investigations is accurate species-level identification. This is crucial if individual species histories are to be reconstructed. A variety of authors have identified anatomical features and/or devised metrics that allow species discrimination to varying degrees of accuracy, predominantly using vole tooth morphology (van der Meulen 1973; Nadachowki 1984; Chaline *et al.* 1999). However, inter-specific similarities, coupled with high intra-specific variation in dental patterns within and among many species of arvicoline rodents (Guthrie 1965; van der Meulen 1973; Nadachowski 1984), can often make

identification difficult. This can be particularly challenging with *Microtus* species where commonly recovered dental elements are undiagnostic to species level. Furthermore, discrete vole species with similar tooth morphologies often have sympatric distributions and show overlap in their ecological requirements (Jurdíkova *et al.* 2000; Santos *et al.* 2001). This scenario may have further been enhanced by dramatic Pleistocene climate change, which brought previously allopatric species into contact with one another. Identification and discrimination of species which show this relationship to one another in the fossil record thus presents a challenge to researchers. The separation of subfossil specimens of the common vole *Microtus arvalis* and the field vole *Microtus agrestis* from isolated specimens of the lower first molar, (henceforth M_1), is one such example where species delimitation has proved particularly problematic (Nadachowki 1984).

The difficulties in discriminating between *Microtus* species can be related to the evolutionary history of the arvicolid sub-family. The Arvicolinae is highly diverse and shows one of the most extraordinary radiations observed within placental mammals in the Northern Hemisphere (Musser & Carleton 2005). Within it, 28 genera containing *c.* 151 species are described (Musser & Carleton 2005). In turn, the *Microtus* genus is one of the most diverse of the arvicolid lineages, characterized by dramatic radiations estimated to have produced *c.* 70 species within the last 1.2-2 myr (Fink *et al.* 2010). This rapid rate of diversification has resulted in the existence of partially-sympatric species, which in many instances have undergone genetic speciation but which often lack conspicuous differences in morphological characters (Markova *et al.* 2012). This evolutionary history can potentially confound morphological species identification where characters are either highly conserved or rapid convergent evolution between species has occurred. High incidences of adaptive convergence have been reported within *Microtus* (Chaline & Graf 1988; Nadachowski & Zagorodnyuk 1996;). In addition, non-heritable environmental components can have a strong influence on the phenotype of a species, which can obscure the phylogenetic signal (Moen *et al.* 2013). Rapid or near-simultaneous divergences can also hinder genetic species differentiation in cases where substitution saturation has occurred

throughout speciation events. This can lead to a weak or absent phylogenetic signal, observed in the data as a polytomy (Halanych & Robinson 1999).

These challenges in species identification are amplified when dealing with ancient *Microtus* remains. This material is often physically damaged and the high level of molecular degradation can be problematic for sequencing good quality ancient DNA. However, the ubiquity of *Microtus* first lower molars (M_1) makes it highly desirable to be able to use this tooth in studies of evolutionary and population level processes. The M_1 has been used extensively and successfully to discriminate between certain *Microtus* species (Nadachowski 1984; Chaline *et al.* 1999). In these instances, existing methods of gross morphological examination and/or simple measurements of the tooth have provided enough taxonomic information to accurately assign species IDs. However, *Microtus agrestis* and *Microtus arvalis* are one of a large number of pairs of *Microtus* species that share near-identical M_1 anatomy and where existing biometric methods may not provide precise taxonomic delimitation. Dental remains from these two species are abundant in the Late Pleistocene – Early Holocene British fossil/subfossil record and, along with other *Microtus* species, often comprise the majority of recorded micromammals (for example Schreve 2011). However, the difficulties associated with accurate identification of the M_1 have led to them often being recorded only as *M. agrestis-arvalis* (Yalden 2010; Schreve 2013; Lebreton 2015*)*, limiting their usefulness in downstream analyses. The sensitivity of small rodents to climatic variability can lead to rapid range shifts, population turnover events, local extinctions and adaptation to novel environmental conditions (Brace *et al.* 2012; Lima-Ribeiro *et al.* 2014; Brace *et al.* 2016; Palkopoulou *et al.* 2016; Smith *et al.* 2016). In specimens from sites that pre-date the British Holocene expansion in agricultural practices, they are also free from the confounding factor of human-mediated effects. This makes them ideal model species with which to examine the effects of abrupt climate change on a British small mammal over the Pleistocene-Holocene boundary, provided accurate species identifications can be made.

One solution to improve morphological identification methods is to assess the applicability of training sets of modern data taken from the M_1. These can be used

to create shape variables that can in turn be used to assign species identification to ancient specimens of unknown ID and also be informative with regards to the main factors influencing tooth shape variability. This approach has been used successfully in previous studies of *Microtus* spp. (*e.g* McGuire 2011). The application of highly sensitive geometric morphometrics and multivariate statistics may also provide greater resolution in this area than has previously been achieved. Additionally, ancient DNA extraction, amplification and Next Generation Sequencing technologies can be used in conjunction with phylogenetic analyses to infer directly taxonomic identification, thereby providing a valuable second line of evidence alongside the morphological data.

With the aims of conducting further analyses of climate-induced effects on *Microtus* specimens from the British Lateglacial – Holocene period (see Chapter 4), this chapter re-examines existing, and generates novel methods in separating *M. agrestis* from *M. arvalis* using the M_1. Any novel approaches to species discrimination generated here will then be applied to a dataset of ancient M_1 material sourced from Gully Cave, a site in South-West England with a faunal age range spanning the Lateglacial (commencing *c.* 14, 700 cal BP) to the Early Holocene (commencing *c.* 11, 500 cal BP; Schreve 2014). This will allow examination of climate driven population-level change/stasis in these species over the Lateglacial-Holocene periods in Britain (see Chapter 4).

The broad aims of this chapter are to:

1. Examine the accuracy of existing biometric methods with which to separate modern *M. agrestis* and *M. arvalis* on the basis on the M_1 anatomy alone
2. Apply 2D landmark and extended eigenshape geometric morphometric methods to modern species separation using the M_1 and evaluate their relative performance both to each other and to biometric results obtained in (1)
3. Identify the main factors involved in tooth shape morphology and quantify the strength of the taxonomic signal
4. Use the generated modern biometric and geometric data as training sets in order to identify ancient M_1 specimens of unknown species ID

5. Use ancient DNA and Next Generation Sequencing (NGS) technology in an attempt to sequence full-length cytochrome *b* sequences from ancient *Microtus* spp. M_1s and identify them to species level on a genetic basis

3.1.2 Phylogeny, distribution and ecology

The common vole and field vole are placed in the subfamily Arvicolinae, a large monophyletic clade within the extremely diverse Superfamily Muroidea (Wilson & Reeder 2005). The Arvicolinae comprise some 140 lineages of voles and lemmings representing *c.* 37 extinct and extant genera and around 143 extant species; although taxonomic revisions are frequent (Carleton & Musser 1984; Chaline *et al.* 1999; Wilson & Reeder 2005; Abramson *et al.* 2009). This group is widespread across the Holarctic, surviving in a wide range of grassland and Arctic habitats. The northern extent of their contemporary range is bounded by the Arctic ocean/ice coverage with a southern distribution limited by the subtropics (Fig 3.1; Wilson & Reeder 2005).

Traditionally taxonomic discrimination within the group has been based on morphological character assessment of the skull, jaw musculature and dentition. The arvicoline rodents show a deepening and shortening of the mandible compared to their cricetine ancestors, thereby providing a stronger bite force (see Repenning 1968 for a detailed description; Chaline *et al.* 1999). Within the Arvicolinae, the broad division between voles and lemmings is generally defined by dental morphology. For example, voles typically exhibit a longer lower incisor (I_1) that crosses from the lingual to the labial side of the molars between the bases of the first and second lower molars (M_1 and M_2; Chaline *et al.* 1999). Lemmings possess a short I_1 that is 'entirely lingual relative to the molars and germinates posteriorly ahead of the lower third molar (M_3) capsule' (Chaline *et al.* 1999 and references therein). Molecular assessment of the relationships within the group has been based largely on the cytochrome *b* (cyt*b*) locus of the mitochondrial genome (Conroy & Cook 1999, 2000; Jaarola *et al.* 2004). This gene has generally performed well at lower taxonomic levels but has led to unresolved topologies among arvicoline genera and among some species of *Microtus,* most probably as the result of site saturation at third codon positions (Jaarola *et al.* 2004;

Gelewski *et al.* 2006). The incorporation of slower-evolving nuclear genes such as exon 1 of the growth hormone receptor (GHR) has improved the deep level phylogenetic relationships within the Arvicolinae (Gelewski *et al.* 2006). The most recent molecular evidence supports the *Microtus* genus as belonging to a clade within the Arvicolini tribe including *Arvicola* and *Chionomys* (Gelewski *et al.* 2006), which is in general agreement with palaeontological evidence (Abramson *et al.* 2009).

Common voles and field voles belong to separate lineages within the genus *Microtus*, which contains several subgenera 'species groups' (Jaarola *et al.* 2004; Table 3.1). At present, there are some 62 -70 *Microtus* species within the genus as a whole (Musser & Carleton 2005; Fink *et al.* 2010) but taxonomic inconsistencies are common and the systematics of the clade are continually under review. All of these species are grassland specialists and survive on a diet consisting largely of grasses and sedges but which varies seasonally and geographically and may also contain a proportion of seeds, grains, roots and bark (Musser & Carleton 2005). The *M. agrestis* lineage is basal in the *Microtus* clade and is not related closely to *M. arvalis* (Martínková & Moravec 2012). Two lineages, *Agricola* and *Microtus* + *Terricola,* form sister clades with two well-supported sub-lineages within them. The common vole *M. arvalis* is nested within the *Microtus* sub-genera while the field vole *M. agrestis* belongs to the *Agricola* clade, which diverged earlier in the history of the genus (Jaarola *et al.* 2004). Both species occupy well-supported positions within their respective groups.

Heckel *et al. (*2005) identified the common vole (*M. arvalis*) as the most abundant European mammal and it has a modern distribution that includes open grass and farmland habitats from the Atlantic coast of France to the Ukraine and central Russia (Mitchell-Jones *et al.* 1999). The field vole (*M. agrestis*) is also abundant throughout Europe, with a distribution stretching from the Atlantic coast to Lake Baikal. The ecological preferences of *M. agrestis* are similar to those of *M. arvalis*, although it tends to prefer damper grassland habitats (Mitchell-Jones *et al.* 1999). It also has a more northern modern distribution, surviving up to the northern-most extremities of

Table 3.1 Subgenera within the *Microtus* genus. Phylogenetic divisions are based on those described by Jaarola *et al.* 2004 using cytochrome b. *Microtus arvalis* (common vole) places within the *Microtus* subgenus and *Microtus agrestis* (short-tailed field vole) places within the *Agricola*.

Genus	Subgenus
Microtus	*Agricola*
	Neodon
	Microtus
	Terricola
	Agricola
	Pallasiinus
	Volemys
	Alexandromys
	Chionomys
	Stenocranius

Scandinavia. The two species have a sympatric distribution within central Eurasia (Fig. 3.1) but are known to partition their use of the environment to minimize interspecific competition, a phenomenon termed 'habitat partitioning' that is commonly observed in sympatric vole species (Jurdíkova *et al.* 2000; Santos *et al.* 2001; Ord & Klomp 2014). While the field vole is a found throughout grassland habitats in Britain the common vole is no longer present on the British mainland and has been presumed absent throughout the Holocene (Yalden 1999). The species is present on Orkney, likely the result of introduction by humans during the Neolithic (Martíková *et al.* 2013; see Chapter 4).

3.1.3 Species discrimination

The limitations of current methods used to separate these two species from the M_1 alone have created problems for researchers. Both are distinguishable using morphological characters when several key anatomical features are considered in combination (Dienske 1969; Nadachowski 1984). They can usually be separated on the basis of their upper second molar (M^2), which has an additional postero-lingual salient angle in *M. agrestis* (the '*agrestis*' loop; Roberts and Parfitt 1999), although this feature only distinguishes *M. agrestis* and its absence is characteristic of other species of *Microtus* beyond *M. arvalis*. The M_1 dental morphology is very similar in both species

and they cannot be separated to a high degree of accuracy by visual inspection of M_1 gross-anatomy alone (Chaline 1999; Roberts & Parfitt 1999). Fedyk & Ruprecht (1971) found that the *M. arvalis* M_1 tends to be smaller and lighter than that of *M. agrestis* and that over 50% of specimens could be identified on the basis of weight-crown height correlations. However, this method has limitations. The proportion of specimens that can be identified is much lower than is desirable. Furthermore, its applicability to palaeontological specimens that are often damaged and/or partially digested, or which have experienced post-depositional diagenesis, is likely to be very limited as accurate weight values will be difficult, if not impossible, to obtain. Therefore, measurements focused solely on the occlusal surface of the tooth, which is often well preserved, are more likely to allow identification of larger numbers of individuals and potentially give more accurate results.

Nadachowski (1984) developed an index, the LT4/LT5 index, based on the observation by Chaline (1971) that the degree of asymmetry between the buccal and lingual enamel triangles is greater in *M. agrestis* than *M. arvalis*. This index alone was found to have little taxonomic value, but proved somewhat useful when plotted against maximum length of the tooth. However, it should be noted that this study was performed with a relatively small sample of modern specimens (n =30 per population) and only with individuals from within Poland (Nadachowski 1984). Nevertheless, such plots have since been adopted as the principal technique for separating these species in archaeological and palaeontological studies (*e.g.* Roberts & Parfitt 1999).

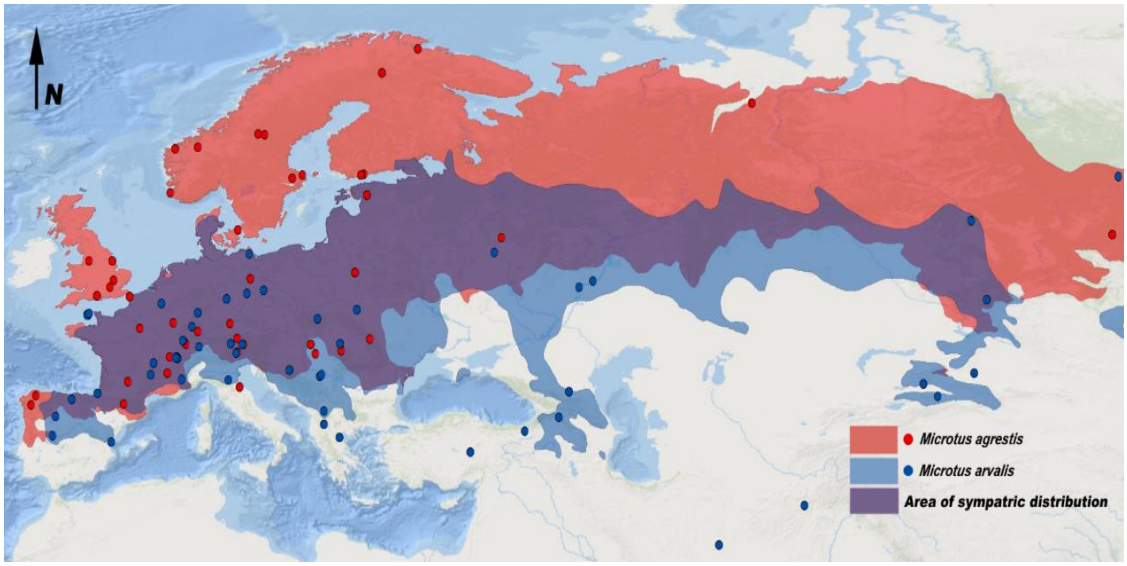


Figure 3.1. Sampling distribution and Eurasian range of field vole *Microtus agrestis* and common vole *Microtus arvalis*. Short-tailed field vole *M. agrestis* range is shown in red and common vole *M. arvalis* is shown in blue. Area of sympatric distribution throughout Central Europe shown in purple. Note isolated populations of *M. arvalis* populations in Turkey and Iran where suitable habitat is available. While the common vole *Microtus arvalis* is present in Orkney, specimens from this region were not included in this study as the population likely represents a Neolithic introduction. *N* of *M. agrestis* = 102; *N* of *M. arvalis* = 124. Basemap Source: ESRI.

An additional confounding factor in identifying *Microtus* spp. from M_1 morphology lies in the amount of *intra*-specific variation relative to the amount of *inter*-specific variation (disparity) between them. High levels of *intra*-specific variation within arvicoline rodents are commonly observed (Nadachowski & Zagorodnyuk 1996; Chaline & Graf 1988; Chaline *et al.* 1999). If this is coupled with low levels of morphological distinction among species, identification can become problematic as true evolutionary relationships are obscured due to morphological overlap. Since a species' phenotype is the sum of its genotype plus environmental effects, one solution is to quantify the strength of the taxonomic signal in each species by partitioning the total amount of shape variation into heritable and environmental components via the quantitative genetics equation:

$$\boldsymbol{P = G + E}$$

Equation 3.1. Components of ***P*** *(phenotype)*

where *P* is the total amount of phenotypic variance, *G* is the genetic (heritable) variance and *E* is the environmental variance. All non-genetic sources are contained in *E,* including epigenetics and all internal and external developmental influences. The left and right M_1s of *Microtus* ssp. are examples of matching symmetry, where an object is present as two discrete copies on either side of the jaw, with a central line of symmetry passing through the mid-sagittal plane across which the genetic component of variation can be assumed to be identical. Thus, an estimate for *E* can be obtained by quantifying the asymmetric component – inconsistent, randomly-distributed differences termed 'fluctuating asymmetry' (Polak 2001). By this means, it is possible to make estimates regarding the extent of environmental effects on tooth shape and to assess, subsequently, whether these effects are likely to influence species identification.

In order to examine questions relating to species separation of *M. agrestis* and *M. arvalis* using morphology, this study uses a geometric morphometric approach. This is the first time this methodology has been applied with the explicit aim of identifying shape variables that accurately separate common and short-tailed field voles. This approach has several advantages over linear distance measurements. Based on the comparison of constellations of aligned shape co-ordinates, geometric morphometrics will allow information capture, pattern recognition and visualization of the geometric form of the lower first molar to a higher degree of sensitivity and completeness than form-factor ratios. When used in combination with multivariate statistics and direct visualisation of geometric transformations through shape models, this method provides a potent tool for the discrimination of related taxa and has been shown to be effective down to the population level (MacLeod & Forey 2002). Furthermore, a multitude of previous studies have demonstrated the efficacy of the geometric morphometric method when applied to molars of other vole species (Navarro *et al.* 2004; Laffont *et al.* 2009; McGuire 2010; Cucchi *et al.* 2014), including studies on species separation (Barčiová 2009; McGuire 2014).

In scenarios where species cannot be easily discriminated on the basis of morphology, or where evolutionary processes and/or environmental influence have obscured the

taxonomic signal in a particular structure, genetic evidence offers an alternative means of species identification. This potentially allows a rapid and highly effective means of inferring species directly from molecular data. Mitochondrial markers have been shown to be extremely accurate at identifying arvicolid species from modern DNA with the mitochondrial protein coding gene cyt *b* one of the most commonly used loci (Jaarola *et al.* 2004; Pfunder *et al.* 2004; Tobe *et al* 2008; Barbosa *et al.* 2013). It has also proved informative at elucidating intraspecific variation and population level processes in both modern and ancient samples (Haynes *et al.* 2003; Hadly *et al.* 2004; Martínková *et al.* 2013).

3.1.4 Aims and research questions

Specific aims addressed in this chapter are:

Morphology:

- To create and compare the relative performance of biometric linear distance measurements and 2D landmark training sets in discriminating between modern *Microtus agrestis* and *Microtus arvalis*
- To test whether significant differences in tooth shape exist between *Microtus agrestis* and *Microtus arvalis* against a null hypothesis of no variation
- To test whether significant size differences in tooth shape exist between *Microtus agrestis* and *Microtus arvalis* against a null hypothesis of no difference
- To test for a relationship between size and shape against a null hypothesis of complete independence between variables
- To quantify the amount of tooth shape variation accounted for by genetic versus environmental effects
- To use the modern training set to assign species identifications to ancient *Microtus* M_1s as yet unattributed to species

Ancient DNA:

- To evaluate methods of identifying species from ancient DNA using Next Generation Sequencing technology and a single mitochondrial locus – cytochrome *b* using:

- Basic Local Alignment Search Algorithms (BLASTN) and
- distance-based v model-based tree building methods with regards to creating accurate species-level phylogenies

3.2 Methods

3.2.1 Morphology

All morphometric analyses were conducted using Mathematica v.10 (Wolfram Research Inc. 2014) with notebooks written by Prof. Norman Macleod. Analyses of allometry and asymmetry were carried out in MorphoJ (Klingenberg, 2011).

Sampling strategy and creating digital images

A training set of modern specimens identified *a priori* as *Microtus agrestis* and *Microtus arvalis* was obtained from the Zoology Collection, Division of Vertebrates, Department of Life Sciences at The Natural History Museum, London. The full dataset represents individuals collected from the mid-19th through to the early 21st century (Appendix B, Table BT1). These were identified using the full suite of soft tissue, dental and skeletal species indicators, thereby giving considerable confidence to the attributions. Specimens were selected on the basis of excellent M_1 anatomical condition and also to represent populations from the full latitudinal and longitudinal range of each species (Fig. 3.1 – Appendix B, Table BT1). Morphological traits may vary at the *intra*-population level across the geographic range of a species. Inclusion of specimens representative of the full latitudinal and longitudinal gradients of a species maximizes the likelihood that all M_1 tooth shape varieties are represented and thus creates a more informative training set.

The left and right M_1 of each adult individual was photographed with a Leica MZ16 FA fluorescence stereomicroscope attached to a Zeiss AxioCam digital camera. For each image, the tooth was mounted on plastozote foam on top of a universal stage. The tooth was orientated so that the occlusal surface was parallel with the focal plane of the microscope. The focus and lighting of each image were optimized to create a clear and crisp outline of the enamel boundary of each tooth. Once each tooth outline

was in focus, a micrometer, positioned at the same distance from the lens as the occlusal surface, was used to calibrate the camera objective and allow insertion of a 1mm scale bar annotation. Left and right datasets are congruent as to the number of specimens used, with 102 specimens of *M. agrestis* and 124 specimens of *M. arvalis* represented in all analyses. As the aim of this investigation is inter-species delimitation and no sexual dimorphism was expected between males and females, sexes were pooled for all analyses.

Dental terminology of the Microtus M_1

The following terminology of the *Microtus* M_1 is based on descriptions by van der Meulen (1973) and illustrated in Figure 3.2. Arvicoline molars are generally high-crowned (hypsodont) with a prismatic geometry and grow continually throughout life. Anatomical descriptions can be subdivided into three main areas: *i*) the three-dimensional features of the crown, *ii*) the relatively flat occlusal surface, and *iii*) the enamel pattern of the occlusal surface. The buccal and lingual sides of the crown are composed of a series of vertical prisms termed *anticlines* (AN), which are alternately interposed with *synclines* (SN). The crown shows enamel-free areas that may assist in diagnostic studies (*M. agrestis* and *M. arvalis* have three; Fig. 3.2). These are vertical strips of dentine originating at the base of the crown and covered with a thin layer of cementum. The M_1 occlusal surface is worn by the grinding action of the jaw, resulting in the prismatic columns becoming truncated and creating a two-dimensional plane of open and closed triangles (T). The number and shape of these triangles varies. *Microtus agrestis* and *M. arvalis* both exhibit seven, the first five (T1-T5) of which are closed, with the sixth and seventh (T6 and T7) being open and largely confluent with the anterior cap dentine (AC). The anterior cap is a rounded region at the front end of the tooth and shows great within-species and among-species variation. Three anatomical forms exist, with *M. agrestis* and *M. arvalis* both of the 'AC3' type which has a well-developed T6 and T7. The entire portion of the tooth comprising all elements anterior to the T3 is referred to as the anteroconid complex (ACC). In the posterior region of the tooth, the talonid-trigonid complex (TCC) is composed of the posterior lobe (PL) and the first three triangles (T1-T3). This area has not been

previously shown to exhibit morphological variation and contains the same elements in all arvicoline species. The enamel pattern of the worn occlusal surface shows a series of lingual and buccal folds that point inwards and outwards in an alternating pattern and are referred to as re-entrant (RA) and salient (SA) angles respectively. Again, only the anterior-most areas of enamel patterning have been observed to exhibit pronounced shape variation.

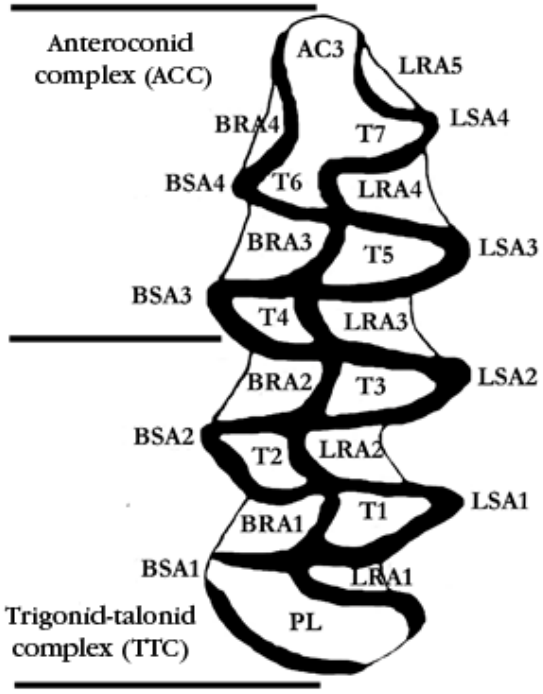


Figure 3.2. Generalised dental anatomy of the *Microtus agrestis - arvalis* left lower first molar (M_1). AC – anterior cap, ACC – anteroconid complex, BSA – Buccal salient angle, BRA – Buccal re-entrant angle, LSA – lingual salient angle, LRA – lingual re-entrant angle, PL – posterior lobe, T1-T7 – open and closed triangles, TCC – trigonid-talonid complex. Illustration by author.

Biometric measurements

Chaline (1971) and van der Meulen (1973) devised a series of metrics based on the symmetry of the buccal and lingual triangles in *Microtus*. Chaline (1971) proposed that *M. agrestis* possesses distinctly asymmetric triangles, with the buccal triangles being smaller than the lingual ones, whereas in *M. arvalis* buccal and lingual triangles are approximately the same size. This led to development of the LT4/LT5 index

(Nadachowski 1984) which, when plotted against length, has been shown to be a somewhat accurate metric with which to discriminate between the two species, albeit with some degree of overlap (Nadachowski 1984; Roberts and Parfitt 1999). Here, length (L) is measured as in van der Meulen (1973), where the maximal length in millimeters (recorded to 2 decimal places, d.p.) of the occlusal surface (L – L') is taken rather than the length of the longitudinal axis, which can be difficult to determine with any degree of accuracy. The lengths of the T4 (LT4) and T5 (LT5) are taken, also in millimeters (to 2 d.p), as the distance from the maximal point of curvature on each triangle to the point of intersection of the T4 and T5 enamel boundary, as in Nadachowski (1984; Figure 3.3). The LT4/LT5 index was calculated as a percentage LT4/LT5 x 100. This index was then plotted against maximal length of the M_1 in order to investigate the accuracy of this method for discriminating between the two species.

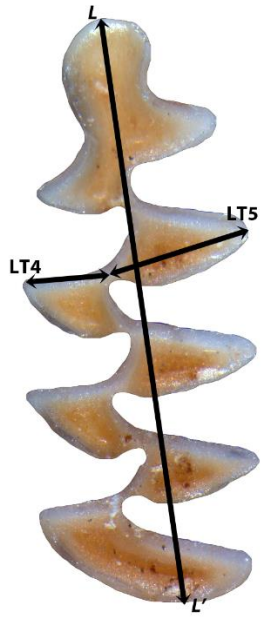


Figure 3.3 Maximal length measurement (L-L') and the LT4 and LT5 measurements

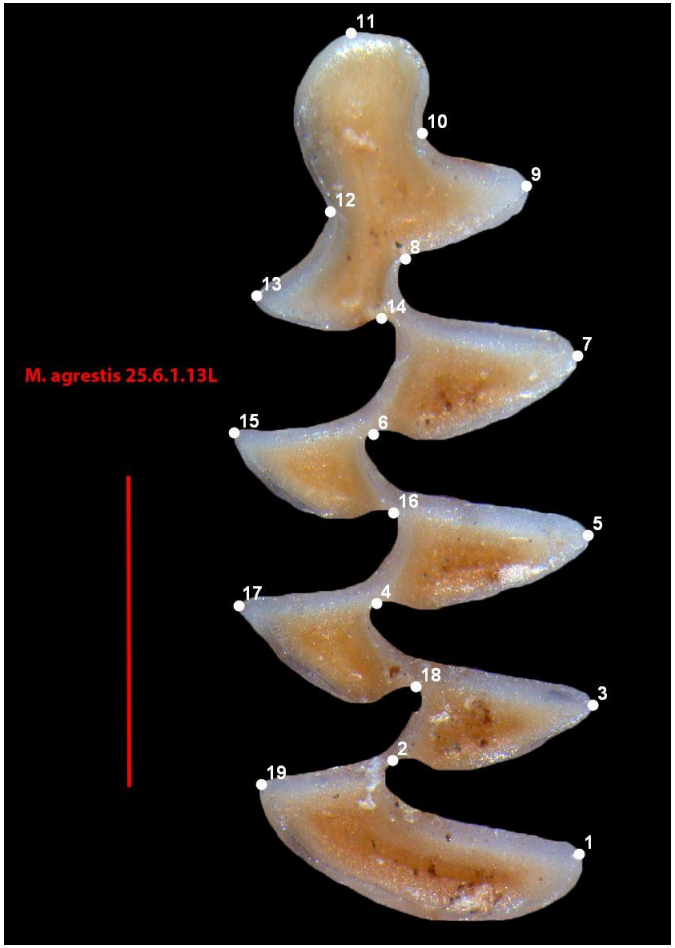
Geometric morphometrics

In order to compare the performance of training sets of increasing complexity, two approaches were compared, a 'simple' configuration of 2D landmarks and extended eigenshape analysis (MacLeod 1999) - a procedure for analyzing a set of outlines represented by landmarks and an interpolated series of semi-landmarks. The latter

approach is expected to offer more information regarding shape change between the two species but is more time-intensive and computationally complex. In order to extract an accurate and standardised size metric, each digital image was scaled prior to extracting landmark co-ordinates via the scale bar annotation included in the image. A series of 19 standard landmark points typically used for analyses of *Microtus* molars (*e.g.,* McGuire 2011; Martínkova *et al.* 2013) was digitised using Media Cybernetics Image-Pro Plus v.1.2 software. These represent the points of maximal curvature of the salient and re-entrant triangles, except for landmark 10, which is located on the point of maximal curvature of the anterior cap (Fig 3.4). These represent Type II landmarks as defined by Bookstein (1991).

Extended eigenshape analysis

The series of curves and angles that define the enamel boundary of the M_1 means that it is particularly suitable for analysis of outline shape variation. Extended eigenshape is a form of semilandmark outline analysis derived from standard eigenshape (Lohmann 1983; MacLeod 1999). As it allows the synthesis of landmarks and inter-landmark points, it has the potential to capture far more shape information on the complex curves and angles of the Microtus M_1 than 2D landmarks and linear distance measurements alone. Extended eigenshape is based on eigendecomposition of pseudolandmark (or semilandmark) co-ordinate points placed along an outline boundary. Unlike standard eigenshape analyses, these are constrained with a series of homologous anchored landmark points located along the semilandmark boundary series. This ensures that artificial variation is not introduced via the digitization process and guarantees geometric (not biological) correspondence of semilandmarks across objects. Another advantage to this method is that it allows differential weighting of each outline segment per its inherent complexity. As each landmark contributes equally to overall variance, placing more landmarks on a segment will increases its contribution to the outcome. This allows each segment to influence the multivariate analysis in direct proportion to the amount of shape variation it exhibits across the sample set (MacLeod 1999; Macleod 2012). The major steps in extended eigenshape analyses are (adapted from MacLeod 1999):



1. Maximal point of curvature on the lingual terminal edge of the posterior loop.

2, 4, 6, 8, 14, 16, 18. Point of maximal curvature of the re-entrant angles at the boundary between enamel and cement.

3, 5, 7, 9, 13, 15, 17, 19. Point of maximal curvature on the salient angles between leading and trailing edges.

10. Maximal point of curvature on the lingual re-entrant angle of the anterior cap.

11. Maximal point of curvature on the most mesial point of the anterior cap.

12. Maximal point of curvature on the buccal re-entrant angle of the anterior cap.

19. Maximal point of curvature on the buccal terminal edge of the posterior loop

Figure 3.4. Anatomical placement of landmarks on the lower left molar (M_1) and their anatomical descriptions

1. Specify a series of standard, homologous, landmark points on the object of interest.

2. Digitise a curve or outline of interest by locating a series of Cartesian co-ordinate points (x,y) along the boundary of the object to act as semilandmarks, starting from a common landmark point (landmark 1).

3. Find *n* evenly spaced coordinate points around the outline boundary using a recursive search to find the minimum number of coordinates which can represent the object to a desired resolution. As more coordinate points are added to the original boundary, the values for the estimated and original object perimeter will converge to within a user-define tolerance criterion (*e.g* 95%, as is specified in this study). The tolerance criterion value indicates the proportion of the total, original, perimeter that is represented by the estimated perimeter (Figure 3.5a). For complexity-weighted extended eigenshape (used here) this method is employed to estimate the minimum resolution required to represent a set of segments between landmark points using differential weighting (Figure 3.5b).

4. Transform the evenly-spaced coordinate points to their shape functions (Φ) by calculating the net angular deviation between adjacent landmark and semilandmark points (see Figure 4 in MacLeod 1999). These express the curves of the object as a series of successive, net angular deviations, from the starting chord.

The resulting phi functions retain the interspecimen size differences via observing the distance between adjacent points but this information is sequestered in downstream analyses to create a size-free dataset.

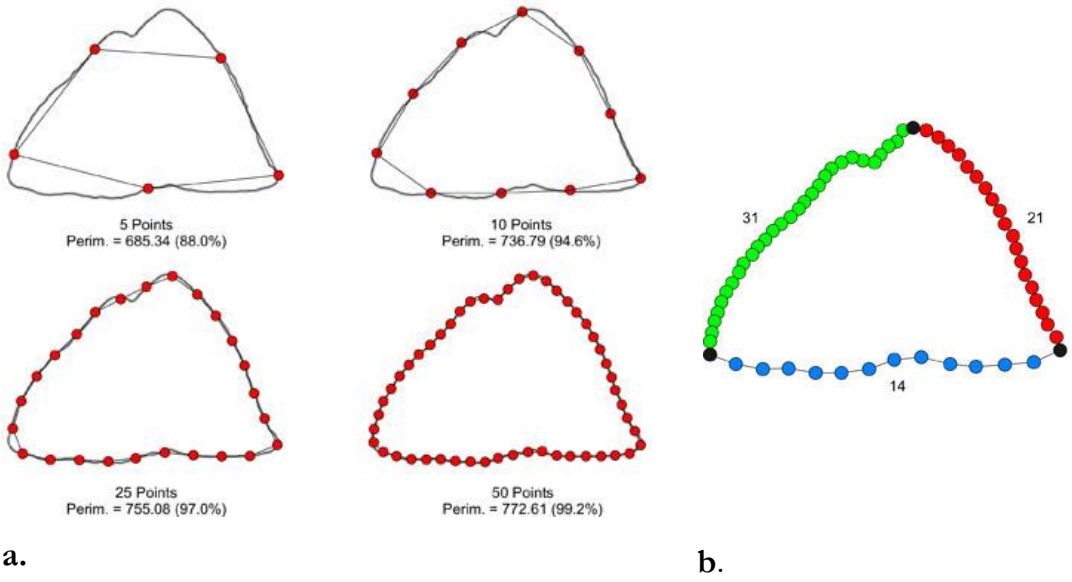


Figure 3.5 a,b. Recursive search for the minimum number of coordinate points required to represent the outline boundary of a planktonic foraminifer species to a variety of tolerance criteria. Shown in (a) is the recursive estimation for the entire outline boundary. Shown in (b) is this method used to determine the minimum number of coordinate points required to represent three homologous segments. Figure from MacLeod (2012).

In this study, a series of nine landmark points located on the tips of the salient and re-entrant angles were used to constrain a starting outline composed of 500 Cartesian coordinate points. The first point represents both the starting point of the outline boundary and the first landmark (Fig 3.6). This resulted in an outline boundary comprising 9 landmarks and 9 segments representing the salient triangles of the lingual and buccal sides, the anterior cap and the posterior lobe of the tooth (Fig 3.6). These were then subjected to an interpolation stage with a 95% tolerance. This ensures each outline segment is represented to a consistent geometric quality resolution and by the same number of semilandmark points per segment, across all specimens within the sample. To achieve this, recursive estimation was applied to weight each segment per its geometric complexity. The number of semi-landmarks allocated to each segment were: segment one - 20, segment two - 18, segment three - 16, segment four -13, segment five – 19, segment six – 14, segment seven – 19, segment eight – 18 and segment nine – 40. These interpolated landmarks were then converted to shape functions for use in downstream PCA and CV analyses. Mean shape was calculated using a Procrustes GLS superimposition. For a full description of the extended eigenshape methodology see MacLeod (1999) and MacLeod (2012).

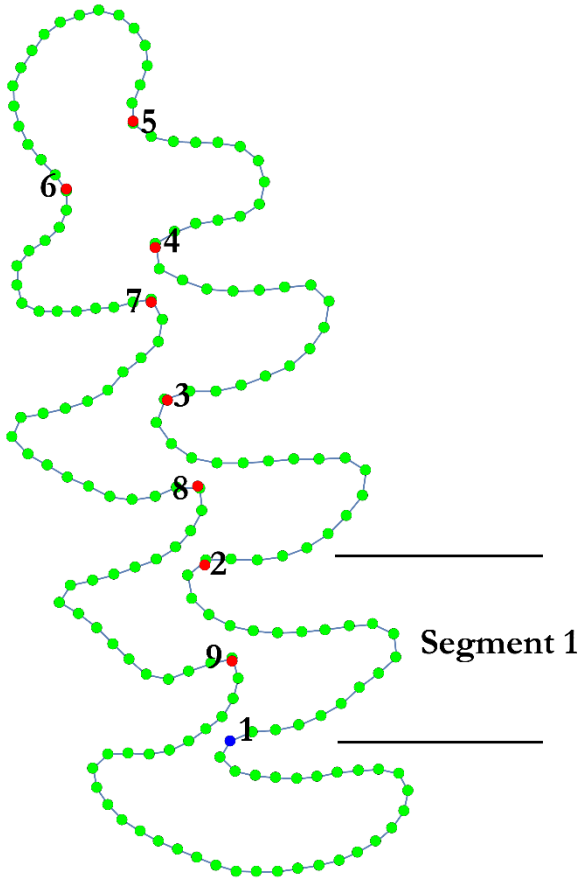


Figure 3.6. Landmarks and semilandmarks used in the extended eigenshape analyses. Representation of the outline of the left *Microtus* M_1 showing the 9 landmark points (starting point shown in blue) used to constrain an interpolated boundary of semi-landmarks broken into 9 segments. Landmarks and outline points run in an anti-clockwise direction (segment 1 indicated).

Multivariate analysis

The 2D landmark co-ordinate data were exported and subjected to Procrustes Generalised Least Squares (GLS) superposition, which removes the effect of position, scale and rotation by minimizing the sum of the squared distance between corresponding landmark points, resulting in a new dataset for each structure of aligned shape coordinates and a set of coordinates representing the mean shape (Rolf & Slice 1990). This also outputs a mean set of aligned shape co-ordinates and a proxy size parameter, centroid size (CS; see Chapter 2 for a description of how this metric is calculated). Phi functions from the extended eigenshape analysis were also subjected to Procrustes GLS to obtain centroid size and mean shape data. As the

operation to convert semilandmark data to their Phi shape functions also corrects for the effects of position, size and rotation, these were used directly in downstream analyses rather than Procrustes shape co-ordinates.

The covariance matrices of the Procrustes aligned shape co-ordinates/Phi shape functions were then subjected to a principal components analysis (PCA), the eigenvectors of which are aligned hierarchically with the major directions of variation of the specimen, set within the multidimensional variable space. Scores of individual specimens projected into this new PCA shape space were then used as the basis for a between-group canonical variates analysis (CVA). Linear distances, roughly analogous to Mahalanobis distances, were calculated from each point in the CVA space to the group centroid using the simple equation:

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Equation 3.2. Calculation of linear distances between points

A distance table and a confusion matrix — which assigns specimens to a group *post hoc*, based on their proximity to the nearest group centroid position in the CVA ordination space — were constructed to estimate the level of group distinctiveness. A log-likelihood ratio test was also carried out to assess the statistical significance of group distinction. Finally, a jackknife misidentification procedure was applied to investigate the stability of the CVA discriminant axes. This procedure removes each specimen in turn, codes it as an "unknown", recomputes the CVA without the "unknown" being present in the training-set data, projects the "unknown" into the resultant CVA space, and then determines if it would be identified correctly in the discrimination analysis. To visualize theoretical shape change, PCA scores were used to generate a series of 5 evenly spaced, along-axes model coordinates for the first 2 principal components and these values combined with the PCA eigenvalues and Procrustes mean shape coordinates, to construct 'strobe' models (MacLeod 2009). These strobe plots represent overlays of hypothetical successive models of M_1 shape

across the five model co-ordinate points in the PCA space and have previously been used successfully to investigate shape variation in various biological structures (*e,g* Lopez-Gutierrez *et al.* 2011).

Fluctuating asymmetry, measurement error and allometry

To allow the partitioning of shape variance into its component parts, each right-side 2D digitisation was flipped prior to adding landmark points so that it was in the same orientation as the left molar. The minimum environmental component of variance was estimated for each species using the 2D landmark dataset only. Shape variance was partitioned into individual (heritable component), side, between sides and measurement error using a Procrustes ANOVA in order to estimate maximum heritability (h^2). This is defined as the proportion of phenotypic variance that is inherited and can be calculated as:

$$h^2 = 1 - (\text{between-sides variance}/\text{total variance})$$

Equation 3.3. Calculation of maximum heritability (h^2)

This is equivalent to $h^2 = G/P$ as between-sides variance is a proxy for E. By Equation 1, $G/P = 1 - (E/P)$. As it is not possible to detect all environmental variation, h^2 can only be considered a maximum estimate and its lower bounds cannot be known without a full estimate for E.

Calculating Measurement Error (ME) - defined as the variability of repeated measurements of a particular character taken on an individual relative to its variability among individuals (Muñoz-Muñoz & Perpiñán 2010) - is an essential component in analyses of landmark data. It is particularly important when shape differences are subtle such as in intraspecific datasets, interspecific datasets between related species with little morphological distinction from each other and when addressing questions of left-right asymmetry (Klingenberg 2011). To calculate ME, ten individuals were repeatedly imaged, landmarked three times and used as replicates in the Procrustes ANOVA analyses.

In its simplest definition, allometry is the statistical association between size and shape (Mosimann 1970). It can be understood as the differential change of a quantitative character that is associated with variation in size and may be produced by a variety of biological phenomena (Seifert 2008; Klingenberg 2016). There are three broad classes of allometry: i) evolutionary allometry which is the result of phylogenetic variation among taxa, ii) ontogenetic allometry due to developmental growth processes and iii), the class that is of interest here, static (or size) allometry, which reflects individual variation within populations and age classes (Gould 1966). Variation in organismal size has been shown to be associated with changes in morphometric variables in many studies of static allometry (*e.g.* Klingenberg 2016 and references therein). Indeed, a nearly ubiquitous property of biological structures is that individuals of different sizes also have different shapes (Adams 2013). This is due to the form of particular structures changing in order to retain their biological function as the biomechanical forces acting on them alter due to changes in size (Seifert 2008).

A strong correlation has been shown to exist between M_1 length and body size in rodents (Martin 1993). Therefore, there may be a strong allometric effect on M_1 tooth shape in *Microtus* species, which could affect the strength of the taxonomic signal, potentially hindering species discrimination. To test for a significant allometric effect on M_1 shape *within* each species, this study uses a multivariate regression of the shape variables, represented by the 2D landmark configurations, on size, represented by a proxy metric - log centroid size (LogCS). MorphoJ uses the regression equation:

$$y = \beta x + e$$

Equation 3.4. Regression Equation

where $\boldsymbol{y}$ represents the random vector of shape variation, $\boldsymbol{x}$ represents the random vector of size, $\boldsymbol{\beta}$ is the matrix of regression coefficients and $\mathbf{e}$ is the random vector of error effects. From this a new variable representing a shape score, S_i, can be defined as:

$$Si = y\beta'(\beta\beta')^{-0.5}$$

Equation 3.5. Calculation of Shape Score, S_i

This score represents the shape variable associated with the shape changes predicted by the regression model, which is equivalent to the shape variable that is most strongly associated with the *i*-th independent (size) variable (Drake & Klingenberg 2008).

Collection and analysis of ancient specimens

Lower first molars of *Microtus agrestis* and/or *arvalis* were obtained from Gully Cave, Ebbor Gorge, Somerset (ST 505508). This site represents a previously unstudied cave-fill sequence that has been excavated at annual intervals by a team from Royal Holloway University of London. The cave stratigraphy is tripartite with the lower and uppermost units characterised by open-framework limestone breccias separated by units of matrix-rich sediment, interpreted as representing two periods of extreme cooling – the Dimlington and Loch Lomond Stadials, separated by a warmer phase of climatic amelioration, the Lateglacial Interstadial. An extensive program of radiocarbon dating and evidence from molluscan and mammalian assemblages has confirmed that the faunal material in the cave spans the Lateglacial through to the Early Holocene (Schreve 2013). Excavation of the cave was carried out in 10 cm spits with bulk samples removed for microfaunal analysis. Sediment was passed through an 8 mm sieve before being removed for wet-sieving and identification.

Right and left lower first molars have been selected from the Gully Cave sequence on the basis of good morphological condition and chosen so as to represent as large a sample size as possible across the three main climatic periods, the Lateglacial interstadial, Younger Dryas and early Holocene. A total of 156 lower first molars, 75 left and 81 right, were recovered from squares F0, F2 and G0 (Tables 3.2a-c). These teeth were digitally photographed and landmarked in the same way as for modern specimens (see above). Co-ordinate point data was treated in the same way as modern data (see above) before the Procrustes GLS superimposition shape co-ordinates/Phi functions were projected into the PCA space defined by the modern training set.

Canonical variates (CVA) and linear distances were then used to assign each unknown specimen to a species according to its proximity to the nearest group centroid, as carried out in the modern dataset.

Table 3.2 a-c. Distribution and number of *Microtus agrestis-arvalis* M1 through the Gully Cave stratigraphic layers.

Square F0			
Depth (cm)	**Left**	**Right**	**Total**
+10 to Datum	0	1	1
0 to -10	4	0	4
-10 to -20	3	3	6
-20 to -30	0	2	2
-30 to -40	2	6	8
-50 to -60	9	8	17
-60 to -70	12	15	27
Total	**30**	**35**	**65**

a. Square F0

Square F2			
Depth (cm)	**Left**	**Right**	**Total**
-120 to -130	22	36	58
Total	**22**	**36**	**58**

b. Square F2

Square G0			
Depth (cm)	**Left**	**Right**	**Total**
+30 to +20	1	1	2
+20 to +10	0	2	2
0 to -10	3	0	3
-10 to -20	6	2	8
-20 to -30	5	1	6
-40 to -60	6	1	7
-50 to -60	2	3	4
Total	**23**	**10**	**33**

c. Square G0

3.2.2 Ancient DNA

Extraction, library builds and NGS sequencing

Taking the results of the morphometric analyses into consideration, 15 lower first molars from the Gully Cave assemblage were selected for ancient DNA analyses. Of these, 14 were categorized as '*arvalis*-like' and 1 was categorized as '*agrestis*-like'. These were prepared for ancient DNA extraction in a dedicated ancient DNA laboratory according to standard protocols described in Chapter 2, Section 2.3.2 and Appendix A Protocol A1. One difference between the extraction preparations of these teeth in comparison to other material was that due to their small size, they were powdered with a pestle and mortar rather than with a drill. Non-disposable plastics were sterilized in a 10% bleach solution before rinsing and being subjected to UV irradiation for 20 minutes between samples. Starting material for the extraction procedure was between 10mg and 15mg which is less than the recommended 40-50mg (Dabney *et al.* 2013). DNA library builds were carried out according to the protocol described in Appendix A Protocol A3 with the exception that DNA extracts were not diluted prior to blunt-end repair, as DNA yields were expected to be low. Libraries were sequenced on an Illumina NextSeq500 platform and raw data was de-multiplexed and adapters trimmed using 'bcltofasta' Illumina conversion software.

Read mapping

Raw paired-end reads were imported into the CLC Workbench software v.8 (CLC Bio-Qiagen, Aarhus, Denmark) where they were paired on import. Sequences were trimmed using quality scores, ambiguity criteria and read length criteria described in Chapter 2, Section 2.3.2.

Determining accurate species identification from shot-gun sequencing without the prior use of capture enrichment is not a straightforward task. In cases where whole mitogenomes or long DNA sequences of the target locus/loci may be present, *de novo* assembly can create contiguous sequences (contigs). These can then be searched using a nucleotide search via Basic Local Alignment Sequence Tool (BLASTN; Altschul *et al.* 1990) against data stored in the National Center for Biotechnology Information

database (NCBI database; Geer *et al.* 2010). This approach has been used effectively in numerous ancient DNA studies (for example Hung *et al.* 2013). However, when the target locus is small (such as for cyt*b* – 1143bp), the number of contigs returned of suitable length to potentially contain the endogenous target sequence can number many thousands. This is prohibitively large to work with. A second approach is to *de novo* assemble the raw reads into contigs and use the assembled contigs as a reference sequence to which the raw reads can be mapped back. Contigs in the size range for the target sequence and which also have large numbers of reads mapping to them may indicate target species DNA. However, this approach also generates large amounts of data and undetected contamination can both create long contigs and generate large numbers of mapped reads, creating misleading results. Nevertheless, both approaches were attempted here. *De novo* assembly did not result in contigs longer than *c.* 6000bp in length for any individual and there were > 10, 000 contigs generated for each individual within a size range consistent with the potential presence of endogenous *Microtus* ssp. cyt*b*. This proved excessively large to work with. When raw reads were assembled to these contigs the viability of working with this data was not improved as many hundreds of contigs had large amounts of reads mapped to them, most likely the result of shot-gun sequence data with short, uninformative reads.

An alternative approach to species identification was therefore undertaken by reducing the stringency of mapping parameters and assembling each dataset to an outgroup judged to be equidistantly related to both *M. agrestis* and *M. arvalis*. Ascertainment bias in assembling shotgun reads to reference sequences can create misleading consensus sequences by preferentially filtering reads that most closely match the reference sequence (Blow *et al.* 2008). Choosing an outgroup equidistant to both species does not by-pass this problem but ensures that any ascertainment bias will be equally applied to both taxa. In this case, a consensus sequence of all published cyt*b* data from *M. oeconomus* was chosen as a reference sequence. This species is placed within the *Pallasiinus/Volemys/Alexandromys* clade, one of the earliest sub genera to diverge in the *Microtus* genus (Jaarola *et al.* 2004). As strict mapping parameters will filter reads so that only short reads with high similarity to *M. oeconomus* will map,

creating a consensus sequence which is likely return a majority of BLAST hits as *M. oeconomus*, parameters were relaxed during read assembly to this species. Mismatch cost was reduced to 1 and gap penalty to 2. Length fraction was reduced from 0.94 to 8 and similarity fraction was reduced from 0.94 to 7. After read assembly, a consensus sequence incorporating low coverage areas and ambiguities was extracted and searched against the BLAST nucleotide rodent database with the first 200 hits being retained.

Consensus sequences of all published cyt*b* sequences for *M. agrestis* and *M. arvalis* were then created. Raw trimmed and merged reads were mapped to the consensus sequences of *M. agrestis* and *M. arvalis* according to which species they had been identified as in the morphometric and BLAST analyses. Mapping parameters were adjusted to more stringent criteria to prevent non-target species reads assembling to the reference; mismatch cost and gap penalty were both set at 2 and length and similarity fraction both raised to 0.94. Where reads gave conflicting base calls that could not be resolved due to the possible influence of sequencing errors at low coverage sites, the base was coded as an ambiguity – 'N'. Consensus sequences with > 90% of the reference covered and which had an average read depth > 4 were then extracted and exported for further analyses. All *Microtus* species have a full-length cyt*b* pseudogene (ψ*cytb*) with has been translocated from the mitochondrial genome to the nuclear genome c. 6 million years ago (DeWoody *et al.* 1999). This gene differs from mitochondrial cyt*b* at 201 (17.6%) of the 1143 sites and so identification of possible ψ*cytb* incorporation into this data was undertaken by comparison of sequences generated here to the ψ*cytb* pseudogene.

Cytochrome *b* sequences for each individual were aligned along codon boundaries with previously published cyt*b* data sourced from Jaarola *et al.* 2004 (Tables 1 & 2 in Jaarola *et al.* 2004) using Mafft to produce an alignment 1143bp in length. This phylogeny was selected over more contemporary phylogenies that use cyt*b*-nuDNA matrices as it has been shown to produce a well resolved tree at lower taxonomic levels using only the cyt*b* locus (Jaarola *et al.* 2004). This allows the rapid construction of a phylogeny that is informative regarding species placement. This phylogeny also contains all the species present in the Gully Cave assemblage belonging to the *Microtus*

genus in addition to *M. agrestis* and *M. arvalis*. These are *Microtus gregalis* (narrow-skulled vole) and *Microtus oeconomus* (northern or tundra vole), two species associated with cold climates. The anatomy of the M_1 makes it highly unlikely that teeth identified at source as *M. agrestis-arvalis* were misidentified and belong to *M. oeconomus*. The former all have 7 M_1 triangles and the latter only has 5. The M_1 tooth morphology in *M. gregalis* is similar to that of *M. agrestis-arvalis* although it does not have a well-developed BRA4 and T6. Again, it is unlikely that *M. gregalis* has been mis-incorporated into this dataset but the presence of this species in the cyt*b* phylogeny will allow a test of possible mis-idenfication. The phylogeny was rooted with *Myodes rutilus* (northern red-backed vole).

Two tree-building methods were used Neighbour Joining (NJ) and Bayesian Markov Chain Monte Carlo (MCMC) analyses, in order to compare their performance (methods described in Chapter 2). The NJ method is quick and allows rapid bootstrapping but may not recover an accurate phylogeny. The Bayesian method is more time-consuming and computationally intensive but is expected to recover an accurate topology due to its increased sensitivity. The appropriate model of nucleotide substitution was determined using JModelTest (Darriba *et al.* 2012). The Akaike, Akaike Corrected and Bayesian Information Criteria (AIC, AICc and BIC) as well as the Decision Theory method all estimated the General Time Reversal (GTR) model with a proportion of invariant sites (pInv = 0.47) and a gamma (G) shape parameter of 0.68 to be the most likely model for this dataset. The NJ tree was constructed in MEGA v. 7.0.18 (Kumar *et al.* 2016) using the Kimura 2-parameter model, which includes transitions and transversions with G set to 0.68. The Bayesian phylogeny was created in MrBayes (Huelsenback *et al.* 2001) on the Cipres portal (Miller *et al.* 2010) using the GRT+I+G parameters suggested by JModelTest (above). Two independent chains ran for $4x10^7$ iterations each, sampling trees and model parameters from the posterior every 4000 iterations. The first 10% of each run was discarded as burn-in. TRACER v.1.5 (Rambaut *et al.* 2014) and Are We There Yet (AWTY: Wilgenbusch *et al.* 2004) were used to check for chain convergence and sufficient sampling before a 50% consensus tree was created. All trees were visualised in FigTree v.1.3 (Rambaut 2009).

3.3 Results

3.3.1 Morphology

Results for left and right teeth were highly congruent and, unless otherwise stated, only results for the right side are reported here.

Biometric analysis

Maximal length (in mm) plotted against the LT4/LT5 index, measured from digital images, is shown in Figure 3.7. While, in general, *Microtus arvalis* occupies the upper left area of this two dimensional ordination space, with *Microtus agrestis* occuping the lower right, there is considerable overlap between the two species. Group assignment using minimal linear distance from each data point to the most proximal group centroid performed better than expected (Table 3.3). The *M. agrestis* M_1 group was generally more distinct than *M. arvalis* with *c.* 78.4% of specimens correctly assigned to their group compared to 73.4% of *M. arvalis* specimens. The range in the LT4/LT5 Index for *M. arvalis* (0.53-0.92 mm) almost completely encompasses the range in the LT4/LT5 Index for *M. agrestis* (0.48-0.88 mm). Similarly, the range of *M. arvalis* length (1.95-3.25 mm) completely encompasses the *M. agrestis* length range (2.24-3.25 mm). A two-tailed Students *t*-test assuming unequal variances confirmed that the mean maximal length of the occlusal surface for these two species differs significantly with *M. agrestis* having, on average, a distinctly longer mean maximal M_1 length than *M. arvalis* (Fig. 3.8: *M. agrestis, n=102,* 2.75mm, *M. arvalis, n=124,* 2.63mm; $p < 0.001$). This result is in accordance with the findings of previous authors (*e.g.,* Nadachowski 1984).

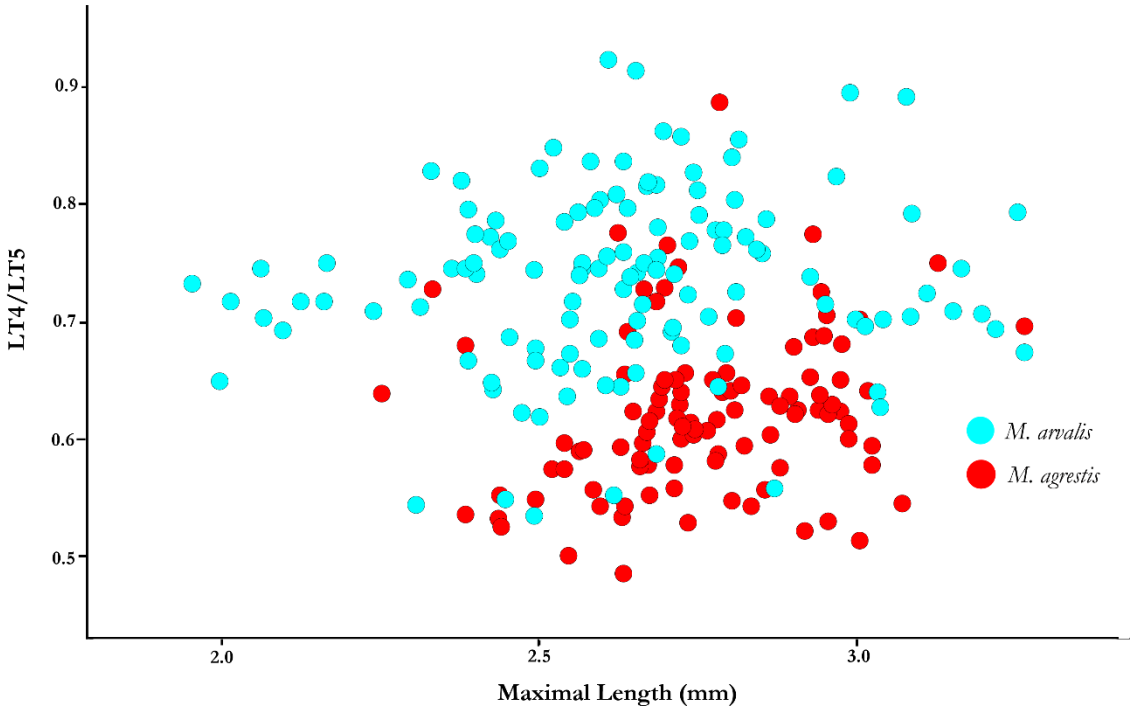


Figure 3.7. Scatter plot showing maximal length (*x*-axis) of the right M_1 plotted against the LT4/LT5 index (*y*-axis). Red – *M. agrestis*, Blue – *M. arvalis*. Note the large region of overlap and presence of many outliers.

Table 3.3. Group assignment of *M. agrestis* and *M. arvalis* based on LT4/LT5 index plotted against maximal length in mm.

Groups	*Microtus agrestis*	*Microtus arvalis*	Total Correct	Group Totals	Percent Correct
Microtus agrestis	80	22	80	102	**78.43**
Microtus arvalis	33	91	91	124	**73.39**
Total Correct	80	91	171	226	**75.66**
Total Estimated	113	113	226		
Percent Estimated Correctly	**70.80**	**80.53**	**75.66**		

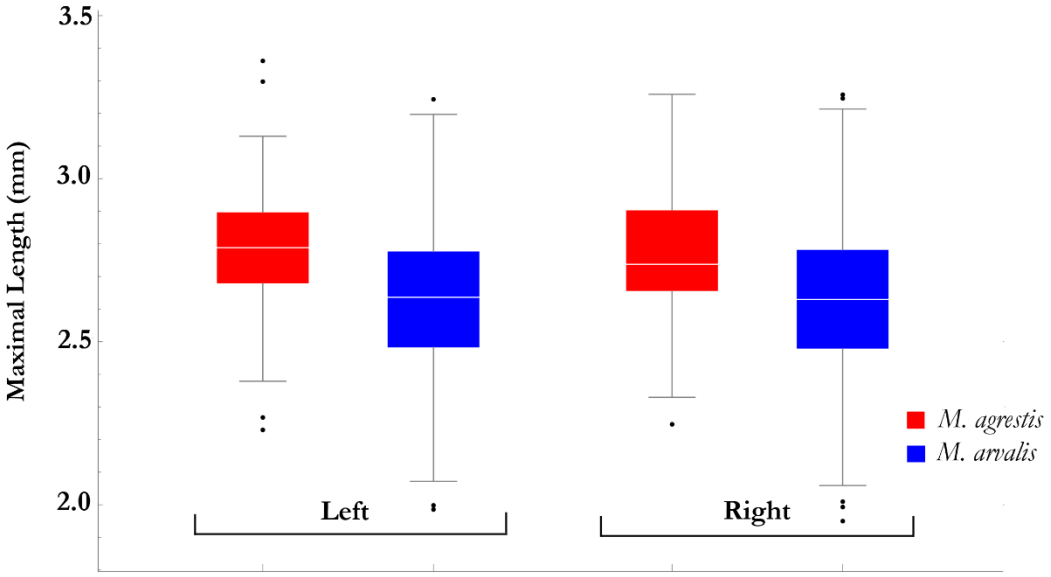


Figure 3.8. Boxplots of maximal length in mm (*y*-axis) of the left and right M_1 for *M. agrestis* (red) and *M. arvalis* (blue).

Geometric morphometrics

2D landmark analysis

Over 30 principal components were generated for the 2D landmark analyses. These each represented only a small fraction of the total variation contained in the data, indicating that this is a high dimensionality dataset (see Figs. BF1 and BF2 in Appendix B). The first two principal components represent 30.81% percent of the total observed landmark shape variation. A scatterplot of these first two components shows that the first axis of these contains information useful in separating out the two species (Fig. 3.9). Observations of the first 35 PC's (which have cumulative variance representing 95% of the information contained in this dataset) showed that only PC1 contained information which discriminated between *M. arvalis* and *M. agrestis* on the basis of M_1 shape. There is a large degree of overlap between each species in the PCA space, confirming that the occlusal surface enamel boundaries in the M_1 of these species (as represented by this set of landmarks) share strong similarity in shape. It should also be noted that this pattern could also be created by portraying a high dimensional dataset in the low-dimensional space of two PC axes that only account for a small amount of the overall variation.

The creation of 'strobe' models – an overlay of landmark points at each of the five model co-ordinates generated in the PCA shape space – illustrates the changes in enamel boundary shape being represented along the first PC axis (Fig 3.10). Only shape changes along PC1 are discussed here as both species occupy a similar shape space along PC2 and all other, subsequent, PCs. Any variation documented by the strobe models along PC2 will be shared by both species and thus not be informative for species discrimination. Along PC1, from the low to high scoring region (moving from the shape space occupied by *M. agrestis* and into the shape space occupied by *M. arvalis*), the lingual triangles (T1, T3 and T5) 'shorten' as the salient angle tips move in a buccal direction and the tips of the lingual re-entrant angles correspondingly move in a more lingual direction. The tip of the lingual salient angle of T7 (LSA4) is located in a more buccal and cranial position in *M. arvalis* than *M. agrestis*. There is little movement in the tips of the buccal salient angles, with only BSA1 and BSA2 moving in a caudal direction and no change in the position of the buccal re-entrant triangles. Interestingly, this lack of shape change in the buccal angles accompanied by a shortening of the tips of the salient and re-entrant lingual triangles across PC1 supports the observation by Chaline (1973) that *M. agrestis* has conspicuously asymmetric buccal and lingual enamel triangles in comparison to *M. arvalis* - whose buccal and lingual triangles are hypothesised to be approximately the same size. This shape-variation trend can be seen by looking at the theoretical shape at model co-ordinate position 1 at the low scoring (*M. agrestis*) region of PC1 compared to the shape of the model at co-ordinate position 5 at the high scoring (*M. arvalis*) region of PC1 (Fig. 3.10). Keeping in mind the degree of overlap between the two species in the PCA shape space, it is worth noting that there are *M. arvalis* individuals that display a greater degree of asymmetry than *M. agrestis* and, *vice versa*, there are *M. agrestis* individuals that exhibit a greater degree of symmetry than *M. arvalis*.

Thirty-five PCs, representing 95% of the variance in this dataset were used as the basis for the group-discrimination approach via CVA. This proved to have greater power in separating out the two species within the confusion matrix (Table 3.4), showing 98.7 percent of training set species could be assigned to the correct group in the CVA space. A log-likelihood ratio test showed between-groups shape differences

as documented by these data to be significant at the $p < 0.001$ level. The jackknife *post-hoc* identification results further supported the robustness of this result with 95% correct group assignment for the training set specimens. Subjecting the log-transformed size parameter centroid size (LogCS) of both left and right datasets to an independent *t*-test assuming unequal variances showed that the mean *Microtus agrestis* M_1 was significantly larger than the M_1 of *Microtus arvalis* (Fig 3.11; left $t = 6.56$, $p < 0.001$, right $t = 5.87$, $p < 0.001$).

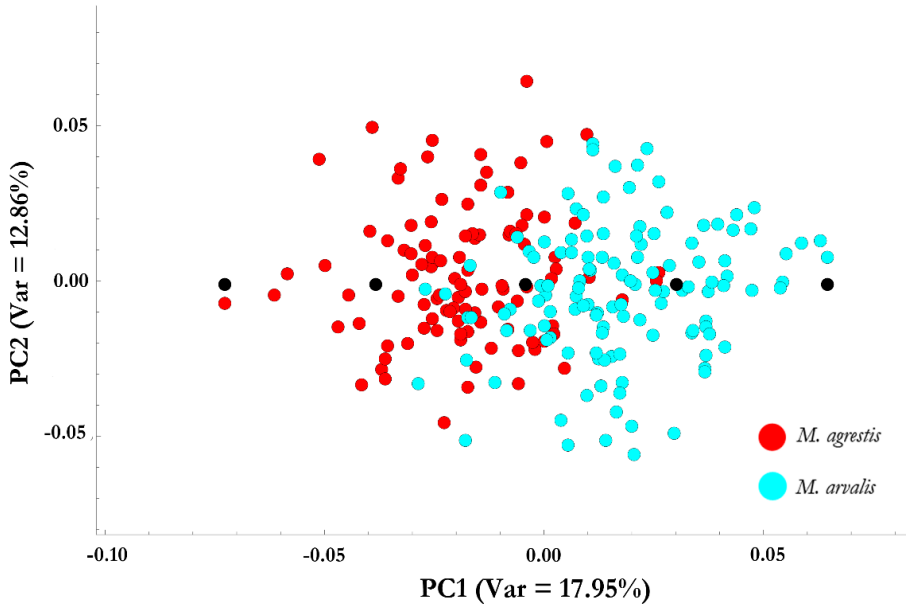


Figure 3.9. Principal components 1 and 2 showing the five model co-ordinates across PC1. Species can only be separated out along the first PC which represents 17.95% of variation in the this dataset, albeit with some degree of overlap.

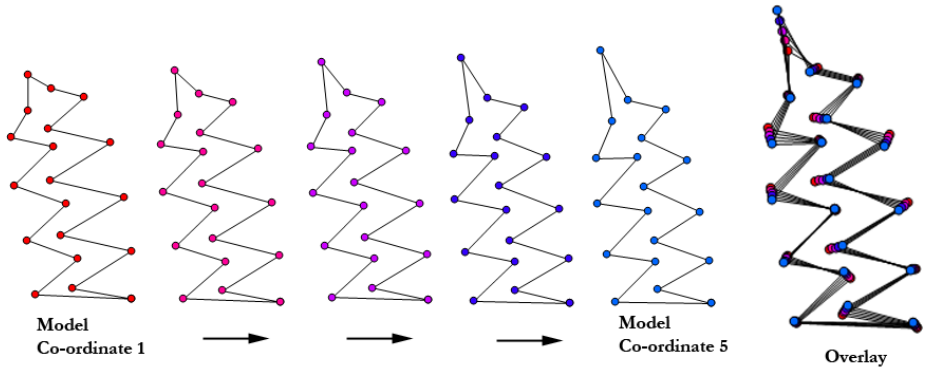


Figure 3.10. M_1 tooth shape as represented by 2D landmark co-ordinates at each model co-ordinate point along PC1 and a strobe overlay of these configurations. Models show an overall heightening of the tip of the anterior cap from the *M. agrestis* space to the *M. arvalis* space with an associated reduction in asymmetry of the buccal and lingual triangles.

Table 3.4. Confusion matrix for the right M_1 showing a total of 98.7% of species in the 2D landmark analysis were assigned to their correct group in the CVA space. Jack-knife values are shown in brackets.

Groups	*Microtus agrestis*	*Microtus arvalis*	Total Correct	Group Totals	Percent Correct
Microtus agrestis	100 *(97)*	2 *(5)*	100 *(97)*	102 *(102)*	**98.04 *(95.1)***
Microtus arvalis	1 *(5)*	123 *(119)*	123 *(119)*	124 *(124)*	**99.19 *(95.96)***
Total Correct	100 *(97)*	123 *(119)*	223 *(215)*	226 *(226)*	**98.67 *(95.13)***
Total Estimated	101 *(102)*	125 *(124)*	226 *(226)*		
Percent Estimated Correctly	**99.01 *(95.1)***	**98.4 *(95.93)***	**98.67 *(95.13)***		

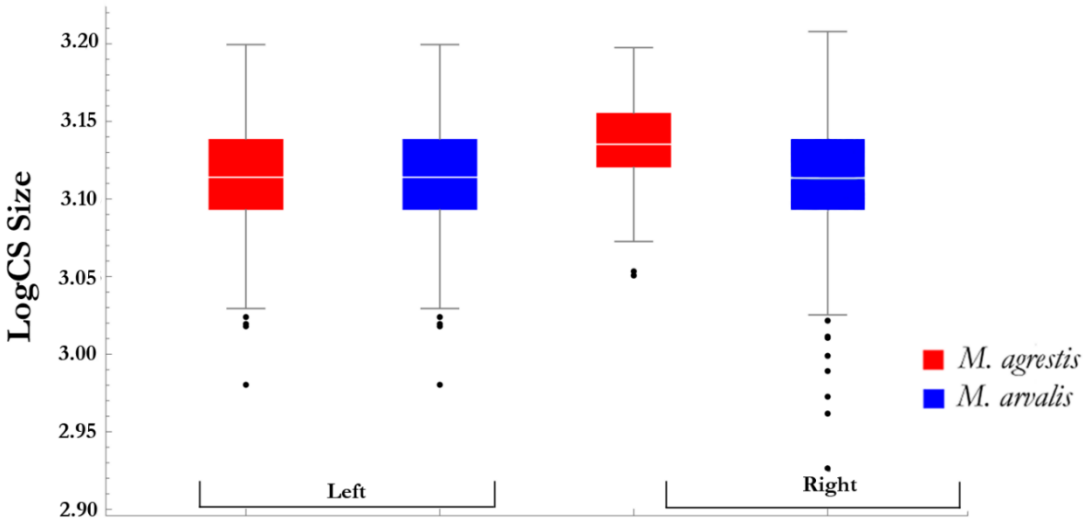


Figure 3.11. Boxplots of logCS size (y axis) for *Microtus agrestis* (red) and *Microtus arvalis* (blue) for the landmark dataset. Note the presence of many outliers in this data.

Extended eigenshape

A large number of principal components (224) were generated for this dataset and, again, each of these typically accounted for only a small amount of the total shape variation (Figs. BF1 and BF2, Appendix B). The first two principal component axes of the extended eigenshape dataset together represent just over 33.61 percent of the

total observed tooth cusp outline shape variation in the dataset. The PCA plot shows a larger degree of overlap across PC1 than that reported for the 2D landmark analysis but shows some separation across PC2 that was not observed in the landmark analyses (Fig 3.12). However, this could also be an artefact of high dimensional data represented in a low dimensional space. The overlay of the complexity-weighted outline segment co-ordinates provides a far more detailed description of the shape variation between the two species (Fig 3.13). Again, the pattern of a 'shortening' of the T1, T2 and T5 triangles is observed between the species, showing that the *M. agrestis* M_1 tends to display a greater degree of asymmetry between the buccal and lingual triangles relative to *M. arvalis*. However, with the additional shape change information provided by the cusp outlines, it can be seen that this division, which exhibited a complex boundary even in the 2D landmark analysis, is not well defined.

This analysis detected additional shape change in the lingual side of the posterior lobe, which becomes narrower in *M. agrestis*, and also in the lingual re-entrant angle 5 and the anterior cap, which become wider and smaller respectively. The general pattern of shape change in the anterior cap from *M. arvalis* to *M. agrestis* is a truncated effect. Along PC2, the anterior cap becomes enlarged and rounder in *M. arvalis* relative to *M. agrestis* and the T6 and T7 triangles show an overall broadening along their posterior edges. The buccal T2 triangle has a more lingually-situated tip and the posterior boundary migrates in an anterior direction, thereby creating a smaller T2 in *M. arvalis* than *M. agrestis*. Despite the increased information on shape variation provided by the extended eigenshape analysis, the CVA group-discrimination did not perform as well as the 2D landmark analysis in separating out the two species, with 93.9% of individuals assigned to the correct group (Table 3.5). The log-likelihood ratio test was significant at the $p < 0.001$ level. Tests for size differences were not performed with this data. Taken together, these results imply that variation in discrete regions alone is not completely diagnostic to separate out the two species. It is more prudent to examine a suite of shape changes in conjunction – namely the degree of asymmetry

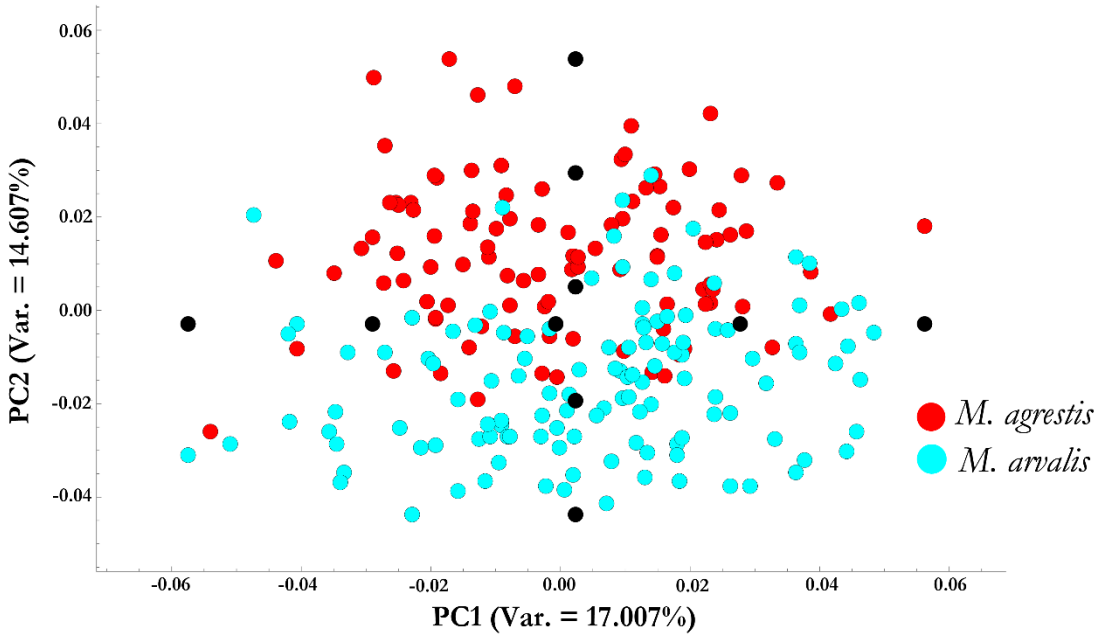


Figure 3.12. Principal components 1 and 2 showing the five model co-ordinates across PC1 and PC2. The first two PC's account for 31.61% of the total variation in the dataset. Model co-ordinates are evenly spaced along the extremities of PC1 and PC2 and represent hypothetical M_1 shape at these locations.

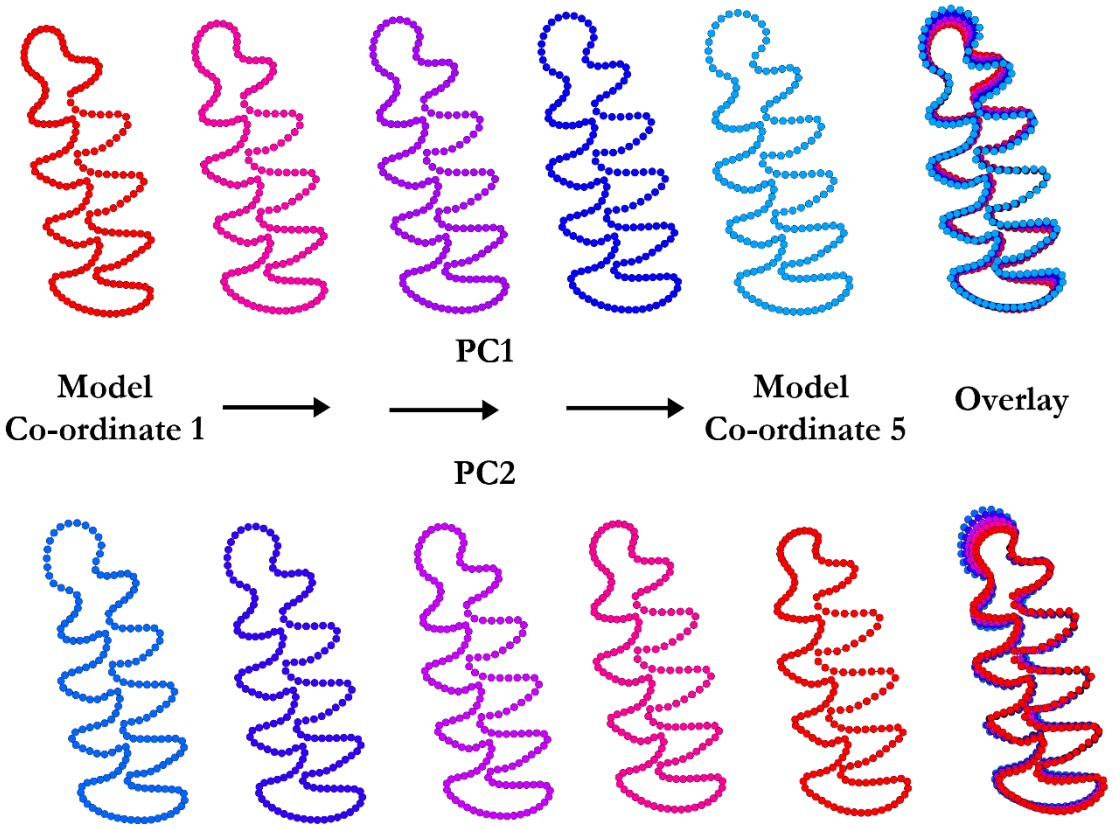


Figure 3.13. Strobe models from the extended eigenshape analysis. Hypothetical M_1 shape at each co-ordinate point along PC1 and PC2 is shown with an overlay of each model.

Table 3.5. Confusion matrix showing a total of 96.88% of species in the extended eigenshape analysis were assigned to their correct group in the CVA space. Jack-knife values shown in brackets.

Groups	***Microtus agrestis***	***Microtus arvalis***	Total Correct	Group Totals	Percent Correct
Microtus agrestis	100 *(100)*	2 *(2)*	100 *(100)*	102 *(102)*	**98.04** ***(98.04)***
Microtus arvalis	5 *(7)*	117 *(115)*	117 *(115)*	122 *(122)*	**95.9** ***(94.26)***
Total Correct	100 *(100)*	117 *(115)*	217 *(215)*	224 *(224)*	**96.88** ***(95.98)***
Total Estimated	105 *(107)*	119 *(117)*	224 *(224)*		
Percent Estimated Correctly	**95.24** ***(93.46)***	**98.32** ***(98.29)***	**96.87** ***(95.96)***		

of the triangles, the relative 'size' and shape of the anterior cap and the shape of the T6, T7 and T2 triangles, which exhibited the most shape variation involved in species separation.

Fluctuating asymmetry, measurement error and allometric effects.

The process of partitioning variation into its components of symmetric variation among individuals uses averages and contrasts of those averages. The averages of left and right landmark configurations represents among-individual variation and the differences between left and right side configurations measures asymmetry (Klingenberg *et al.* 2002). Measurement error is calculated as the differences between replicates of each landmark configuration. The ANOVA analysis partitioned the *Microtus agrestis* and *Microtus arvalis* Procrustes-transformed landmarks into deviations of the original configurations from the overall consensus configuration, identifying discrete components due to *i*) individuals (corresponding to maximum heritability index $h^{2)}$ *ii*) sides and *iii*) individual x sides (fluctuating asymmetry). Partitioning the variation around the symmetric consensus configuration (not shown) indicates that in both species, the variation in symmetric shape among individuals accounts for the largest portion of total shape variation in the M_1 (Tables 3.6 a,b) with maximum h^2 of 0.81 in *M. agrestis* and 0.8 in *M. arvalis*. This suggests that between-individual variation is much greater than variance between left and right teeth within individuals and consequently, that it is unlikely that random environmental effects inflate within-species variation. Measurement error only accounted for 0.91% (*M. agrestis*) and 0.65% (*M. arvalis*) of the total variation and thus can be considered negligible.

Tables 3.6 a - b. Procrustes Analysis of Variance (ANOVA) results for the quantification of left-right asymmetry in lower first molars of *M. agrestis* (a) and *M. arvalis* (b). Degrees of freedom for individual effects are calculated by MorphoJ as: (n - 1) x ($2k$ – 4), where n represents the number of individuals, and k represents number of landmarks (Klingenberg 2002).

a.

Effect	Sum of Squared Deviations (SS)	Mean Square	*df*	F	P	Percentage of Total SS	Proportion - Max h^2
Individual	65.631807	0.01911235	3434	4.49	<0.0001	79.93	**0.80**
Side	1.114023	0.03276539	34	7.7	<0.0001	1.36	
Ind * Side	14.616438	0.00425639	3434	5.83	<0.0002	17.80	
Measurement Error	0.744222	0.00072963	1020			0.91	
Residual	0.0074422	7.2963E-06	1020				

b.

Effect	Sum of Squared Deviations (SS)	Mean Square	*df*	F	*P*	Percentage of Total SS	Proportion - Max h^2
Individual	72.962485	0.1744679	4182	4.72	<0.0001	81.38	**0.81**
Side	0.656359	0.01930467	34	5.23	<0.0001	0.73	
Ind * Side	15.446995	0.0369369	4182	6.42	<0.0001	17.23	
Measurement Error	0.586836	0.00057533	1020			0.65	
Residual	0.00586836	5.7533E-06	1020				

To test for the allometric effects on tooth shape, a multiple regression was carried out between centroid size and the shape components within each species as represented by the 2D landmark configuration. As centroid size was not normally distributed, these values were log-normalized prior to analysis, which also resulted in an improved linear relationship. Significance levels were computed using a permutation test that randomly reassigns observations for the dependent variable to observations for the independent variable to test the null hypothesis of complete independence between variables. For both species, size had a significant, but small, effect (Fig 3.14, a & b). Allometry only accounted for 1.59 percent (left) and 1.78 percent (right) of total shape variation in *M. agrestis* ($n = 102$, $p = 0.05$ for the left M_1 and $p = 0.03$ for the right M_1). For *M. arvalis* allometry accounted for 3.14 percent in the left tooth and 2.08 in the right ($n = 124$, $p < 0.001$ for both sides). It is therefore not considered to be a strong predictor of M_1 shape as represented by the 2D landmark dataset.

Data projection of ancient specimens

As the 2D landmark analyses performed as well as the extended eigenshape with regards to group separation, this shape space was used for projection of the shape variables from the fossil specimens. Using only 2D landmarks also allowed for inclusion of a larger number of fossil specimens as these did not have to have the entire enamel boundary preserved. Figures 3.15 and 3.16 show PCA plots of the projected "unknown" data from Gully Cave into the PCA space defined by the modern, identified dataset for both left and right molars. PC1 and PC2 account for 27.8% of total tooth shape variation in the left M_1 and 30.78% in the right. PC1, PC2 and PC3 together account for *c.* 39% of total tooth shape variation in both datasets. Visual inspection of these plots shows that the majority of the "unknown" data plots in the PCA space defined by *M. arvalis*. Canonical variates analysis (CVA) and linear distances to group centroids (Appendix B, Table BT2) support this conclusion, with 65 left M_1s of unknown ID being assigned as *M. arvalis*-like and only 10 being assigned as *M. agrestis*-like. A similar result was returned for the right M_1 with 62 specimens assigned as *M. arvalis*-like and 19 being assigned as *M. agrestis*-like (Tables 3.7 and 3.8).

These results suggest that the common vole existed in larger numbers at this site than the field vole. The unknown specimens were then assigned to a broad climatic period, either Lateglacial Interstadial (warm), Younger Dryas (cold) or Holocene (warm), based on indirect evidence from radiocarbon dating of other Gully Cave faunal remains and molluscan assemblages (Table 3.9 a, b). This showed that both species were present at this site throughout these climatic periods, suggesting population continuity over the Younger Dryas cold phase. Of particular significance is the presence of *M. arvalis* in the Holocene cave deposits. This the first time this species has been recorded in the British mainland in the Holocene (compare with Yalden 2010). Accordingly, these data, and the specimens from which they are derived, represent the first record of the common vole, *M. arvalis*, in mainland Holocene Britain.

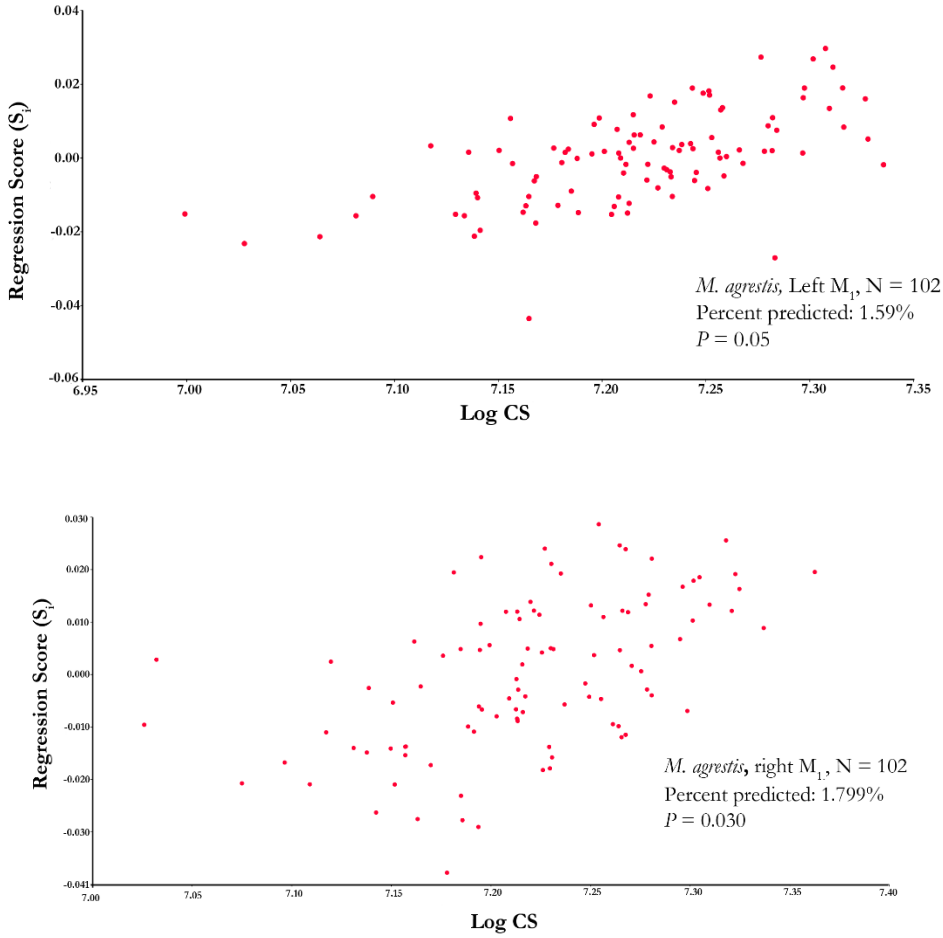


Figure 3.14 (a). Results of a multivariate regression of LogCS on shape for the lower first molar in *M. agrestis* left (above) and right (below).

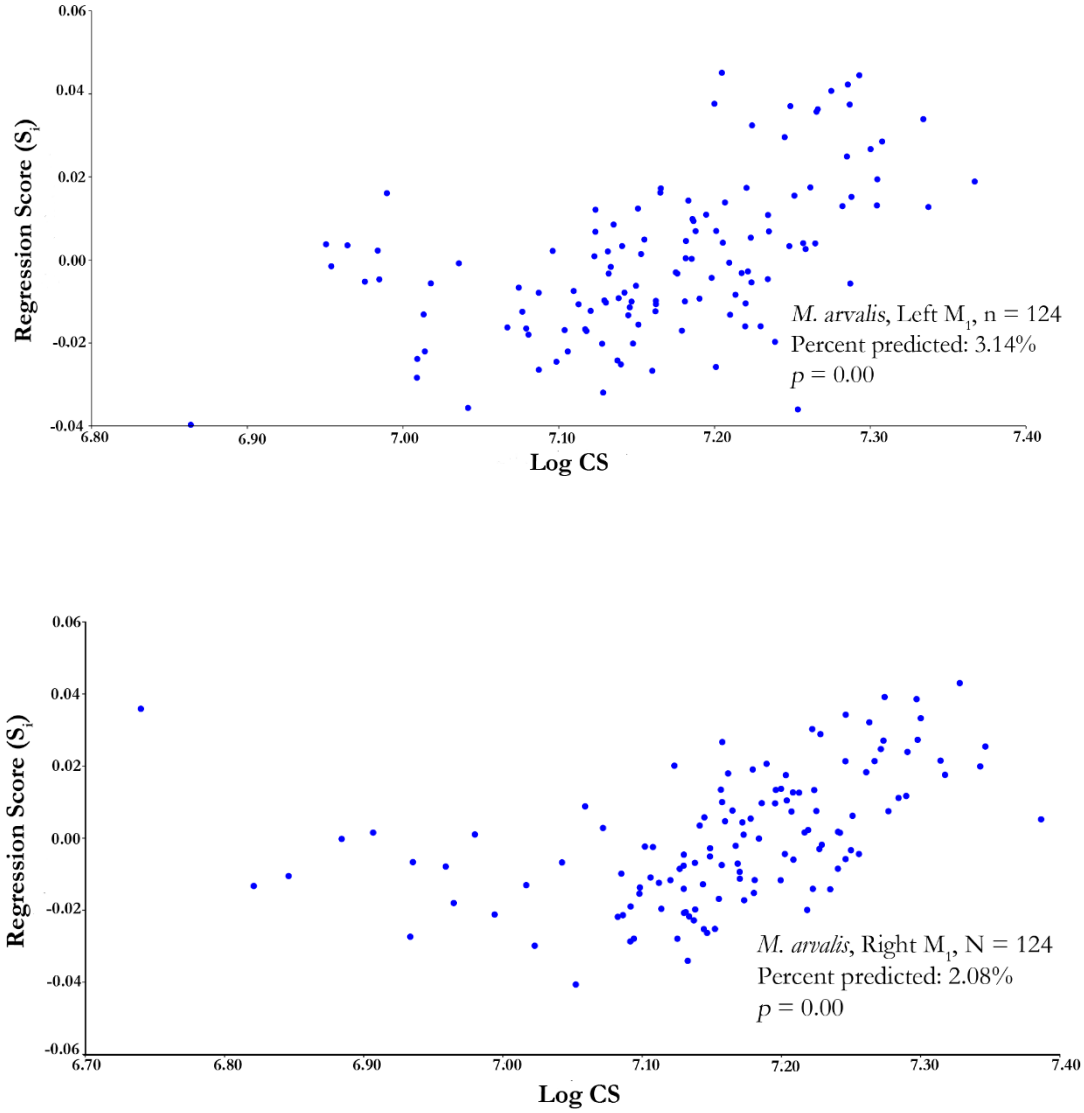


Figure 3.14 (b). Results of a multivariate regression of LogCS on shape for the lower first molar in *M. arvalis* left (above) and right (below).

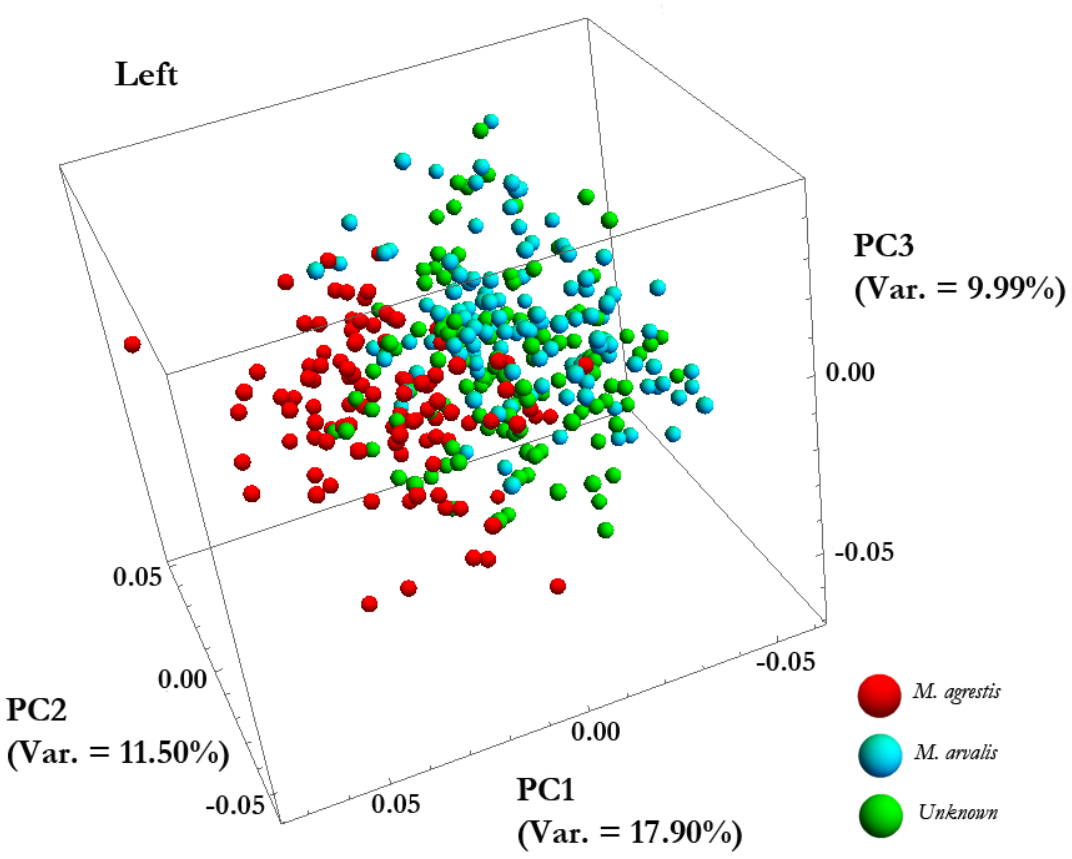


Figure 3.15. Principle components plot showing the unidentified specimens from Gully Cave projected into the space defined by the modern M_1 variables for the left M_1.

Table 3.7. Raw count for ancient specimen assignment in the left dataset.

	Microtus agrestis	*Microtus arvalis*	Total
Raw Count	10	65	75
Percentage	13.33	86.67	100

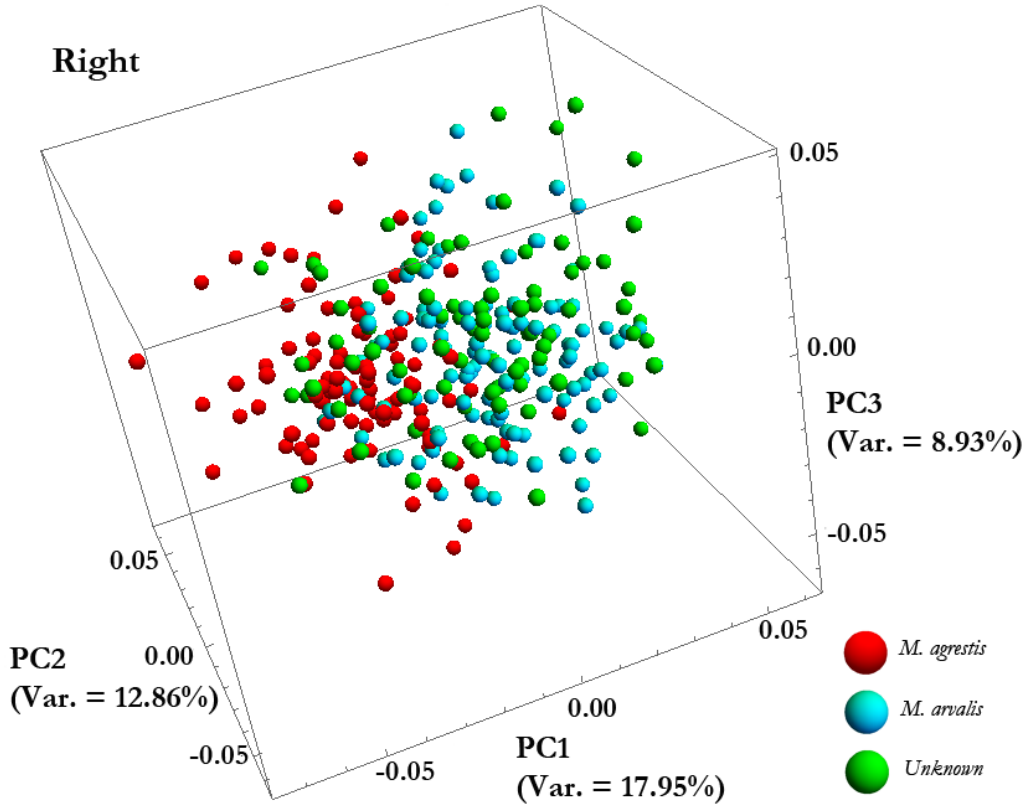


Figure 3.16. Principle components plot showing the unidentified specimens from Gully Cave projected into the space defined by the modern M_1 variables for the right M_1.

Table 3.8. Raw count for ancient specimen assignment in the right dataset.

	Microtus agrestis	*Microtus arvalis*	Total
Raw Count	19	62	81
Percentage	23.46	76.54	100

Table 3.9 a, b. Specimen numbers for each species of the Gully Cave climatic periods.

Microtus agrestis					
Left			**Right**		
Holocene	Younger Dryas	Lateglacial	Holocene	Younger Dryas	Lateglacial
3	6	1	5	7	7
	Total	**10**		**Total**	**19**

a. M. agrestis

Microtus arvalis					
Left			**Right**		
Holocene	Younger Dryas	Lateglacial	Holocene	Younger Dryas	Lateglacial
15	29	21	11	22	29
	Total	**65**		**Total**	**62**

b. M. arvalis

3.3.2 Ancient DNA

Out of the 15 M_1 teeth used in this analysis, 10 yielded full-length (1,143bp) ancient cyt*b* DNA sequences of sufficient quality and quantity to be included in further analyses. None of these contained the characteristic mutations associated with ψ*cytb*. Average coverage, number of reads mapped and fraction of the reference sequence covered are given in Table 3.10. The final dataset consisted of 116 cyt*b* taxa, of which 10 were ancient sequences from Gully Cave. One of these (MM104) was identified as *M. agrestis*-like in the morphometric analyses and nine (MM099, MM100, MM105, MM109, MM110, MM112, MM113, MM114, MM115) were identified as having an M_1 morphology most similar to *M. arvalis*. Summary statistics were calculated in DNAsp v. 5.10 (Rozas 2010). For the pooled dataset, there were 125 sites with gaps or missing data, 540 monomorphic (invariable) sites and 478 variable (polymorphic) sites, of which 406 were parsimony informative. The aim of these analyses was not to assess phylogeography or population processes but rather, to investigate the performance of BLAST searches and tree-building methods in identifying ancient DNA sequences of cyt*b* quickly and accurately to species level. Therefore, summary statistics such as nucleotide diversity, haplotype diversity and pairwise genetic

distances were not generated here. These will be discussed in further detail for *M. arvalis* in Chapter 4.

Table 3.10. Illumina read statistics from Gully Cave Microtus spp. aligned to a consensus cyt*b* reference sequence. Statistics for MM104 are from mapping to a consensus sequence of field vole cyt*b*.

ID	Avg. read length after trim	Number of reads mapped to reference cyt*b* sequence	Average coverage
MM099	55.62	88	3.96
MM100	59.11	73	1.92
MM104	57.24	107	3.69
MM105	54.24	147	4.95
MM109	57.57	129	4.57
MM110	57.09	196	9.5
MM112	49.42	113	3.24
MM113	48.87	125	3.78
MM114	60.07	97	5.02
MM115	56.73	103	5.05

The first step in the analyses was to set up a series of BLASTN query searches to examine which Genbank sequence data had local sequence similarities to the submitted queried sequences. This search was performed with a consensus sequence constructed from mapping trimmed and merged Illumina paired end reads to the consensus sequence of an outgroup taxa, *Microtus oeconomus*. The first 200 hits were retained from each search and in every case these 200 hits were in agreement with the species identification obtained in the morphometric analyses, *i.e* those specimens identified as *M. arvalis* all returned *M. arvalis* hits and the specimen identified as *M. agrestis* returned only *M. agrestis* hits. BLAST E-numbers were all < 1e-20 and percent identities were all > 99%.

Figures 3.17 and 3.18 show the Neighbour Joining tree and the Bayesian MCMC tree used to reconstruct the *Microtus* phylogeny with the target species sequences. The NJ tree did not recover the subgenera phylogeny within genus *Microtus* as shown in Jaarola *et al.* (2004). *Pallasiinus*, *Volemys* and *Alexandromys* were incorrectly placed

towards the base of the tree, diverging earlier than to *Terricola*, *Agricola*, *Microtus* and the Nearctic species. It did, however, recover the correct phylogenetic relationship between *Terricola*, *Microtus* and the Nearctic clades. However, the *Agricola* clade was incorrectly placed as a sister group to *Microtus*. Bootstrap support in deeper levels of the phylogeny was poor and the topology is not well resolved. All of the *Microtus* spp. cyt*b* sequences generated in this study were placed in species groups that were in agreement both with the morphological analyses and with the BLAST search. Specimen MM104 was placed with high bootstrap support with the four *M. agrestis* sequences from Genbank within the well-supported *Agricola* sub-lineage. The branch leading to this specimen is long, possibly indicative of a large amount of genetic divergence between this ancient cyt*b* sequence to modern cyt*b*. Alternatively, this could be due to *i*) a number of ambiguous base calls and missing data for this individual (23 bases) or *ii*) an artifact of the inclusion of erroneous base calls due to the combined effects of sequencing error and ancient DNA damage. As the translation of this sequence did indicate any errors in base calls it is more likely that missing data and ambiguous calls affected the branch length. All nine of the specimens identified as *M. arvalis*-like, MM110, MM113, MM105, MM114, MM115, MM100, MM109, MM112 and MM099 (in order of appearance in the tree) grouped together as a sister group to modern *M. arvalis* cyt*b* sequences. The placement has high bootstrap support, both for the placement of the ancient sequences and for the *M. arvalis* clade as a whole.

The Bayesian tree performed noticeably better at reconstructing the deep level topology of the *Microtus* genus. The *Agricola* genus is placed outside the sister clades of *Terricola* and *Microtus* with the Nearctic, *Pallasiinus*/*Volemys*/*Alexandromys* and *Chionomys* clade all forming monotypic lineages at increasing distance from *Microtus* + *Terricola*. Deep level Posterior Probability support was typically higher than for bootstrap values, with the branches leading to the inner lineages all showing probability values between 0.9 – 1. The exception to this is the placement of *Chionomys* and the branch leading to *Agricola* and *Terricola* + *Microtus*, which have probability values of 0.54 and 0.68 respectively. Lower level taxonomic support was between 0.9 and 1 in nearly all cases. All of the Gully Cave samples in the Bayesian MCMC tree

had phylogenetic positions that fully agreed with both the morphological results, the BLAST results and the placement of these taxa in the NJ tree. Sample MM104 had a long branch length and was again placed in the well-supported *Agricola* clade with the four modern *M. agrestis* cyt*b* sequences. The nine *M. arvalis* samples also formed a discrete group with the *M. arvalis* sub-lineage and this was supported with a posterior probability of 1.

The question over whether *M. agrestis* – *arvalis* M_1s could have been misidentified as the narrow-skulled vole *M. gregalis* was not supported in either the BLAST analysis or the phylogenetic tree construction. This taxon did not feature in any of the returned BLAST hits. It was placed as a poorly-supported sister taxa to *Chionomys* in the NJ tree and distant from all other lineages in a basal position in the Bayesian tree. This taxon has been previously described as 'a phylogenetic enigma' (Jaarola *et al.* 2004) and has shown inconsistent placement in a variety of studies (Conroy & Cook 2000; Abramson *et al.* 2009). It has not, however, been regarded as a close relative of either *M. agrestis* or *M. arvalis* and any error in identification would have been easily illustrated with the cyt*b* sequences used in this study. The absence of such a pattern strongly indicates that none of the teeth sampled for ancient DNA were incorrectly identified at the sorting stage.

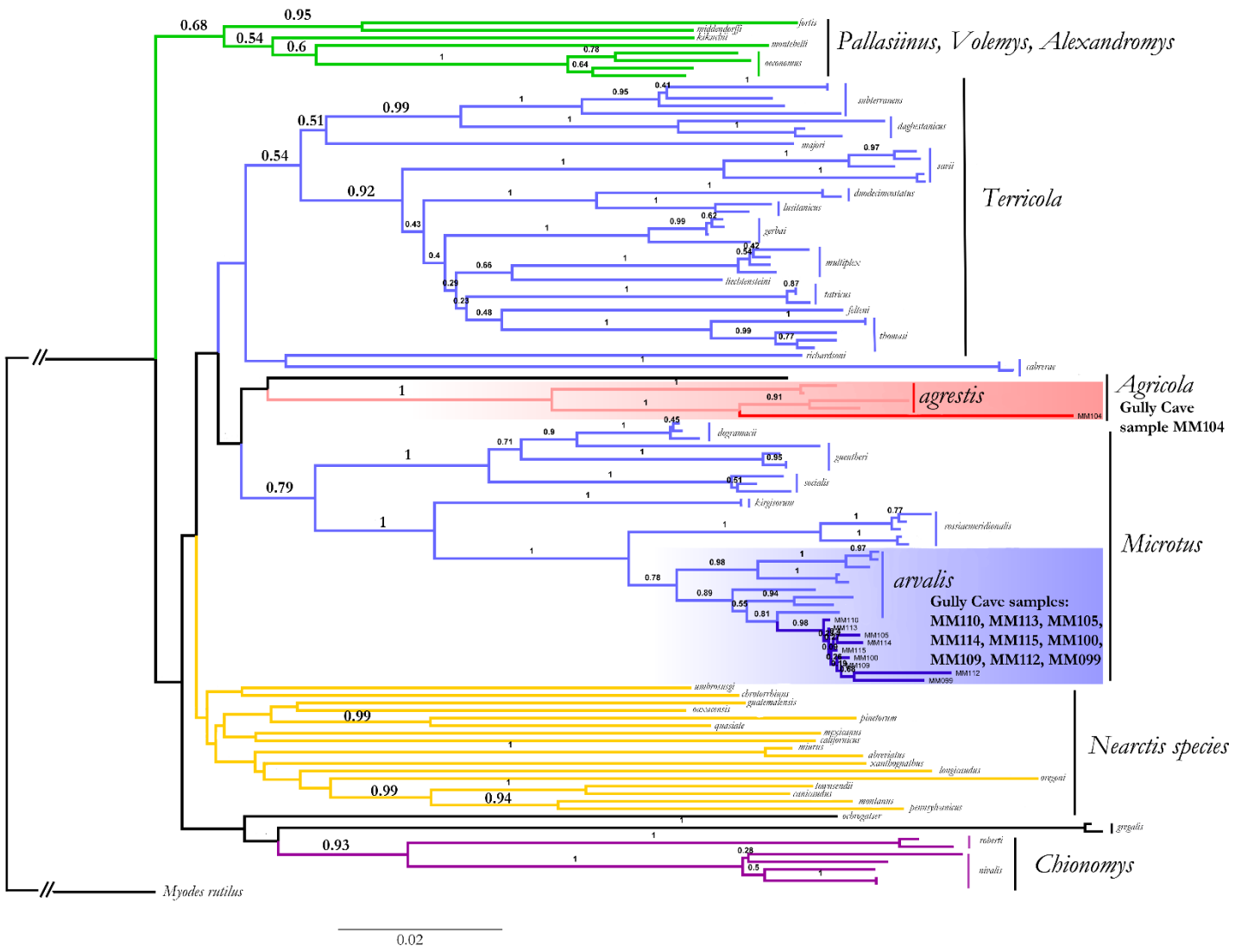


Figure 3.17. Neighbour Joining (NJ) tree of all cytochrome *b* haplotypes from Jaarola *et al.* (2004) and *the Microtus agrestis - arvalis* sequences from the Gully Cave - assemblage. Rooted with *Myodes rutilus*. Number above branches denote bootstrap support from 1000 replicates.

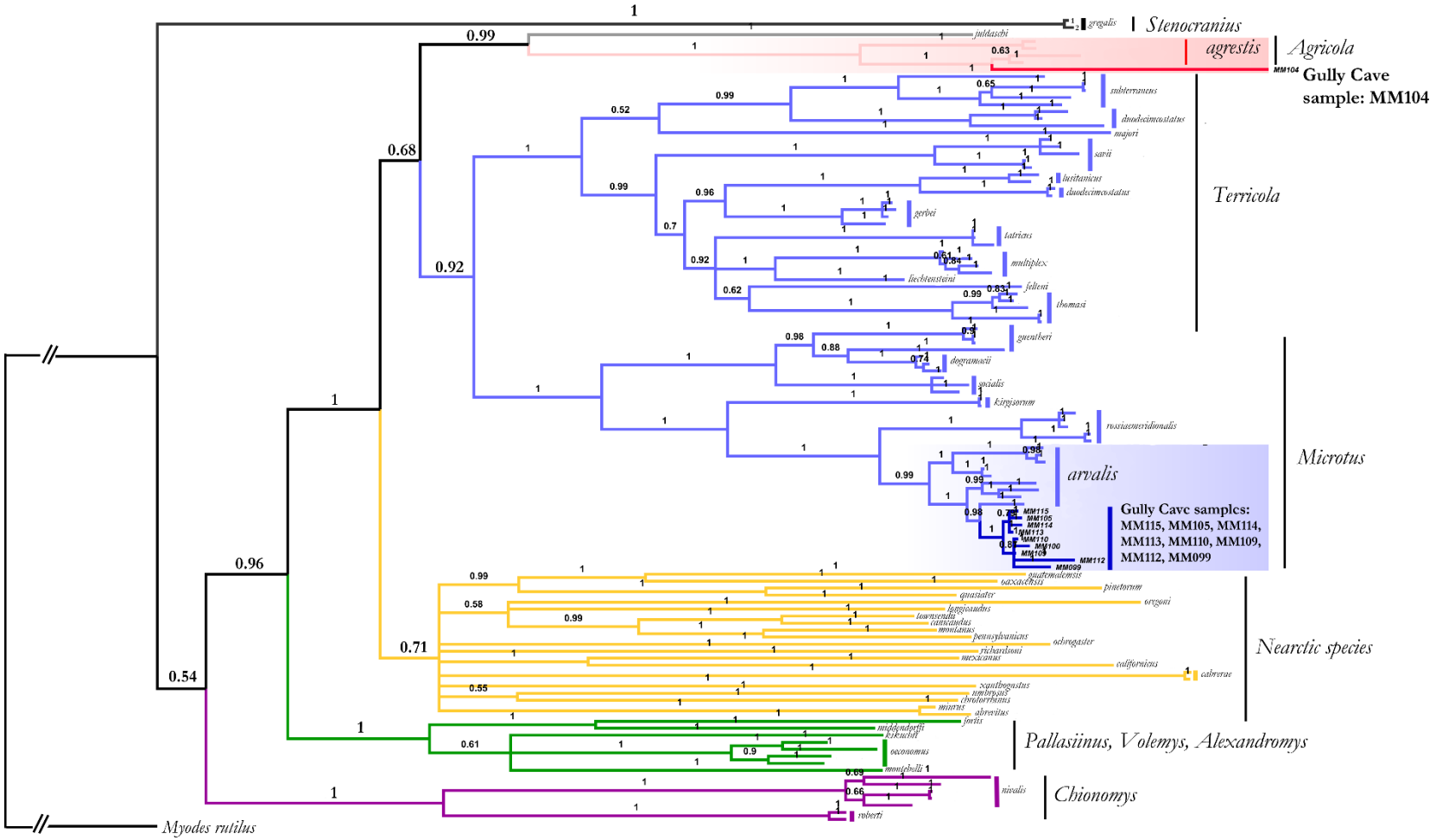


Figure 3.18. Bayesian MCMC tree of all cytochrome *b* haplotypes from Jaarola *et al.* 2004 and *the Microtus agrestis - arvalis* sequences from the Gully Cave assemblage. Rooted with *Myodes rutilus*. Number above branches denote Posterior Probability

3.4 Discussion and main findings

This chapter aimed to identify morphological metrics and use ancient DNA evidence to discriminate between the short-tailed field vole *Microtus agrestis* and common vole *Microtus arvalis*. First, data gathered from modern M_1 specimens of these species was characterised by both existing biometric and novel geometric morphometric methods before being used to assign identifications to fossil Gully Cave *Microtus* M_1s. The relative contributions of genetic and environmental factors were assessed to allow quantification of the maximum heritability of tooth shape and the influence of static allometry on tooth shape was investigated. Lastly, ancient DNA was extracted from a subset of samples and the cytochrome *b* mitochondrial locus sequenced via Next Generation Sequencing (NGS) to explore the utility of BLAST searches and phylogenetic inference methods in species identification.

The main findings in this chapter were:

- Biometric measurements of modern data using the LT4/LT5 index plotted against maximal length (mm) were able to discriminate between species with an estimated 75.6% accuracy for pooled data
- 2D landmark and extended eigenshape geometric morphometrics performed markedly better than linear distance measurements in species separation with 95.55 % and 95.96% accuracy respectively for pooled data
- The effects of static allometry on within-species tooth shape were significant but negligible
- Due to overlap in the ranges of tooth size when measured as maximal length (mm) or when estimated with a proxy value, LogCS, this metric has no power when discriminating between the two species
- Maximum heritability (h^2) for both modern *M. agrestis* and *M. arvalis* was *c.* 80%, indicating a strong taxonomic signal in tooth shape
- The majority of the fossil Gully Cave M_1s were categorized as having *M. arvalis*-like tooth morphology

- Full-length or near full-length ancient cyt*b* was successfully recovered from the majority (10 out of 15) of individuals sampled
- BLAST search results and phylogenetic inference of species placement all showed congruence with the morphological results
- Both Neighbour-Joining (NJ) and Bayesian MCMC phylogenetic analyses returned identical placements for the Gully Cave samples but differed in the level of resolution deeper in the *Microtus* phylogeny
- These analyses records the common vole, *Microtus arvalis*, from the British mainland during the Holocene for the first time

3.4.1 Morphology

Previous attempts to discriminate between *M. agrestis* and *M. arvalis* utilised the biometric technique developed by Nadachowski (1984) - the LT4/LT5 Index plotted against maximal length of the occlusal surface (Roberts & Parfitt 1999) and based on the observation by Chaline *et al.* (1999) that *M. agrestis* displays asymmetry between the buccal and lingual triangles. In these studies, complete separation of the two species was shown. However, when this metric was tested on the database presented here, it was found to be a somewhat accurate species indicator but did not perform as well as the geometric morphometric analyses. Around 75% of individuals were assigned to their correct group and there was a large degree of overlap among species and a number of outliers. While this metric performed reasonably well, micromammal assemblages often contain very large numbers of *Microtus* molars and an error rate of 25% may lead to the misidentification of many specimens, which will adversely affect downstream analyses.

The Nadachowski (1984) study was carried out with a relatively small number of specimens sourced exclusively from within Poland and thus does not provide information about the full range of shape variation present across the full climatic envelope of both species. The molars of *Microtus* species exhibit a rapidly evolving morphology (Chaline *et al.* 1999), which has been shown to co-vary with climate change (Zachos *et al.* 2001; McGuire 2010). Inclusion of specimens from across

Eurasia was essential in this investigation as it allows quantification of how molar shape changes under the full environmental tolerances of both species. Material recovered from different layers in stratigraphic contexts may well have experienced different climatic conditions (within the boundaries of the climate envelope of each species) and, therefore, may show a range of morphologies influenced by prevailing climate. Data obtained from one European region cannot necessarily be extrapolated therefore to the entire Eurasian range of these species. Thus, it is not surprising that the results obtained by the Nadachowski (1984) Polish study are not fully supported here.

The geometric morphometric analyses provided a far more potent means of species delimitation than the use of form-factor ratios. This is due to the independence of these methods from size effects, the representation of form through a geometric configuration (*i.e* no need for *a priori* assumptions about significant features) and the increased power of multivariate statistics. Both analyses showed a large degree of apparent overlap in the PC1 vs PC2 space. The observed changes in the geometry of the triangles, anterior cap and posterior lobe were of a matter of degree, with both species containing some individuals that were more morphologically similar to individuals of the other species than to individuals of their own group. This strongly suggests that species discrimination between *M. agrestis* and *M. arvalis* using the M_1 is not a straightforward task. However, the group-discrimination CV analysis was capable of > 90% correct species allocation and thus provides a much more reliable means of identification. A comparison of the two training sets created here, 2D landmarks and 2D extended eigenshape of the outline boundary, show that a simple set of landmarks representing points of minima and maxima of curvature of the salient and re-entrant triangles can perform as well than an analysis that uses the entire outline of the tooth cusp trace. Either approach can be used to assign unknown specimens to these species (provided the unidentified individuals are genuinely attributable to one or the other). The 2D landmark analyses provide a far quicker and less computationally intense approach to species discrimination. However, the extended eigenshape method provided a far greater depth of information on tooth

shape changes and would be the method of choice when a detailed investigation of comparative morphology is required.

The main shape changes in M_1 between *M. agrestis* and *M. arvalis* were concentrated around the anterior cap, the overall degree of asymmetry between the buccal and lingual triangles and a pattern of change in the most variable triangles – the T1, T2, T5, T6 and T7. It is often difficult to infer the functional significance of such *intra* and *inter*-specific changes in tooth anatomy, the level of biological constraint of different areas of the tooth and the degree to which discrete shape changes are functionally integrated with each other. The occlusal surface of vole molars is primarily involved in the masticatory process, with the series of enamel combs shown to be a highly effective mechanism in chewing (Pozdnyakov 2010). Therefore, it is may be that shape changes observed between the teeth of *M. agrestis* and *M. arvalis* relate to differences in the biomechanical processes involved in mastication. However, other explanations for the results observed here, such as neutral processes, cannot be discounted.

Size discrimination using maximal length in mm and LogCS showed that *M. agrestis* has a significantly longer and larger molar than *M. arvalis*. The degree of overlap in the ranges of these metrics mean that they are not useful in species discrimination and should not be used as such. Allometric effects on tooth shape were significant but weak in both species, a result comparable to what has been reported for other *Microtus* species (*e.g.,* Piras *et al.* 2010). Maximum heritability suggested that the largest proportion of tooth shape variation, 80% for both common and field voles, is accounted for by inherited genetic factors, suggesting that environmental effects do not obscure the taxonomic signal in the training sets and that the taxonomic signal is strong. However, it must be noted that this metric is only an estimate of maximum heritability and not a net value. Polly *et al.* (2011), in a similar study, reported h^2 estimates for *M. agrestis* and *M. arvalis* of 0.77 and 0.89 respectively but this analysis was based on a dataset of $n = 24$ for *M. agrestis* and $n = 38$ for *M. arvalis*. The much larger sample sizes used in this study provide a more reliable indication of this value.

As allometry was weak and measurement error negligible, the remaining environmental effects on tooth shape may therefore be related to climatic factors. Tooth morphology in *Microtus* spp. has been shown to co-vary strongly both along geographic clines in modern extant species (McGuire 2010, Piras *et al.* 2010) and chronologically over the climatic oscillations observed in the Quaternary (Piras *et al* 2009, McGuire 2011). It is possible, therefore, that climate variation is responsible for the proportion of tooth shape variation not accounted for by the combined effects of genetics, allometry and measurement error investigated in this study. This will be investigated in more detail for *M. arvalis* in Chapter 4.

An implicit assumption in this study is that modern specimens of *M. agrestis* and *M. arvalis* are representative of ancient material in terms of tooth shape. However, the high degree of morphological plasticity in vole teeth and their ability to change form rapidly (Hadly *et al.* 2004) suggest that this assumption may not be entirely appropriate. Nevertheless, although the fossil material considered here is only a few thousand years old, the 2D training set performed well in assigning ancient specimens of unknown identity to a group and allowed material from a cave sequence of Lateglacial to Early Holocene age to be identified to species level. It also enabled each fossil individual to be assigned to a climatic period resulting in the identification of *M. arvalis* in the British Holocene record for the first time. The presence and continuity of common vole material throughout the cave sequence would tend to support a pattern of population persistence across a period of abrupt climatic fluctuation at Gully Cave. Radiocarbon evidence from the Eurasian beaver (*Castor fiber*) has shown that cave sites in the South-West of England may have acted as micro-refugia over the Younger Dryas cold phase (see Chapter 5), allowing temperate-adapted species to persist locally until Holocene climatic amelioration was initiated (see also Stewart & Lister 2001). However, Brace *et al.* 2016 used aDNA to demonstrate population discontinuity of collared lemmings where material of this species was present throughout a cave sequence. Therefore, common vole population dynamics at Gully Cave needs to be further investigated. The continuity/discontinuity of *Microtus arvalis* over the Younger Dryas at this site is further investigated in (Chapter 4).

3.4.2 Ancient DNA

This study demonstrated the efficacy of ancient cytochrome *b* in accurately identifying *Microtus* species dating back as far as the Lateglacial from low amounts of starting material derived from the M_1. Ancient DNA has been used as a tool for archaeological species identification for some time, initially using Polymerase Chain Reaction and Sanger sequencing (for example in geese: Barnes *et al.* 2010, and canids: Horsburgh 2008) and more recently using Next Generating Sequencing (NGS), often in conjunction with capture enrichment (*e.g* Templeton *et al.* 2013). Archaeological remains may not always be identifiable to species level from gross anatomy alone, either in situations where the material is fragmentary, or does not possess species-specific characters or, as is the case here, shows morphological similarity with closely-related sympatric species. Ancient DNA therefore offers an opportunity to identify species from remains with little or no morphological information and as can also serve as a second line of confirmatory evidence in cases where prior morphological analyses have taken place.

As high-throughput sequencing amplifies all DNA in a given sample, endogenous target-species DNA (which usually represents only a small fraction of total DNA) is amplified alongside contaminants, both modern and ancient. Identifying species from the resulting short and uninformative read sequence data can prove challenging. The species used in this study benefitted from a number of factors that allowed these challenges to be overcome. First, contextual and morphological information was readily available. The occurrence of *Microtus* species in the British archaeological and fossil record together with the M_1 anatomy indicated that these teeth could belong to one of two species, *M. agrestis* or *M. arvalis* (Nadachowski 1984; Yalden 2010). This allowed the reads to be mapped to an outgroup (*M. oeconomus*), phylogenetically equidistant from both species, to ensure all data were similarly impacted by the effects of read mapping ascertainment bias. The resultant consensus sequence could then be used in the BLAST search analyses. Second, both species have many modern cyt*b* representative sequences stored in the NCBI database, which allowed accurate local alignments to their extant contemporaries. Finally, while the resolution at deeper levels of the *Microtus* genus remain unresolved, lower level ordinal phylogenies are

'known' and cyt*b* has been previously shown to be able to recover these accurately. This allowed the ancient cyt*b* sequences generated here to be placed in an existing phylogeny as a test of the strength of the taxonomic signal in the ancient data. In situations where the above factors are not met, it may be more appropriate to initially use PCR in conjunction with species-specific primers to ascertain an identification prior to NGS, perhaps utilizing species-specific barcode loci (Hebert *et al.* 2003).

Both the Neighbour-Joining (NJ) and Bayesian MCMC tree building methods performed well here with regards to correct species placement and showed highly congruent topologies at lower taxonomic levels. This suggests that when a phylogeny is known, and ancient data can be combined with multiple within-species sequences, faster and less sensitive tree-building methods can be used for ancient species identification. The NJ tree was unable to recover the deep level phylogenetic relationships within the *Microtus* sub-genera. However, when it is desirable to have correct tree topologies at all taxonomic levels, topology can be constrained in the analyses, which will overcome problems associated with unresolved polytomies.

Despite not being considered a species-specific 'barcode' locus such as *cox1* (Hebert *et al.* 2003; Hajibabaei *et al.* 2007), this study has highlighted the utility of cyt*b* DNA in species identification. The *Microtus* genus contains many closely-related species and is characterised by a series of rapid speciation events (Fink *et al.* 2010). This has led to many species showing overlap not just in morphology but also in *intra* and *inter*-specific genetic distances as measured by cyt*b* (Jaarola *et al.* 2004). The fact that this locus performed well here indicates that it is informative enough to delimit species, even in situations where species show low cyt*b* divergence. However, the use of ancient DNA for species identification has a number of logistical drawbacks compared with the use of morphological evidence. When used without prior capture enrichment for target loci, the random nature of shotgun sequencing may not yield reads covering the locus/loci of interest and the depth of coverage required to make correct base calls may not be achieved. Furthermore, ancient DNA sequencing, library builds and NGS sequencing are far less cost-effective than morphological methods. With material as small as the *Microtus* molar, the complete destruction of the

sample is also necessary. This has the effect of reducing the numbers of individuals that can be processed, significantly fewer than can be included in a morphology study. It is therefore not appropriate for identifying datasets of containing hundreds of individuals.

These findings have provided support for the results of the morphometric analyses in this chapter and also presented an opportunity to investigate millennial scale, climate-driven impacts on voles over the Lateglacial-Holocene in Britain (see Chapter 4). Additionally, they have demonstrated the efficacy of aDNA extraction and NGS sequencing technology in successfully extracting full length cyt*b* sequences from isolated specimens of the *Microtus* lower first molar. As these teeth are found in many terrestrial Quaternary sequences (Escudé *et al.* 2013) this suggests these methods may also be used to further investigate the enigmatic history of this speciose genus.

Chapter 4: Climate driven ecomorphological and genetic change in the common vole (*Microtus arvalis*) over the Pleistocene-Holocene boundary in Britain.

4.1. Introduction

4.1.1 Overview

Ancient mitochondrial DNA and morphological analysis of small mammals have provided a wealth of information on the patterns and process of post-LGM recolonisation of Britain and north-west Europe (Searle *et al.* 2009; Cucchi *et al.* 2014; Brace *et al.* 2016). This has led to the identification of periods of ecosystem instability (Brace *et al.* 2009), morphological responses to island isolation (Cucchi *et al.* 2014), and the development of the theory of the 'Celtic Fringe' (Searle *et al.* 2009; Brace *et al.* 2016; Chapter 1, Section 1.1.9). Following on from the identification of the common vole *Microtus arvalis* in the Lateglacial to early Holocene sediments in Britain (see Chapter 3) this chapter seeks to investigate further the phylogeographic and population history of this species in Britain. The broad aims are:

- To identify the phylogeographic origins of British populations of the common vole *Microtus arvalis* using cytochrome *b*
- To describe population continuity/turnover events from *M. arvalis* excavated from a single site, Gully Cave, Ebbor Gorge, Somerset
- To identify ecomorphological size and shape changes in the tooth that may be associated with either changes in climate or population turnover events

4.1.2 Ecology and phylogeography of the common vole, Microtus arvalis

Fossil evidence shows that the common vole has been present in Europe in its current morphological form since at least the Late Cromerian (0.500 – 0.450 Myr; Kowalski 2001). The species has been identified as being the most abundant modern European mammal (Mitchell-Jones *et al.* 1999) and is so numerous as to be considered an

agricultural pest in many regions (Jacob *et al.* 2013). These voles are mainly folivorous and typically inhabit open steppe-grassland ecotypes with a large, continuous distribution stretching from the Atlantic coast of Spain in the west to central Russia in the east (Chapter 3, Fig 3.1). Populations persist at elevations from sea-level to around 2600m in the Alps (Mitchell-Jones *et al.* 1999; Jacob *et al.* 2013). Its northern range which extends as far north as the extreme south of Finland is delimited by mean July temperatures below 16^{o}C and its southern boundary extends to Turkey where it is delimited by arid environments (Spitz 1977). These voles are absent from Scandinavia (bar the extreme south of Finland) and mainland Britain, although populations exist on the Guernsey and Orkney archipelagos, the latter likely via a human-mediated introduction during the Neolithic (Martínková *et al.* 2013).

Two 'forms' of the common vole exist with a hybrid zone in central European Russia: the '*arvalis*' form which occupies the western part of its range and the '*obscurus*' form which is found in the east (Fig 4.1, Meyer *et al.* 1997; Haynes *et al.* 2003; Jaarola *et al.* 2004). These fall into well-defined, distinct mitochondrial lineages and differ from each other cytogenetically (Meyer *et al.* 1997; Bulatova *et al.* 2010). Given that it is highly unlikely that the British Lateglacial or Holocene populations derive from the '*obscurus*' form (Fig 4.1) this study will only be concerned with a geographic sub-sample; the '*arvalis*' type. The common vole shows a pattern of phylogeographic structuring typical of small mammals. These taxa tend to have a high level of spatial structuring with multiple and often complex divergent mitochondrial lineages, usually attributed to periods of differentiation in glacial refugia followed by recolonization events (Hewitt 2000). For example, the bank vole (*Myodes glareolus*) includes eight lineages (Wójcik *et al.* 2010), the weasel (*Mustela nivalis*) five (McDevitt *et al.* 2012; Rodrigues *et al.* 2016) and the field vole (*Microtus agrestis*) also five (Herman *et al.* 2014).

The phylogeography of *M. arvalis* has been inferred largely from analyses of the protein-coding cytochrome *b* gene and non-coding Control Region (CR) of the mitochondrial genome (Jaarola *et al.* 2004; Stojak *et al.* 2015). While the non-coding CR likely meets the criteria for selective neutrality cyt*b,* a protein coding gene involved in the bc1 complex, has shown significant deviations from neutrality in some

organisms (Neilsen & Weinreich 1999; Rand 2000). Therefore, this gene is potentially varying under selective pressure, which may lead to adaptive responses in novel environments. This, in turn, can create distinct lineages among geographically separated populations that may be interpreted erroneously as the result of repeated bouts of glacial isolation and recolonisation.

For example, Mishmar *et al.* (2003) and Ruiz-Pesini *et al.* (2004) showed that humans exhibit elevated variability in cytochrome *b* sequences in temperate regions coupled with significant changes in functionally important amino acid positions. Evidence that adaptive responses to unique environments have had an impact on the phylogeography of the common vole has been shown by Fink *et al.* (2004) who demonstrated that functionally important cyt*b* protein sequences were often affected by selection. However, relationships between protein sequences in the common vole differ from relationships between mtDNA lineages, indicating that selection pressure did not create the strong phylogeographic structuring seen in this species (Fink *et al.* 2004). The most likely scenario is that the climate-driven demographic history of *M. arvalis* populations has been the primary driver of diversity in mtDNA, in association with repeated episodes of adaptation to novel environments (Fink *et al.* 2004).

Microtus arvalis exhibits five main cyt *b* haplotype lineages in Europe that are largely allopatric, show a high degree of variation in geographic range and have a number of hybrid suture zones (Fink *et al.* 2004; Jaarola *et al.* 2004; Tougard *et al.* 2008; Bužan *et al.* 2010; Stojak *et al.* 2015). The large Western clade contains two well-defined sub-lineages – Western-North, which encompasses individuals from north-east France, Belgium, Luxembourg and the Orkney archipelago and Western-South, which encompasses individuals from Spain and western France. This lineage is purported to have a glacial refuge on the Iberian Peninsula (Haynes *et al.* 2003). The Central lineage is composed of individuals from Switzerland, France, the Netherlands, Denmark and Germany with a refugial origin in the Balkan Peninsula (Stojak *et al.* 2015). The Italian lineage, with a refugial area on the Italian peninsula, contains individuals from northern Italy and southern Switzerland. Individuals from Slovenia, Bosnia Herzegovina, Montenegro and Serbia from a distinct clade in the Balkan lineage while

specimens from Finland, Russia, Ukraine, Poland, Slovakia and Hungary group together in the Eastern lineage. A highly divergent sub-lineage has also been discovered in Freiberg, west-central Germany (Fig 4.1; Haynes *et al.* 2003; Fink *et al.* 2004; Jaarola *et al.* 2004; Tougard *et al.* 2008; Stojak *et al.* 2015).

These intraspecific lineages differ from one another through their highly heterogeneous evolutionary and demographic histories, with evidence of some originating in cryptic northern refugia and differential impacts of landscape on the timing and pattern of post-glacial recolonisation (Tougard *et al.* 2008). Evidence of population persistence outside of the classic refugial areas of the multiple Mediterranean peninsulae is provided by fossil evidence showing that *M. arvalis* was present in the Carpathian Basin during the last ice-sheet advances along with other typical temperate-adapted species such as brown bear (*Ursus arctos*), red deer (*Cervus elaphus*), roe deer (*Capreolus capreolous*), red fox (*Vulpes vulpes*), weasel (*Mustela nivalis*) and pine marten (*Martes martes*) (Stojak *et al.* 2015). That this population contributed to subsequent extant post-glacial populations was demonstrated by Stojak *et al.* 2015, using cyt*b,* who suggests this as the source population for the Eastern clade. The German Freiberg lineage also infers a second area of cryptic northern refugia for the common vole as it is highly divergent from the proximal Central and Eastern lineages (Stojak *et al.* 2015).

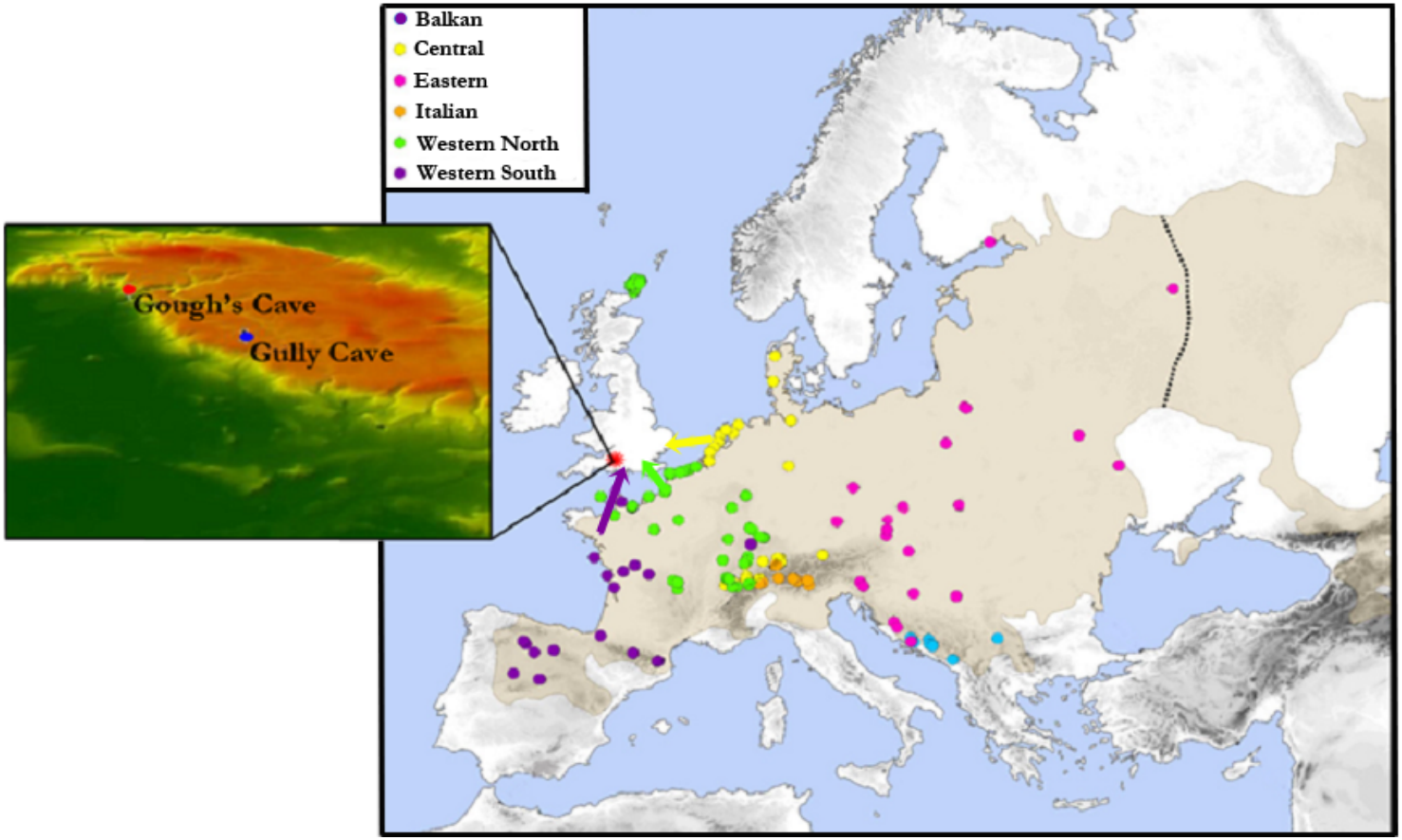


Figure 4.1. Main mitochondrial lineages for common vole *Microtus arvalis* in Europe. Figure modified from Stojak *et al.* (2015). Boundary between '*arvalis*' and '*obscurus*' forms indicated in dashed lines. Gully Cave site indicated in red in South-West England with 3D terrain of the region showing proximity of Gully Cave to Upper Palaeolithic site Gough's Cave. Arrows indicate potential geographic source populations for Britain.

4.1.3 Common vole, Microtus arvalis, in Britain

The discovery of common vole remains in Holocene deposits at Gully Cave, Ebbor Gorge, UK (see Chapter 3) was the first evidence of this species in mainland Britain during this interval (Corbet 1961, 1986; Yalden 1999). Until then, the absence of *M. arvalis* from Britain was regarded as an anomaly given that the temperate Atlantic climate would have provided suitable grassland habitat for the species and that other small mammals with similar ecologies and dispersal abilities (*e.g.,* field vole *Microtus agrestis* and bank vole *Myodes glareolus*) not only colonized mainland Britain but have persisted to the present day (Yalden 1999). Currently *M. arvalis* is absent from the entirety of the British mainland but occurs on the Guernsey Isles and the Orkney archipelago. In the latter region divergent evolutionary processes associated with isolation on offshore islands have led to a high level of genetic divergence and morphological change relative to its mainland European counterparts (Martínková *et al.* 2013; Cucchi *et al.* 2014), leading to its taxonomic status as a distinct subspecies, *M. a. orcadensis*.

The presence of Holocene common vole remains in south-west England leads to the question of whether a 'sweepstake' dispersal method of colonisation from mainland Britain to Orkney is a possible alternative to the generally accepted theory of human-mediated introduction. However, the evidence for an artificial introduction is overwhelming. Orkney was covered by the North European ice-sheet during the Last Glacial Maximum (LGM: *c.* 26, 000 – 21, 000 cal BP; Peltier & Fairbanks 2006) during which the species was undoubtedly absent from those islands. No landbridge existed post-LGM when the climate had warmed sufficiently for suitable habitats to become available (Yalden 1982). Thus, over-land colonization would not have been possible. The sole other *Microtus* species present in Britain, the field vole *M. agrestis*, naturally colonised many off-shore islands that were connected to mainland Britain via land-bridges but did not colonise Orkney (Corbet & Harris 1991). Furthermore, a large-scale radiocarbon dating program of *M. a. orcadensis* archaeological material from a variety of Orkney islands placed this species' interval of occupation as extending from the Late Neolithic until the present day, with the oldest ^{14}C date recovered being

5050-5437 cal BP (Martínková *et al.* 2013). Subsequent molecular analyses using a combination of cyt*b* and microsatellites have identified the likely source population as the North-West coast of Europe, within the Western-North clade (Martínková *et al.* 2013).

The geographical proximity of the south-west of England to the north-west coast of Europe suggests a geographical origin for the British Gully Cave common voles as either the Western-North or Western-South clades. However, more complex patterns of recolonisation driven by abrupt changes in climate have been observed in other small mammals and cannot be ruled out for *M. arvalis*. The water vole *Arvicola amphibious* recolonized Britain in two temporally separated waves over the Late Pleistocene and Holocene and has a remnant population structure in the UK that reflects its demographic history (Brace *et al.* 2016). Collared lemmings (*Dicrostonyx torquatus*) experienced a series of climate-driven extinction-recolonisation events at a regional level in the Late Pleistocene (Brace *et al.* 2012). As the common vole specimens included in this study span two warm phases (the Lateglacial Interstadial and early Holocene) separated by a brief cold phase (the Younger Dryas) it is possible that this species also experienced multiple climate driven extinctions and recolonisation events.

4.1.4 Abrupt climate change and ecomorphology

The term 'ecomorphology' describes the field of study that deals with how a species' morphological adaptations relates to its ecology and can be subdivided into two branches – the effect of environment on the functional design of an organism or the effect of an organism's functional morphology on its capability to interact with its environment (Wainwright 1991; Sievwright & Macleod 2012). Using an ecomorphological framework, hypotheses relating to how a species' morphological traits constrain or facilitate its interactions with its environment may be formulated and tested (Sievwright & Macleod 2012). The current period of anthropologically-mediated climatic warming has precipitated a new wave of research using past climes to predict effects of abrupt climate change on biota. To this end, the study of

morphology has been suggested as unifying approach within which the interactions between species and climate can be assessed across taxa, ecosystems, space and time (Polly *et al.* 2011). Particular morphological traits that have a clear functional relationship to an organism's physical (*e.g* climate) or biological (*e.g* macrovegetation) environment can be said to mediate species-environment interactions and, if so, these traits will experience selection pressures. Therefore, possession of such traits can influence the environmental tolerances of species, even to the extent of playing a major role in the determination of a species' geographical distribution, its survival and its fitness potential (Polly *et al.* 2011). The measurement of such key ecomorphological traits, their distributions and their interactions with the environment forms the basis of a novel approach for understanding feedbacks between climate and biotic elements, termed 'ecometrics' (Eronen *et al.* 2010).

Several studies using ecometric or ecomorphological approaches have demonstrated clear relationships between key phenotypic traits and environment and/or ecology. For example, various bovid morphological characters have been used to infer palaeo-environments (Macleod & Rose 1993, DeGusta & Vrba 2005, Plummer *et al.* 2008), mammal teeth (which have been used extensively in such studies) have been shown to co-vary with specific climatic variables and be accurate indicators of past environmental change (Damuth *et al.* 2001; Polly *et al.* 2011*)* and morphological variation in the falconiform humerus has been shown to reflect species-specific ecological functions (Sievwright & Macleod 2012). Distinct ecomorphs have been identified in Late Pleistocene mammal populations and have been associated with population extinctions. Tooth wear analyses of Beringian wolves suggested that this population was ecologically specialised for hunting and scavenging megafauna, subsequently becoming extinct at the same time its prey disappeared from this region (Leonard *et al.* 2007). Arctic fox populations living at lower latitudes in Europe possessed smaller paws than modern arctic foxes, a hypothesized ecological adaptation to reduced snowfall (Stewart 2008). These southern arctic fox populations became extinct as the climate ameliorated and they failed to track their declining habitat northwards (Dalén *et al.* 2007).

The teeth of small mammals directly interact with the environment via their role in the masticatory process (Ungar 2010). They have been shown to be excellent proxies for both detecting past changes in climate and also in species turnover events (*e.g* Cerling *et al.* 1997; MacFadden 2000; Fortelius *et al.* 2002). For example, geometric morphometric analyses of the Orkney vole *M. a. orcadensis* M_1 allowed detection of a rapid increase in size associated with shape changes on the occlusal surface post-colonisation of the archipelago (Cucchi *et al.* 2014). The rapidly evolving form of the molars of *Microtus* species (Guthrie 1965) makes them ideal proxies with which to infer population-level change in tooth size and shape which may be associated with turn-over events in the Gully Cave *Microtus arvalis* population(s) throughout the cave sequence.

4.1.5 Aims and research questions

Specific aims addressed in this chapter are:

Morphology:

- Evaluate patterns of tooth shape variation for deterministic association with patterns of abrupt climatic change. Specifically:
    - Use changes in tooth shape and size to investigate population turn-over events through the early Lateglacial – Younger Dryas – Holocene layers of the Gully Cave sequence against a null hypothesis of no significant difference between climatic periods
- Examine the effects of sample size on levels of group distinctiveness via a series of data augmentation simulations

Ancient DNA:

- Use ancient cytochrome *b* sequences to:
    - Determine the European geographical source population for the Gully Cave specimens from the known *M. arvalis* European mtDNA clades

- Use haplotype networks and Bayesian MCMC phylogenetic analyses to investigate possible population turnover events for this species in the Gully Cave sequence
- Infer the phylogenetic relationship of the British Gully Cave population(s) to the Orkney population

4.2 Methods

4.2.1 Site and sampling

A full description of the Gully Cave site, excavation techniques and methods of inferring ages/climatic periods for the layers is given in Chapter 3, Section 3.2.1. Figure 4.2 shows the broad divisions of the cave sequence into Lateglacial Interstadial, Younger Dryas and Holocene. A total of 40 M_1's had a sufficient degree of morphological preservation to be included in the 3D geometric morphometric analyses and 14 lower first molars were used for DNA analyses (Table 4.1). Samples were processed, ancient DNA extracted, sequenced, assembled and species identifications determined as described in Chapters 2 & 3. Standard procedures for checking for contamination and NUMTs were carried out, also as described in Chapter 3, Section 3.2.2. Sequences obtained here were combined with a larger existing database of modern and ancient published cyt*b* sequences from the GenBank repository. Combinations of data varied depending on the analysis; accession numbers, unique haplotypes and citations for all Genbank data are provided in Appendix Table CT1.

4.2.2 Ancient DNA

To address the first aim of determining the European source population for British Late Pleistocene and early Holocene common vole populations both a Median-Joining network and Bayesian MCMC inference of phylogeny were used. The MJ network was here calculated in PopART (http://popart.otago.ac.nz) with (ε) set to zero. The dataset contained 388 taxa, including duplicate haplotypes, representing all six sub-lineages of the common vole within Europe with *M. obscurus* excluded.

Figure 4.2. Interior of Gully Cave, Ebbor Gorge, showing sediments of Lateglacial Interstadial, Younger Dryas and Holocene age (bottom to top). Image (Schreve 2014).

Table 4.1. Sample table of all *Microtus arvalis* lower first molars used in this study. Square that each tooth was excavated from in Gully Cave along with spit depth in cm and type of analyses each tooth was used in are shown. Only specimens for which aDNA extraction was successful are listed.

Specimen ID	Square	Depth (cm)	Climatic period	Analyses (GMM/aDNA)
MM027	F0	+10 to Datum	Holocene	GMM
MM104	F0	-20 to -30	Holocene	GMM
MM106	F0	-20 to -30	Holocene	GMM
MM108	F0	-30 to -40	Holocene	GMM/aDNA
MM112	F0	-30 to -40	Holocene	GMM/aDNA
MM114	F0	-30 to -40	Holocene	GMM/aDNA
MM115	F0	-30 to -40	Holocene	aDNA
MM148	G0	+20 to +10	Holocene	GMM
MM149	G0	+20 to +10	Holocene	GMM
MM105	F0	-60 to -70	Younger Dryas	aDNA
MM113	F0	-60 to -70	Younger Dryas	aDNA
MM123	F0	-60 to -70	Younger Dryas	GMM
MM124	F0	-60 to -70	Younger Dryas	GMM
MM125	F0	-60 to -70	Younger Dryas	GMM
MM127	F0	-60 to -70	Younger Dryas	GMM
MM129a	F0	-60 to -70	Younger Dryas	GMM
MM129b	F0	-60 to -70	Younger Dryas	GMM
MM131	F0	-60 to -70	Younger Dryas	GMM
MM132	F0	-60 to -70	Younger Dryas	GMM
MM133	F0	-60 to -70	Younger Dryas	GMM
MM136	F0	-60 to -70	Younger Dryas	GMM
MM156	G0	-50 to -60	Younger Dryas	GMM
MM169	G0	-40 to -60	Younger Dryas	GMM
MM060	F2	-120 to -130	Bølling	GMM
MM062	F2	-120 to -130	Bølling	GMM
MM064	F2	-120 to -130	Bølling	GMM
MM065A	F2	-120 to -130	Bølling	GMM
MM069	F2	-120 to -130	Bølling	GMM
MM070	F2	-120 to -130	Bølling	GMM
MM071	F2	-120 to -130	Bølling	GMM
MM072	F2	-120 to -130	Bølling	GMM
MM074	F2	-120 to -130	Bølling	GMM
MM076	F2	-120 to -130	Bølling	GMM
MM077	F2	-120 to -130	Bølling	GMM
MM082	F2	-120 to -130	Bølling	GMM
MM085	F2	-120 to -130	Bølling	GMM
MM088	F2	-120 to -130	Bølling	GMM
MM090	F2	-120 to -130	Bølling	GMM
MM091	F2	-120 to -130	Bølling	GMM
MM093	F2	-120 to -130	Bølling	GMM
MM094	F2	-120 to -130	Bølling	GMM
MM095	F2	-120 to -130	Bølling	GMM
MM096	F2	-120 to -130	Bølling	GMM
MM099	F2	-120 to -130	Bølling	aDNA
MM100	F2	-120 to -130	Bølling	aDNA
MM109	F2	-120 to -130	Bølling	aDNA
MM110	F2	-120 to -130	Bølling	aDNA

Bayesian MCMC inference of phylogeny was carried out in MrBayes v3.2.6 (Huelsenbeck *et al.* 2001). with the best-fit model of nucleotide substation determined by JModelTest2 (Darriba *et al.* 2012). The Akaike, Akaike Corrected and Bayesian Information Criteria (AIC, AICc and BIC) as well as the Decision Theory method all calculated the General Time Reversible Model with a proportion of invariant sites and a gamma-distributed rate variation among sites (GTR+G) as the best-fit model. Data representing all six sub lineages were included, encompassing both modern sequences and those sampled from ancient *M. arvalis* specimens from Orkney and Belgium (Table CT1). A high ratio of taxa to uninformative characters can create an excess of polytomies and unresolved branches in the resulting phylogeny. Therefore, only unique haplotypes were used. This resulted in a final dataset consisting of 287 individuals. *Microtus obscurus* sequences were included and the tree was rooted with *M. rossiaemeridionalis* (Accession AY513819). Two independent chains were run for 10^6 iterations each, sampling trees and model parameters from the posterior every 1000 iterations. The first 10 percent of each run was discarded as burn-in. TRACER v.1.5 (Rambaut *et al.* 2014) and Are We There Yet (AWTY: Wilgenbusch *et al.* 2004) were used to check for chain convergence and sufficient sampling before a 50 percent consensus tree was created and edited in FigTree v.1.3 (Rambaut & Drummond 2010).

To investigate the second aim of identifying population continuity/replacement over the Gully Cave Bølling, Younger Dryas and Holocene stratigraphic layers an attempt was made to improve the phylogenetic resolution of the tree using Bayesian genealogy sampling and time-stamped data. A coalescent model of evolution was implemented in BEAST v.1.8.3 (Drummond *et al.* 2012). Previous studies have attempted to produce accurate posterior distributions of the time to most recent common ancestor (tMRCA) for the major lineages within this species using similar methods and datasets (*e.g.,* Stojak *et al.* 2015) but interpretation of these results have been confounded by large confidence intervals in divergence dates between clades. The addition of the 10 Gully Cave sequences with broad age range estimates is unlikely to result in improved resolution in this area. Thus, the aim of this investigation was not to estimate lineage

divergence times but to examine the effect that inclusion of time series data and different demographic models have on phylogenetic positioning.

A dataset of 165 partial cyt*b* sequences, 1040bp in length and including duplicate haplotypes, was assembled for individuals belonging to the Western-North lineage only. This included all ancient data sampled from the Orkney archipelago of Neolithic to modern age and ancient samples from the Belgium coast of Medieval Age (Martínková *et al.* 2013). Where ancient specimens did not have specific radiocarbon dates available, they were assigned the median age for their period with a range extending to the upper and lower bounds of that period (Table CT2).

BEAST simulations were carried out in parallel on the Cipres Web Portal (Miller *et al.* 2010). Sequences were aligned along codon positions in Mesquite (Maddison & Maddison 2016). Two molecular clocks were compared, a strict molecular clock and an uncorrelated lognormal relaxed molecular clock, which allows rates variation among tree branches (Drumming *et al.* 2006). For each of these clock types, two demographic models were also compared; constant population size and simple expansion growth. This resulted in four BEAST simulations: *i*) strict clock with constant population size, *ii*) relaxed log-normal clock with constant population size, *iii*) strict clock with a coalescent exponential population size and *iv*) relaxed log normal clock with a coalescent population size. The clock rate was set at 3.27×10^{-7} based on a robust estimate of the cyt*b* substitution rate determined by Martínková *et al.* (2013) from time series of ^{14}C dated *M. arvalis* specimens. Each simulation was run for 100 million generations, on four independent MCMC chains sampling from the posterior every 10000 generations. To select the most appropriate model, the integrated BEAST option of path sampling and stepping stone sampling was used in order to estimate marginal likelihoods (MLEs) for each model. These were run on the default parameters of four independent MCMC chains comprised of 1000 steps of 100 000 generations following a burn-in period of 10 million generations. Bayes factors were then calculated from these data and model preference determined using the criteria of Kass & Raftery (1995).

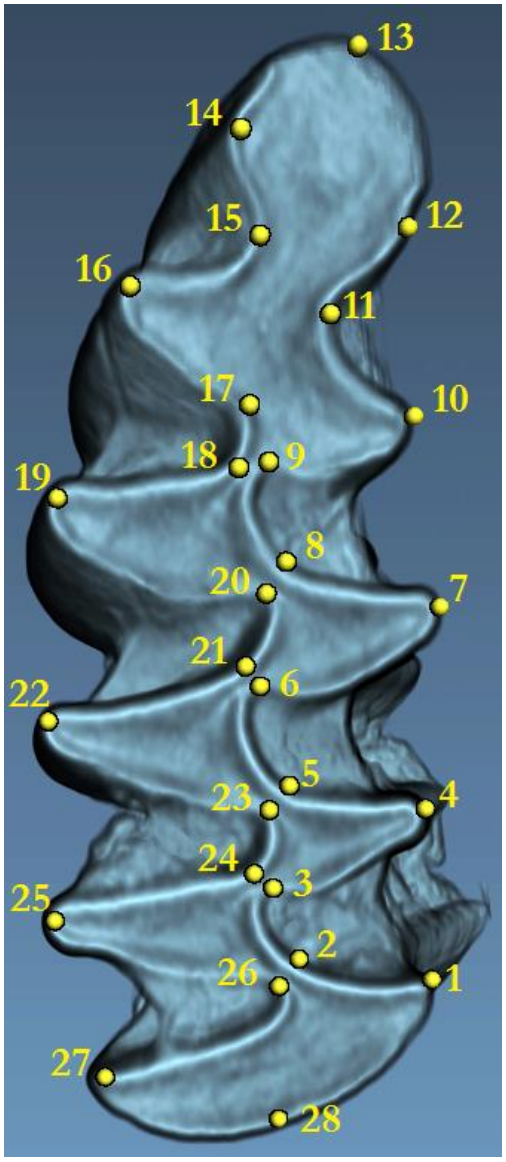
To examine the possibility of population subdivision between Holocene, Younger Dryas and Bølling periods an Analysis of Molecular Variance (AMOVA) was performed in Arlequin v.3.5 (Excoffier & Lischer 2010). Individuals were grouped by population and the significance of the AMOVA was obtained via 1000 non-parametric permutations

4.2.3 3D Geometric morphometrics

Computed microtomography (MicroCT) and 3D landmarks

Forty right *Microtus* M_1s were subjected to X-ray microtomography scanning on a Nikon Metrology HMX ST 225 instrument within the Imaging and Analyses Centre (IAC) at the Natural History Museum London. Molars were batch-scanned using a 3D printed grid which gave a voxel (volume pixel) resolution of *c.* 13 microns per tooth. Volume files were converted to greyscale tag image format (tiff) stacks using Fiji (Schindelin *et al.* 2012). Each individual .tiff stack, which ranged from between 123 to 174 slices per specimen, was volume rendered to visualise the 3D representation of M_1 morphology. Initially, the intention was to process each volume to remove CT scan artifacts and to extract the 3D enamel boundary. This would then be converted into a stereo lithography file (STL) and exported as a wire-mesh boundary outline. This would be processed to create a clean, remeshed 3D model on which to place 3D landmarks and 3D semi-surface landmarks. However, despite various manual and automated protocols being carried out, there was insufficient greyscale density differentiation between the enamel boundary of the buccal and lingual triangles and the matrix of cementum and sediment that fills the re-entrant triangles (Fig 4.3). Therefore, an alternative approach was developed. A series of 28 3D landmarks were collected on the outline of the tooth over a series of five ortho-slices orientated parallel to the occlusal surface of the tooth (Figs 4.4 a-b). A larger number of landmarks were used in this study than in Chapter 3 as this dataset represent intra-specific variation and, as such, less shape variation may be present in the dataset. Therefore, the use of more landmarks will ensure as much shape information as possible is extracted from the tooth. As these molars are hypsodont and rootless no anatomical structure that would conform to type I or II landmarks

was present from which to terminate the ortho-slice landmark configuration. Therefore, the maximal point of curvature across all lingual salient angles was used as the final ortho-slice location (Appendix C, Fig CF3). This was calculated as the point of maximal curvature from the line representing maximal



1. Maximal point of curvature on the lingual terminal edge of the posterior loop.

2, 5, 8. Posterior maximal curvature of the lingual edge of the re-entrant triangle at the points of intersection with the buccal enamel boundary.

3, 6, 9. Anterior point of maximal curvature of the re-entrant angles at the boundary between enamel and cement on the buccal side.

4, 7, 10. Point of maximal curvature on the salient angles between leading and trailing edges on the buccal side.

11. Maximal point of curvature on the lingual re-entrant angle of the anterior cap.

12. Maximal point of curvature on the buccal side of the anterior cap.

13. Maximal point of curvature on the most mesial point of the anterior cap.

14. Maximal point of curvature on the buccal side of the anterior cap.

15. Maximal point of curvature on the buccal re-entrant angle of the anterior cap.

16, 19, 22, 25. Points of maximal curvature on the salient angles between leading and trailing edges on the lingual side.

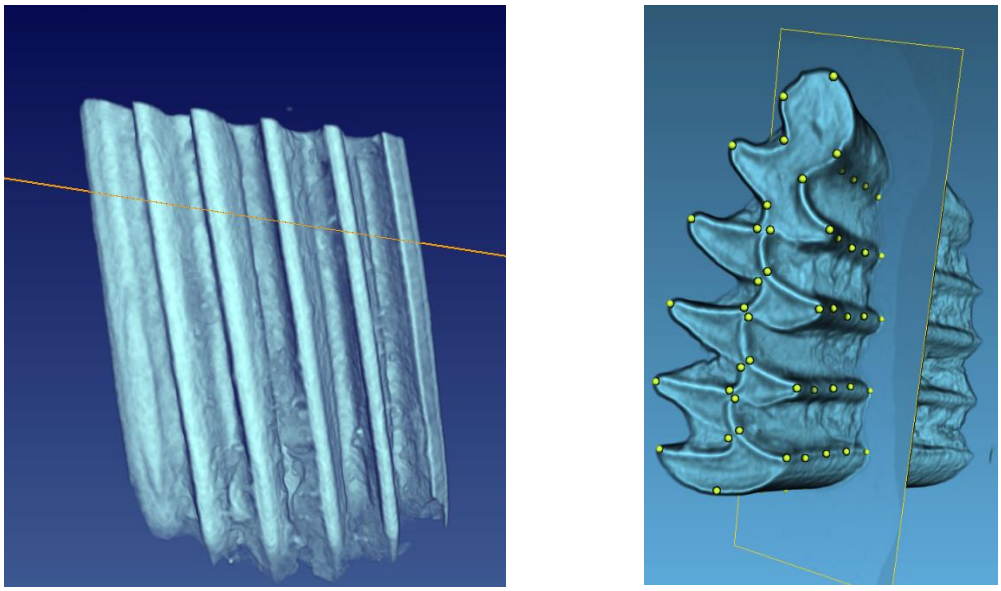
17, 20, 23, 26. Anterior points of maximal curvature of the re-entrant angles at the boundary between enamel and cement on the lingual side

18, 21, 24. Posterior point of maximal curvature of the re-entrant angles at the boundary between enamel and cement on the lingual side.

27. Maximal point of curvature on the lingual terminal edge of the posterior loop.

28. Point of maximal curvature on the enamel boundary of the posterior lobe.

Figure 4.3. Anatomical locations and descriptions of the 28 landmark points on each orthoslice of the *Microtus arvalis* M_1.



a. **b.**

Figure 4.4. Ortho-slices taken at equal intervals throughout the *Microtus arvalis* M_1. (a) shows the ortho slice situated paralell to the occlusal surface of the tooth and on which the 2D landmarks were places. In (b) the landmarks located on the re-entrant triangles are not visible due to the cementum-sediment matrix.

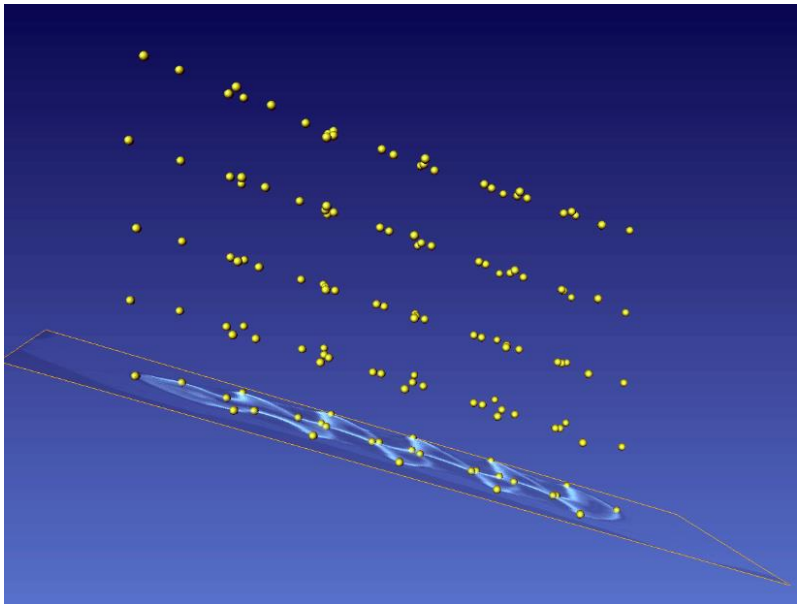


Figure 4.5. The resulting 3D landmark constellation which served as a 3D shape proxy for the *Microtus arvalis* M_1. This results from a series of 3D landmarks collected over 5 ortho-slices and which served as a 3D shape proxy for the *Microtus arvalis* M_1.

length with the tooth orientated so that the buccal surface is uppermost. Once this point had been determined, a series of 5 equally spaced ortho-slice transects from the occlusal surface to the termination transect was taken through the 3D model of the M_1. On each of these an identical series of 28 2D landmark point co-ordinates was obtained (Fig 4.3 - 4.4) to create a final geometric configuration of 3D landmarks to serve as a shape proxy for M_1 3D tooth morphology (Fig 4.5). Due to the need for excellent morphological preservation, the sample size in this dataset for each climatic period was lower than if the occlusal surface alone had been used, on account of the taphonomic processes to which the remains have been subjected. A total of 40 teeth were suitable for analyses; 8 belonging to the Holocene, 12 to the Younger Dryas and 20 to the early part of the Lateglacial (Bølling).

Discriminate function analysis, data augmentation and correlation with climatic variables

Raw 3D co-ordinates were subjected to Procrustes Generalised Least Squares (GLS) superimposition before the aligned shape co-ordinates were decomposed using a co-variance based principal components analysis (PCA). The PCA serves to reduce the total amount of shape variance into a smaller number of composite variables. Projected scores on the PC eigenvectors necessary to retain 95 percent of the observed shape variation were then subjected to Canonical Variates Analyses (CVA), the significance of which were tested with a bootstrapped log-likelihood ratio test. The level of distinctiveness in shape and centroid size were examined in a three group scenario in which the Bølling, Younger Dryas and Holocene periods all represent morphologically distinct sets. Results from the aDNA analyses suggested that a further, and unexpected, second possible grouping, a two-group scenario in which the early Lateglacial individuals form a discrete group separate from the Younger Dryas and Holocene set, which form one, more-or-less morphologically homogenous group.

The PCA dataset produced here exhibits high dimensionality with a large number (>100) of PCs, the majority of which represent only a small proportion of the total

amount of variance contained in the data. When dimensionality is high, the volume of the space it occupies increases so fast that the available data become sparse. With such high dimensional data all objects in the dataset may appear sparse and dissimilar from each other (Macleod 2013). Moreover, the amount of data required to support a result often grows exponentially with the dimensionality of the dataset (termed 'the curse of dimensionality'), creating problems with analytical techniques that rely on searching for common data organisation strategies. The effect of this on CVA can be the production of highly significant discriminant groupings based on high dimensionality PCA scores that are not particularly stable (Macleod 2013). In this study, central to this test of group distinctiveness, was the use of the jack-knife misidentification procedure, which was applied in an attempt to overcome this problem. The jack-knife procedure removes each specimen in turn, codes it as an "unknown", recomputes the CVA without the "unknown" being present in the training-set data, projects the "unknown" into the resultant CVA space, and then determines if it would be identified correctly in the discrimination analysis. This allows a test of the stability of the observed CVA discriminate functions.

Due to the low sample sizes in this analysis, a novel method of artificial data augmentation was used in conjunction with the jack-knife procedure to assess the impact of sample size on the accuracy and stability of the CVA classification (van Dyke & Meng 2001; Tanner & Wong 2010; Macleod, 2016). In this procedure, minor variations in the size and representativeness of the dataset are induced by increasing the sample size (number of cases) and/or copies of the original data in which simulated co-ordinate point locations are adjusted to within x percent of the variance of the original data (Figure CF2). This can vastly improve the robustness of group-characterization in discriminate function analyses by introducing more structure into the analysis in the form of simulations of slight variants of the original data (MacLeod, 2016). Here, a large number of simulations were performed both on the three group (Holocene, Younger Dryas, Bølling) and two group (Holocene & Younger Dryas, Bølling) scenarios. Sample sizes were increased two-fold and five-fold and for each of these increases in sample size 'drift' from the original data points was set both at a variety of small increments; 5 percent, 10 percent, 15 percent and then at larger

increments of 50 percent and 100 percent. The jack-knife procedure was then performed on these simulated datasets to investigate whether increases in sample size and variance from the original data points could lead to more robust between-group discrimination.

An analysis of variance (ANOVA) and Students T-test were used to test for significant differences in log-centroid size (LogCS) over the climatic periods. Lastly, to investigate whether a significant correlation exists between shape variables and climate variables, the mean of the PC scores for each of the three groups (Holocene, Younger Dryas, Bølling) were correlated against the mean temperature North Greenland Ice Core Project (NGRIP) $\delta^{18}O$ Greenland Ice Core Chronology 2005 (GICC05) (Svensson *et al.* 2006, 2008) for the corresponding period.

4.3 Results

4.3.1 Ancient DNA

DNA recovery and phylogenetic analyses

Out of 14 molars that were sampled for ancient DNA, 10 yielded DNA that, when mapped to the common vole cyt *b* gene, was of sufficient quality and quantity to be used in downstream analyses (Table 4.2). Four sequences belonged to the Early Lateglacial (MM099, MM100, MM109, MM110: Bølling, *c.* 15 000 -14 000 cal BP), four to the Younger Dryas cold phase (MM105, MM112, MM113, MM114: 12 900 – 11 600 cal BP) and two to the Early Holocene (MM108, MM115: 11 600 – 8 200 cal BP; Walker *et al.* 2012; Table 4.2). In general, these were low coverage assemblies with an average fragment length under 60bp (after quality trimming and removal of duplicate reads) suggesting that, overall, DNA preservation is relatively poor.

In order to incorporate the vast number of modern and ancient cyt*b* sequences available for *Microtus arvalis*, and to eliminate ambiguities at the start and end of sequences generated here, alignments were trimmed to 1040bp. When ambiguities are excluded, all of these represented unique haplotypes. For all sequences, there were no

nonsense translations and no stop codons present within the sequences. The proportion of variable sites at each codon position showed the typical pattern for protein coding mtDNA genes with most of the variation at third codon positions. Transitions were more common than transversions in all datasets.

Table 4.2. Gully Cave *Microtus arvalis* from which ancient DNA was successfully extracted and sequenced.

ID	Climatic period	Avg. read length after trim	No reads mapped to reference sequence	Average coverage
MM099	Bølling	55.62	88	3.96
MM100	Bølling	59.11	73	1.92
MM105	Younger Dryas	54.24	147	4.95
MM108	Early Holocene	47.23	111	3.76
MM109	Bølling	57.57	129	4.57
MM110	Bølling	57.09	196	9.5
MM112	Younger Dryas	49.42	113	3.24
MM113	Younger Dryas	48.87	125	3.78
MM114	Younger Dryas	60.7	97	5.02
MM115	Early Holocene	56.73	103	5.05

MrBayes returned a 50 percent consensus tree topology that was congruent with the phylogenetic relationships of all the major European clades in previously published studies (*e.g.,* Stojak *et al.* 2015: Fig 4.6). All estimated sample sizes (ESS) were greater than the recommended value of 200 and the Potential Scale Reduction Factor (PSRF) for all parameters was >1. Chain convergence, analysis of splits and all other MCMC outputs were examined in TRACER v.1.5 (Rambout *et al.* 2014) and AWTY (Wilgenbusch *et al.* 2004). No measures of genetic diversity were calculated for this dataset as only unique haplotypes were considered (which bias the statistical estimates). A well-supported (posterior value = 1) split was observed between the monophyletic Western clade and the large clade representing the Central, Eastern, Italian and Balkan lineages. The posterior probability value of the split between the two major sublineages, Western-North and Western-South, was moderately high at 0.89. All of the Gully Cave British samples were placed well within the Western-North group and formed a large, unsupported polytomy with samples from Orkney (modern

and ancient), France, Germany, Luxembourg, the Netherlands and Belgium (modern and ancient).

A database of 388 haplotypes that had been assigned to either the Western-North or Western-South clade was used to create the Median-Joining network in POPart (Figure 4.7) with measures of genetic diversity calculated in DNAsnp (Rozas 2010). Excluding the 35 sites with gaps or missing data, there were a total number of 153 variable (informative) sites (*S*) and 167 mutations. Unique haplotype number (*h*) was 147 with a haplotype diversity index (*Hd*) of 0.962. Nucleotide diversity per site (π) was 0.00959, a low value possibly reflecting the large number of duplicate haplotypes used in this database. The network recovered the well-defined split between the Western-North and Western-South clades (Fig 4.7) with a minimum distance of nine mutations separating out these lineages. As in the Bayesian analyses, the Gully Cave samples were all placed well within the Western-North lineage, constituting strong evidence that the north-western coast of Europe was the geographical source area for both Lateglacial and Holocene populations of the British common vole. The Western-North lineage in the network is characterised by two paraphyletic star-shaped clusters, one predominantly including the ancient and modern Orkney vole populations and the other predominantly representing specimens from France, Belgium (ancient and modern), Luxembourg, the Netherlands and West Germany, although these are not monophyletic. Both of the Holocene Gully Cave specimens (MM108 and MM115) and the majority of the Younger Dryas samples (MM105, MM113 and MM114) were placed in the large star-shaped cluster that principally includes specimens from modern and ancient (Neolithic onwards) Orkney. Interestingly, samples from the Bølling (MM099, MM100, MM109, and MM110) do not group with the rest of the Gully Cave and Orkney samples but instead appear to be clustered together, and separately, branching with six similar haplotypes from Germany and France. The exception to this is MM112, which belongs to the Younger Dryas but which also groups in this branch. Specimen MM112 possesses three fixed mutations that are found only in samples from the early Lateglacial: a C at position 378, a T at position 519 and an A at position 619.

The BEAST analysis contained a total of 165 individuals representing the Western-North clade only and included duplicate haplotypes so as not to skew the demographic analyses. There were 119 variable sites in this dataset exhibiting a total of 126 mutations. Eighty-four haplotypes were unique with an *Hd* index of 0.97 and π value of 0.0066. Bayesian genealogy sampling was carried out with a calibrated molecular clock rate of 3.27×10^{-7} (Martínková *et al.* 2013; Stojak *et al.* 2015) using the following parameter combinations: *i*) strict clock with constant population size, *ii*) relaxed log-normal clock with constant population size, *iii*) strict clock with a coalescent exponential population size and *iv*) relaxed log normal clock with a coalescent population size. 'Null' simulations, where the BEAST simulation samples only from the prior and not the data, all had ESSs < 200 indicating that the results obtained here are driven by the data and not the priors (Appendix C, Table CT3). All ESS's for simulations sampling from the data were > 200 (Appendix C, Table CT3). Bayes factors calculated for each of these analyses preferred scenario '*iv*' with a relaxed log normal clock and a coalescent model of population size. However, the aim of this analysis was to test the effect of adding time calibrated data on the tree phylogeny and not to test demographic models of population history. All BEAST maximum clade credibility (MCC) trees for all demographic scenarios returned identical topologies, the MCC tree presented is from scenario *iv* (Fig 4.8). This tree exhibits a lack of structure with extremely low support for any sub-lineages within the Western-North clade. The exception to this trend was the high level of support for the Bølling samples forming a discrete, basal lineage to the rest of the clade with samples from ancient Belgium, France and Germany (*i.e.,* the same individuals the Bølling samples clustered with in the network analyses; Fig 4.7). Interestingly, the Younger Dryas individual MM112 was also placed in this discrete sub-lineage.

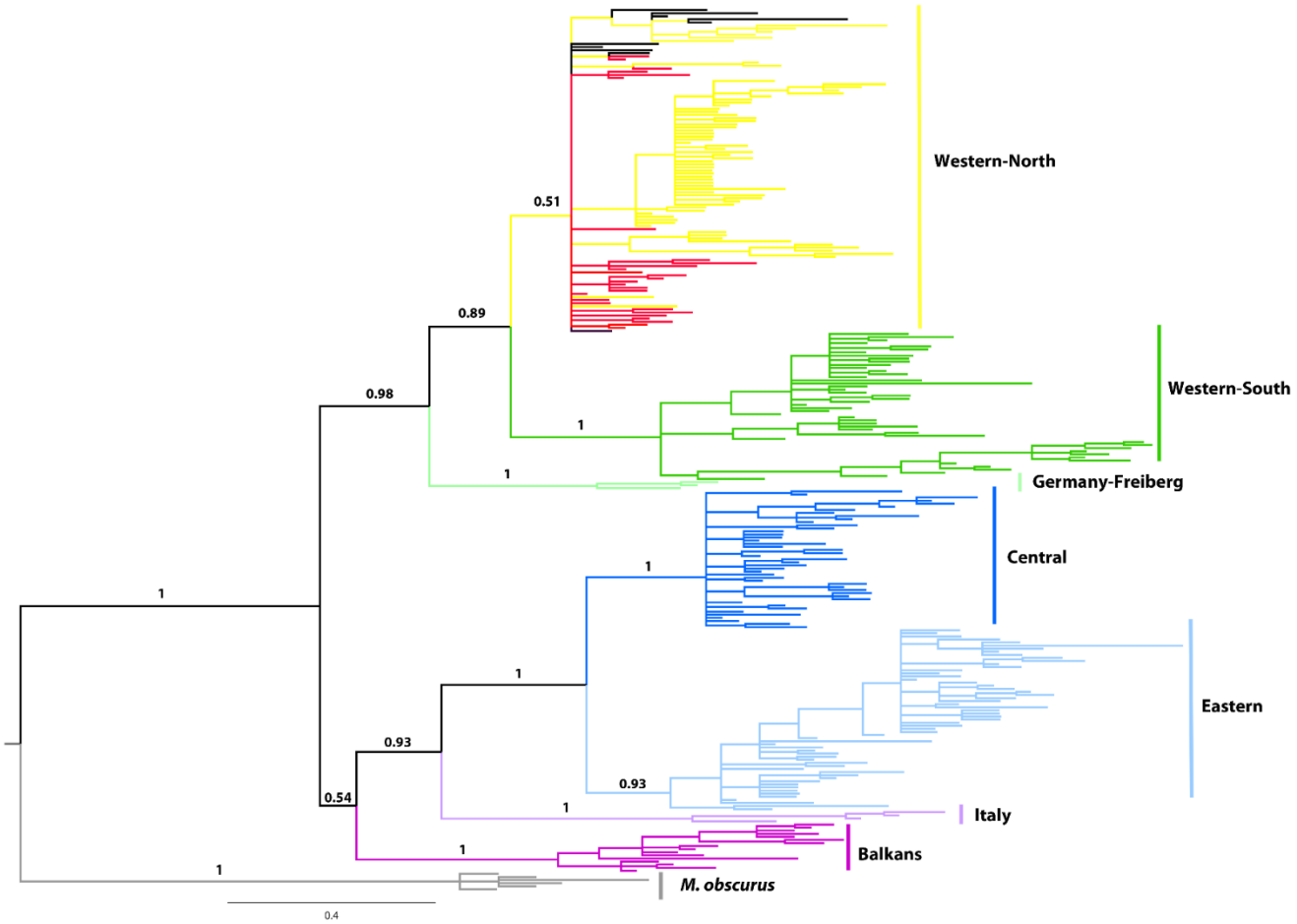


Figure 4.6. Cytochrome *b* phylogeny (1040bp) generated in MrBayes showing relationships among the European lineages of the common vole *Microtus arvalis*. A 50% consensus tree with posterior support values for major lineages is shown, rooted with *Microtus obscurus*. High posterior support was recovered for the broad division between the Western clade and the clade representing the Central, Eastern, Italian and Balkan clades. High support was also recovered for all major sublineages. Orkney cyt*b* sequences are shown in red and British mainland Gully Cave sequences in black. All Gully Cave sequences place within the large polytomy of the Western-North lineage.

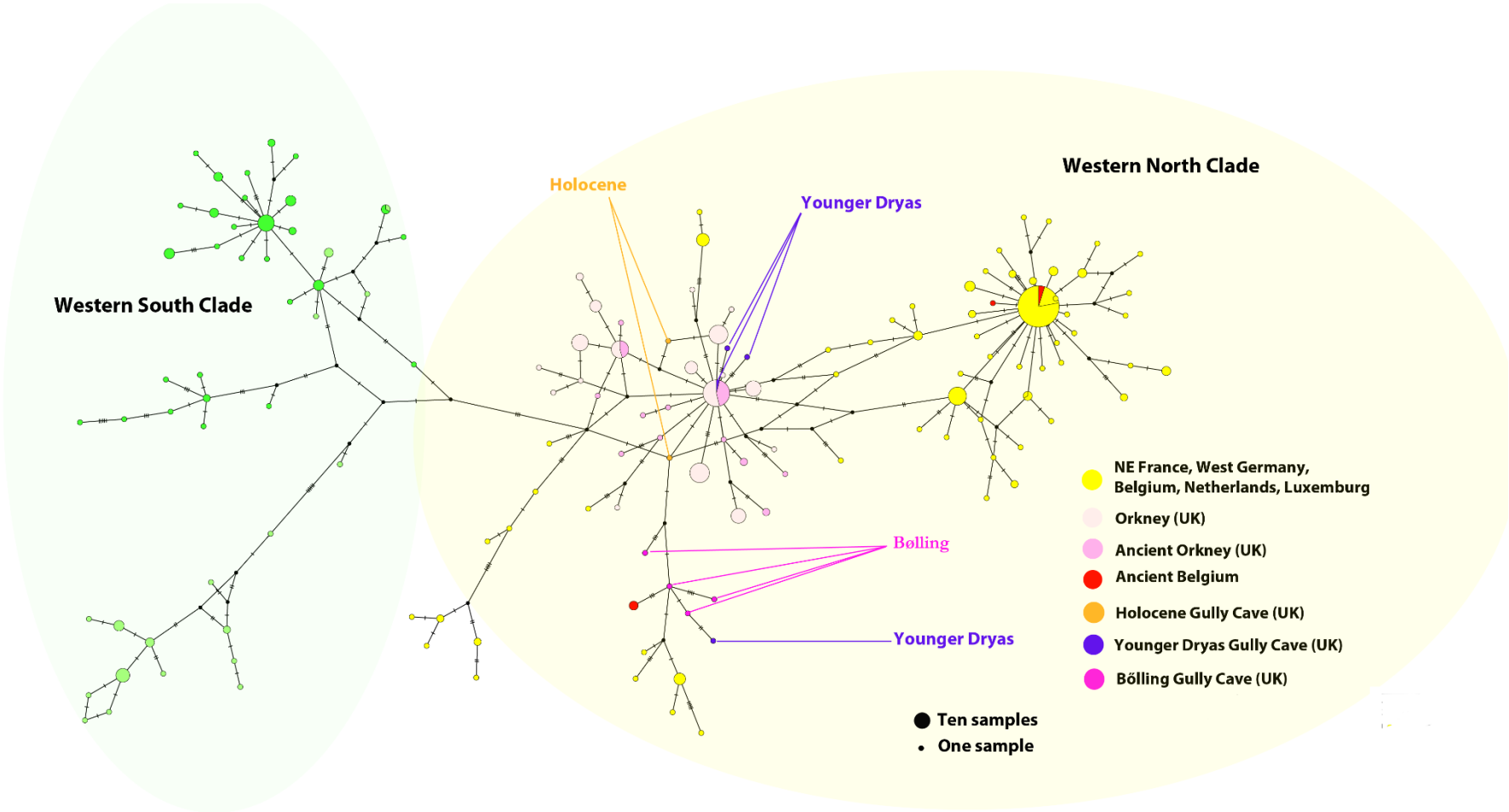


Figure 4.7. Median-Joining (MJ) network generated in POPart for the Western-North and Western-South lineages using partial cyt*b* sequences (1040bp). Relationships are shown between the Gully Cave Holocene, Younger Dryas and Bølling samples and populations from the Western-North clade. Holocene and Younger Dryas samples group with the ancient and modern Orkney samples while the Bølling samples clustering together.

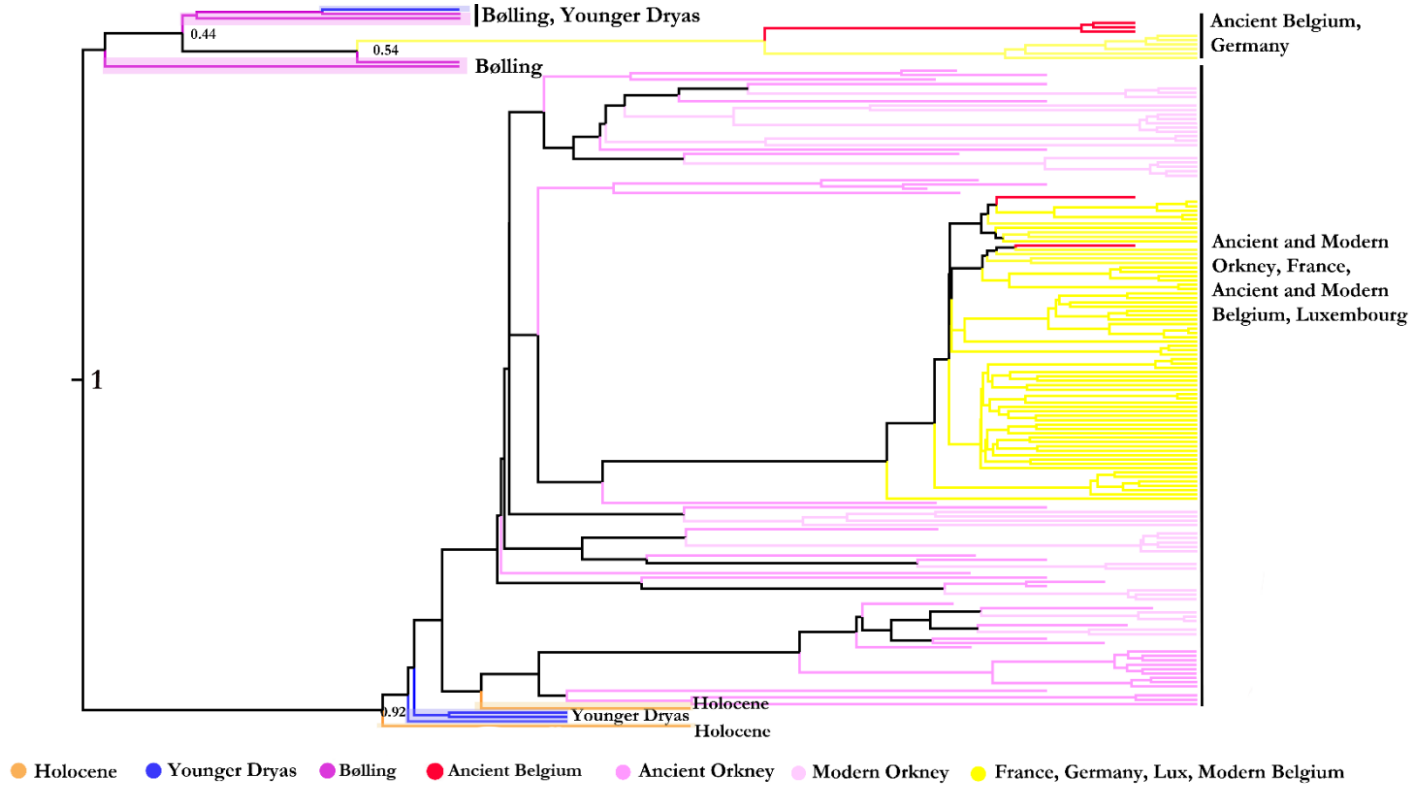


Figure 4.8. Time-calibrated maximum clade credibility (MCC) tree generated in BEAST showing intraspecific relationships among the Gully cave cyt*b* (1040bp) sequences and cyt*b* sequences from Orkney and the north-west coast of Europe. The genealogy was generated using a molecular clock rate of 3.27 x 10^{-7} (Martínková *et al.* 2013) and uses a relaxed log normal clock with a coalescent population expansion prior (although topologies did not differ with other priors). The only high nodal support which was returned was for a split between the Bølling haplotypes and the remainder of the lineages.

The AMOVA showed that the majority of variation in the dataset was within populations (83.17 percent), with only 16.83 percent of variation between populations. Futhermore, no significant between-population differentiation (p = > 0.05 for all pairs) could be identified, suggesting chronological gene flow between these groups.

4.3.2 3D Geometric morphometrics and ecomorphology

The covariance-based PCA of 40 Procrustes aligned shape co-ordinate configurations matrix for *Microtus arvalis* M_1 molars generated a total of 40 eigenvectors with non-zero eigenvalues. The first eigenvector in this series represented 44 percent of the total shape variation observed in these data but did not result in any substantive clustering of discrete climatic groups. This result was not unexpected, as PCA is not a statistical tool designed specifically for detecting between-group separation. Rather, it is designed to reduce the dimensionality of a dataset of potentially correlated variables into a set of linearly uncorrelated variables that are aligned orthogonally to each other. Here, the principal components are used to reduce the dimensionality of the 3D *Microtus arvalis* landmarks into a smaller number of discrete variables that can then be used as the basis for a group discrimination analyses with Canonical Variates Analyses (CVA). However, the absence of any differentiation in M_1 shape in the PCA shape space suggest shape differences are very subtle.

The first 20 PCA eigenvectors retained 95 percent of the total shape variation present in the dataset. The scores of landmark configurations projected onto these eigenvectors were used as the basis for the CVA. This resulted in 2 CV axes (representing 67.09 percent and 32.91 percent of the total variation respectively; Fig 4.9) for the three-group scenario and 1 CV axis (representing 100 percent of the total variation) for the two-group scenario. The log-likelihood ratio was non-significant for the three group scenario (p = 0.054; Appendix C, Fig CF1-a), although only marginally so. While this would technically accept the null hypothesis, this result may be an artefact of low sample size and problems created with high dimensionality in the PCA space. The log-likelihood ratio test was significant at the 0.05 level for the two-group scenario (p = 0.025, Appendix C, Fig. CF1-b).

In the three group scenario (Bølling, Younger Dryas, Holocene), the groups appear reasonably well differentiated in the CV-space (Fig. 4.9) with CV-1 involved in separating out the Holocene and Younger Dryas groups from the Bølling and CV-2 involved in separating out the Holocene from the Younger Dryas and Bølling. The test of group distinctiveness (using linear distances of each point to the most proximal group centroid, see Chapter 3, Section 3.2.1 for a description of the calculation) also showed good support for each group with the Holocene having 87.5 percent of specimens correctly assigned to the group, the Younger Dryas having 91.66 percent, the Bølling 90 percent and the overall group total 90 percent (Table 4.3a). However, these percentages were significantly reduced when the jack-knife stability test was performed; the Holocene correct specimen assignment dropped to 25 percent, the Younger Dryas 58.3 percent, the Bølling 45.00 percent and the overall total 45.00 percent. Thus, while the CVA analysis alone would infer strong between-group differentiation the reduction in correct group assignment in the jack-knife results strongly suggests the influence of problems associated with high dimensionality (*i.e.,* unstable CV axes).

The two-group CVA exhibited marginally better group distinctiveness percentages with The Holocene + Younger Dryas having 95 percent of specimens correctly assigned to the correct group and the Bølling having 90 percent with the overall total at 92.5 percent (Table 4.3b). The jack-knife results were considerably better here with the Holocene + Younger Dryas having 73.68 percent of specimens correctly assigned and the Bølling having 60 percent with a group total of 65.00 percent (an 18 percent increase over the three group scenario; Table 4.3b). This may be an artefact of simply having a smaller number of groups (as each specimen will inevitably have to assigned to either one group or the other) and a larger sample size in the Holocene + Younger Dryas group. Furthermore, these values are not much larger than what would be expected by chance alone. No significant results were returned for size differences in the ANOVA for the three-group scenario (Fig 4.10a, $d.f$ =2, F-stat 1.33, p = > 0.05) or in the two-tailed Students *T*-test for the two-group scenario (Fig 4.10b, p = > 0.05)

Data augmentation and climate correlation

After data augmentation, the linear distance group assignment procedure produced 100 percent correct and significant specimen assignments for all datasets in the CVA space. Therefore, only results for the jack-knife analyses are presented here. The augmentation of the datasets to increase both the sample size and the 'drift' percentage of those new variables from the original data increased the jack-knife percentages in all cases (Fig 4.11, Tables 4.4 a and b). Overall, group assignment percentages tended to be higher when the drift percentage was under 50 percent and dropped considerably after this threshold. Correct assignment percentages at 100 percent did not show any improvement over smaller drift values and in some cases were lower. In terms of sample sizes, the 5-fold increase, as expected, generally yielded better group assignment rates than the 2-fold increase. The two-group (Holocene + Younger Dryas) scenario was generally better supported than the three group (Holocene, Younger Dryas, Bølling), as was observed in the CVA space.

While the data augmentation exercise does not assist in selecting whether a three group or two group scenario is more likely, it does strongly suggest that augmenting the datasets significantly improves stability in CVA space. If the assumptions associated with this form of augmentation are accepted (that the augmented dataset is representative of the population) then there is strong evidence to support M_1 shape change between climatic periods. The pattern of shape change and the implication of these results in elucidating patterns of population turnover associated with the climatic intervals remain somewhat ambiguous based on these analyses.

The mean NGRIP GICC05 $\delta^{18}O$ values for the Early Holocene (11 600 cal BP – 8 200 cal BP), Younger Dryas (12 900 cap BP – 11 600 cal BP and Bølling (14 000 cal BP – 15 000 cal BP were -35.63, -40.38 and -38.77 respectively. The mean PC1 scores for each group were: Holocene -0.0052, Younger Dryas -0.014, Bølling -40.38. A correlation coefficient (r) between these variables of 0.5 (p = > 0.05) was returned, indicating a moderately positive relationship between them. A three-point correlation contains the minimum number of points with which to carry out this calculation.

Therefore, it is not possible to be certain that a true linear relationship exists between mean PC scores and $\delta^{18}O$ or to investigate this relationship further.

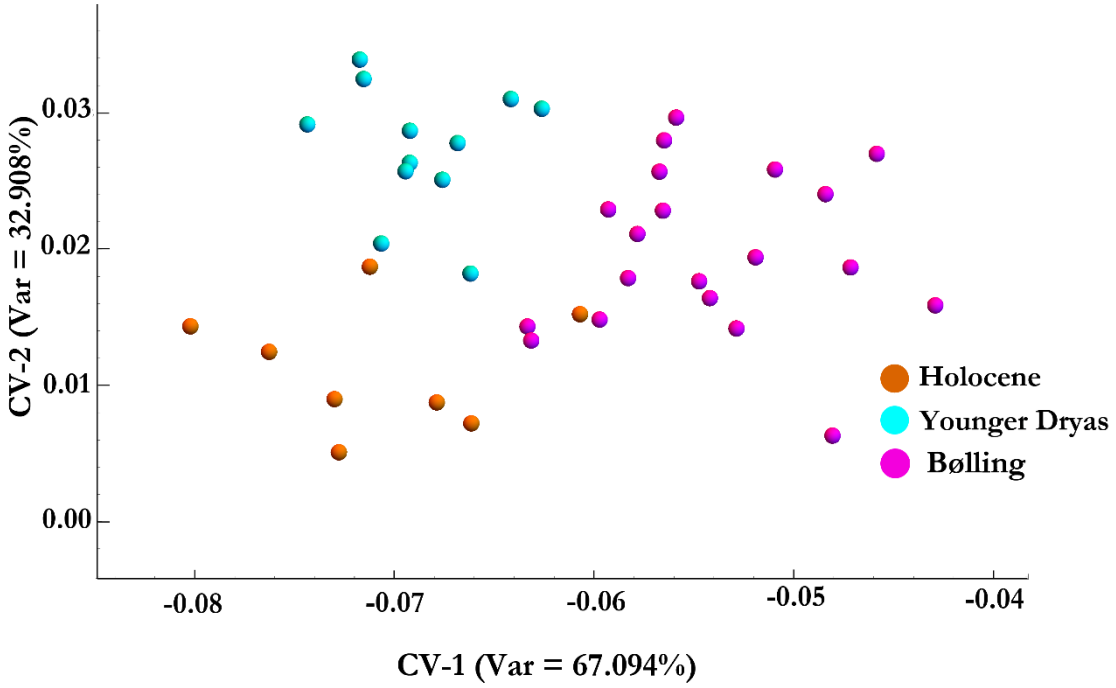


Figure 4.9. Canonical Variates 1v2 for the three-group scenario. Shown is the maximum between-group separation and minimum within group variance for the three climatic phases.

Table 4.3 (a) and (b). Linear distance analyses of group distinctiveness and jack-knife percentages (in brackets) for the (a) three-group scenario and (b) two-group scenario.

(a)

Groups	Holocene	Younger Dryas	Bőlling	Total Correct	Group Totals	Percent Correct
Holocene	7 *(2)*	0 *(3)*	1 *(3)*	7 *(2)*	8 *(8)*	**87.5 *(25)***
Younger Dryas	1 *(2)*	11 *(7)*	0 *(3)*	11 *(7)*	12 *(12)*	**91.66 *(58.3)***
Bolling	2 *(4)*	0 *(7)*	18 *(9)*	18 *(9)*	20 *(20)*	**90 *(45.00)***
Total Correct	**7 *(2)***	**11 *(7)***	**18 *(9)***	**36 *(18)***	**40 *(40)***	**90 *(45.00)***
Total Estmated	**10 *(8)***	**11 *(17)***	**19 *(15)***	**40 *(40)***		
Percent Estimated Correctly	**70 *(25)***	**100 *(41.17)***	**94.74 *(60)***	**90 *(45.00)***		

(b)

Groups	Holocene + Younger Dryas	Bőlling	Total Correct	Group Totals	Percent Correct
Holocene + Younger Dryas	19 *(14)*	1 *(5)*	19 *(14)*	20 *(19)*	**95 *(73.68)***
Bolling	2 *(9)*	18 *(12)*	18 *(12)*	20 *(20)*	**90 *(60.00)***
Total Correct	**19 *(14)***	**18 *(12)***	**37 *(26)***	**40 *(40)***	**92.5 *(65.00)***
Total Estmated	**21 *(23)***	**19 *(17)***	**40 *(40)***		
Percent Estimated Correctly	**90.48 *(60.87)***	**94.74 *(70.58)***	**92.5 *(65.00)***		

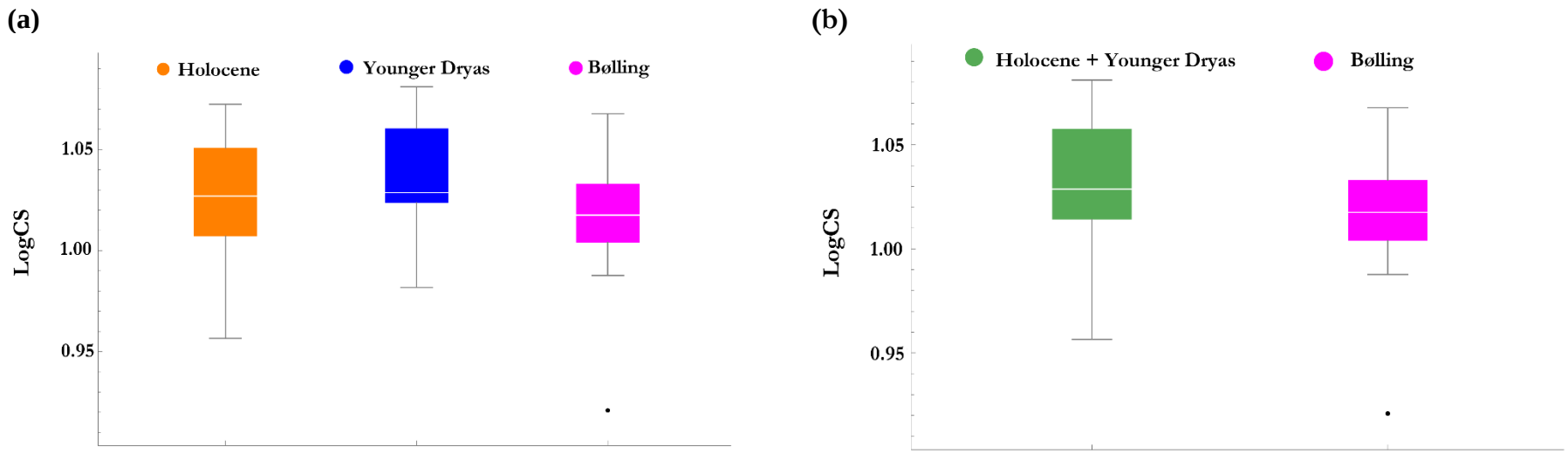


Figure 4.10a, b. Boxplots of Log Centroid Size for the three (a) and two (b) group scenarios. An Analyses of Variance (ANOVA) was performed to test for significant differences in LogCS size between the three main climatic groups but returned a non-significant result. A Student's T-test on significant difference of mean LogCS size also returned a non-significant result for the two-group scenario.

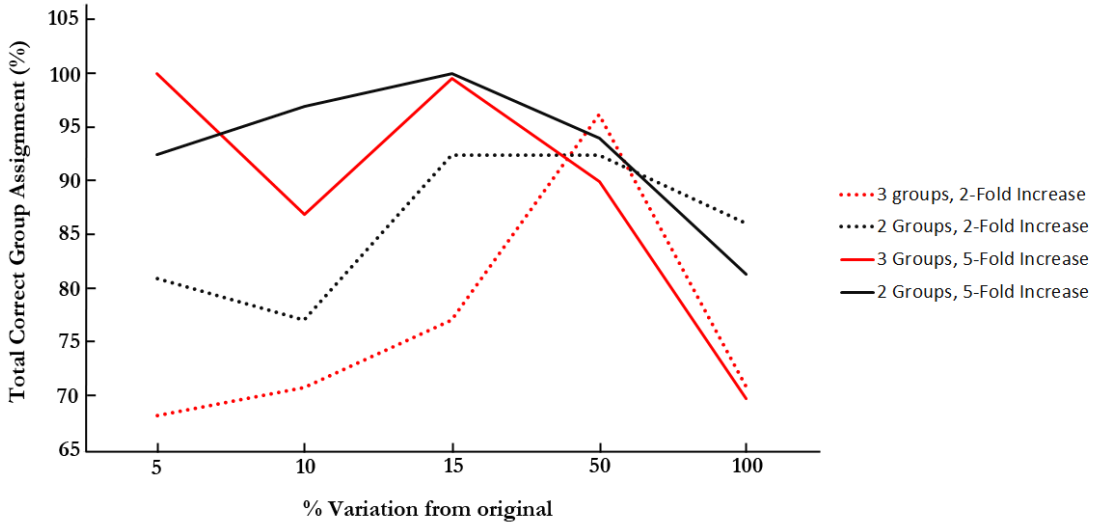


Figure 4.11. Jack-knife group assignment percentages for *Microtus arvalis* 3D landmarks. Red lines represent the 3 group scenario and black lines represent the 2 group scenario.

Tables 4.4a, b. Jack-knife percentages for *Microtus arvalis* 3D landmarks. (a) shows results for the 3 group scenario and (b) shows results for the two group scenario.

(a)

	Increase in sample size									
	2-fold					5-fold				
Drift	**5**	**10**	**15**	**50**	**100**	**5**	**10**	**15**	**50**	**100**
Holocene	68.75	68.75	62.5	93.75	62.5	100	75	100	92.5	62.5
Younger Dryas	75	79.16	83.3	100	70.83	100	91.6	100	86.66	71.67
Bølling	64.1	66.66	79.48	94.87	74.36	100	88.8	98.98	90.9	71.71
Total Correct	**68.35**	**70.88**	**77.21**	**96.2**	**70.88**	**100**	**86.9**	**99.49**	**89.95**	**69.85**

(b)

	Increase in sample size									
	2-fold					5-fold				
Drift	**5**	**10**	**15**	**50**	**100**	**5**	**10**	**15**	**50**	**100**
Holocene & Younger Dryas	80	72.5	95	92.5	90	90	100	100	93	82
Bølling	82.05	82.05	89.74	92.31	82.05	94.95	93.94	100	94.95	80.8
Total Correct	**81.01**	**77.21**	**92.4**	**92.41**	**86.07**	**92.46**	**96.98**	**100**	**93.97**	**81.4**

4.4 Discussion and main findings

The broad aims of this chapter were to investigate population continuity/discontinuity over the Lateglacial to Early Holocene period in the common vole *M. arvalis,* using a combination of evidence from ancient DNA and 3D geometric morphometrics. Specimens were excavated from the stratigraphic layers at a single site in South-West England - Gully Cave, Ebbor Gorge, Somerset. Three-dimensional landmarks were used as a proxy for the shape of the lower first molar (M_1) and Log Centroid size (LogCS) as a size proxy. Group discrimination analysis, the robustness of which was tested with jack-knife simulations, was employed to detect structured variation in aspects of M_1 shape and/or size between different age groupings defined *a-priori* as either *i*) Holocene, Younger Dryas, Bølling or ii) Holocene + Younger Dryas, Bølling. Ancient DNA data obtained by NGS shotgun sequencing were mapped to the cytochrome *b* gene before being combined with existing modern and ancient data for this species to examine both patterns of continuity/replacement through the Gully Cave sequence and their phylogeographic affinity to the main European lineages.

The main findings in the chapter were:

- The Bayesian phylogenetic analysis and the Median-Joining (MJ) network robustly place the British Gully Cave samples within Western-North lineage
- Cytochrome *b* sequences from the Bølling clustered separately from the Younger Dryas and Holocene sequences in both the MJ network and the BEAST simulations
- No significant population differentiation was detected by an AMOVA between Holocene, Younger Dryas and Bølling populations
- No significant difference in M_1 size was detected over the climatic periods in either the two or three group subdivision scenarios
- Marginally non-significant shape differentiation exists among the three age/climatic groups for the *M. arvalis* M_1, which was moderately but positively correlated with climate

- A significant M_1 shape differentiation result was returned for the two age/climatic group scenario
- Data augmentation tests suggest that, for both two-group and three-group scenarios, a five-fold increase in the M_1 sample size would result in an improvement of the group-discrimination results with the most reasonable simulations indicating these age/climatic groupings are significantly different in terms of overall tooth shape

4.4.1 Ancient DNA

During the early Lateglacial, the North Sea basin landbridge (Doggerland) would have allowed post-glacial recolonisation of Britain from a number of geographical areas along the north-west and south-west European coastline (Clark *et al.* 2012). Colonisation events from refugial populations in the Iberian Peninsula, from France/Belgium or from German Central lineage populations were all therefore possible. The phylogenetic analyses carried out in this study show, for the first time, that British populations of the common vole belong to the large Western European clade and, specifically, to the Western-North sub-lineage of this group. The Lateglacial source population was, therefore, certainly from the NW coast of France or Belgium and in the event of subsequent recolonisation waves as a result of subsequent abrupt climatic changes, these also originated from this region.

Authors differ on the age of the 'time to most recent common ancestor' (tMRCA) of the main European lineages, depending on whether fossil or molecular clock calibrations have been used (Tougard *et al.* 2008; Stojak *et al.* 2015). Tougard *et al.* (2008), using fossil calibrations obtained tMRCAs for the common vole lineages of between 317 000 and 75 000 years BP, while Stojak *et al.* (2015), using molecular clock estimates, obtained much younger tMRCA dates (detailed below). As many authors report that using fossil estimates placed deep in the nodes of a phylogeny can be misleading in intraspecific phylogenies (Herman & Searle 2011; Ho *et al.* 2011) the tMRCAs obtained by Stojak *et al.* (2015) should be preferred, at least provisionally. This investigation found that all six common vole lineages have similar measures of genetic diversity, suggesting they all experienced a demographic expansion at around

10 000 cal BP. The main lineages also all have similar tMRCAs with a median range of 12 482 cal BP (Italian) to 23 735 cal BP (Western-North); albeit with large confidence intervals (Stojak *et al.* 2015). This suggests that the British samples, as constituents of the Western-North clade, have a history consistent with LGM refugia in the Mediterranean peninsulae, prior to population re-expansion and re-colonisation of Europe.

The MJ network showed two star-shaped phylogenies within the Western-North clade, one composed mainly of ancient and modern cyt*b* samples of the Orkney vole and a second composed mainly of cyt*b* sequences from France, Germany, the Netherlands, Luxembourg and Belgium (Fig. 4.7). Star-shaped phylogenies are a signature of a rapid demographic expansion with little geographic structuring (Slatkin and Hudson, 1991). This is congruent with the findings of Stojak *et al.* (2015) that all common vole lineages experienced large population expansions at around, or just prior to, 11 700 cal BP (*i.e.,* just prior to, Holocene warming). As a grassland specialist it seems counter-intuitive that this species would experience an expansion during a period associated with reforestation. However, during the early Holocene open grassland ecotypes are known to have persisted in Europe on fine-grained soils where seasonal dryness prevented the establishment of woodland (Theuerkauf *et al.* 2013). The placement of all of the Holocene common vole sequences and the majority of the Younger Dryas sequences within the large star-shaped phylogeny of the Orkney vole suggests that these populations were a part of this rapid expansion, and while also indicating that Orkney and British mainland populations shared a common ancestry.

Martínková *et al.* (2013), using cyt*b* and Approximate Bayesian Computation (ABC), established the most likely origin of the (introduced) Orkney population as the NW coast of Belgium. The close phylogenetic relationship of the Younger Dryas and Holocene age Gully Cave haplotypes to the Orkney sequences begs the question of whether the British mainland may have served as an alternative source population for a natural 'sweepstake' colonisation of the Orkney archipelago. However, the data collected for this study do not provide sufficient resolution to answer to this question. While cyt*b* has proved informative for resolving the phylogenetic and population-level

history at the intra-specific level for a number of rodent species (*e.g.,* wood mouse *Apodemus sylvaticus* Michaux *et al.* 2003; red squirrel *Sciurus vulgaris* Grill *et al.* 2009; and field vole *Microtus agrestis* Herman *et al.* 2014), it may obscure some finer-scale population processes due to incomplete lineage sorting, hidden introgression events, site saturation and/or the matrilineal pattern of inheritance (McGuire *et al.* 2007). To examine the alternative theory of Orkney colonisation from mainland Britain, inclusion of whole mitogenome, a range of nuclear markers (particularly from the Y-chromosome that would allow sex-specific contributions to be investigated) and more sensitive analytical techniques such as coalescent modelling would need to be employed.

Both the Median-Joining network and Bayesian genealogy sampling in BEAST showed the Gully Cave Bølling sequences branching separately from both the Gully Cave Holocene and Younger Dryas haplotypes and from the remainder of the Western-North clade. In both analyses, these sequences group with a subset of samples which have geographic locations towards the North-East of the geographical range covered by the North-West clade but which show no phylogenetic affinity to the most proximal (Central) lineage. The results obtained here were unexpected as the most obvious scenario of population turnover would have been to observe genetic differences between the Lateglacial interstadial, Younger Dryas and Holocene, all of which represent distinct phases of climatic development (Svensson *et al.* 2008). However, all of the Lateglacial interstadial samples used in this study belong to the Bølling (*c.* 15 000 – 14 000 cal BP), which occurs between two cold phases: the Oldest Dryas (ending *c.* 15 000 cal BP) and Older Dryas (commencing *c.* 14 000 cal BP). It is possible that the South-West of England was recolonised by the common vole following the Oldest Dryas where populations had little opportunity to become established prior to the onset of the brief Older Dryas climatic interval. This, is turn, may have caused these founder populations either to become locally extinct or to track their habitat back to the mainland from where they subsequently recolonised and became established during the Allerød interval. However, the AMOVA did not support a scenario chronological population differentiation between these putative

populations. It is likely that further sampling would be required to clarify the genetic relationship between these groups.

4.4.2 Geometric morphometrics and ecomorphology

The 3D geometric morphometric results are somewhat more difficult to interpret. The (marginally) non-significant result in the CVA analyses which was correlated with mean NGRIP GICC05 $\delta^{18}O$ values for each period (Holocene, Younger Dryas, Bølling), suggests that there was ecomorphological M_1 shape change that can be related to the abrupt climatic patterns of warming and cooling. However, M_1 shape did not show any distinct replacement in the form or size variation of the tooth over climatic phases, which would have been indicative of a sudden population turnover event. The scenario suggested by the aDNA results was tested by using a two-group scenario in which the Holocene and Younger Dryas-age samples were pooled and the Bølling-age samples retained as a distinct group. This group structure was better supported, but only marginally.

The lower first molar of voles has been used in a number of studies addressing climate-related effects on tooth shape (*e.g.,* clinal variation McGuire 2010, evolutionary change over the Quaternary Piras *et al.* 2008 and island evolution Cucchi *et al.* 2014). This tooth is known to have distinct anatomical areas that can change rapidly in relation to a changing climate (Chaline *et al.* 1999; Abramson *et al.* 2009) and population replacement has been inferred from morphological change alone in a variety of studies (*e.g* Barnosky *et al.* 2003). However, incongruence between morphological and genetic evidence has been commonly observed. Monophyletic groups as defined by mitochondrial DNA may take many thousands of years to form while tooth morphology may change rapidly over a much smaller timescale. Conversely, morphological stasis has been observed in many species, including voles, which show considerable genetic structuring (Piras *et al.* 2008). Due to the differential factors acting upon the genotype and phenotype, it is not surprising that the morphogical and genetic results obtained here are not in complete agreement.

No significant results were returned for differences in size (inferred from a proxy – LogCS size) over the climatic phases under either the two-group or three-group scenarios. This suggests that the subtle differences in M_1 shape observed through the sequence are not related to allometric effects. Climate change can influence morphology by inducing physiological responses (such as is seen in species that follow Bergmann's Rule; McNab 1971) and additionally by the creation of new ecological opportunities via modification of a species' habitat which, in turn, can initiate morphological adaptation (Renaud *et al.* 2005). Changes in vole molar size have been observed to be key indicators of adaptation to environmental change (Renaud *et al.* 2005) and can occur over millennial timescales. For example, Cucchi *et al.* (2014) observed a rapid size increase in the Orkney common vole subspecies (*M. a. orcadensis*) related to evolutionary pressures associated with isolation on offshore islands. If the Gully Cave population was experiencing size-related adaptation or developmental changes in relation to abrupt climatic change, it is reasonable to expect that these would have been observed in this dataset. However, Gully Cave is situated in very close proximity to the geographical source population for the British common vole (the NW coast of Europe). If population replacement events occurred, it is likely that founders would have originated from populations situated in very similar environments to those found in the south-west of England. These may have had a pre-existing morphology and molar size very similar to those of the Gully Cave population, essentially obscuring any population turnover events. This, however, remains speculative.

Low sample size is a pervasive problem when applying geometric morphometric techniques to zooarchaeological material both because of the overall scarcity of material and the small proportion of material that does exist and which exhibits an adequate degree of structural preservation. This problem is exacerbated when attempting to use the 3D morphology of a small structure such as the lower first molar of voles. While Cardini and Elton (2007) have shown that mean size, standard deviation of size and shape variance are fairly impervious to the effects of low sample size, parameters such as mean shape and angles between static allometric trajectories have been identified to be strongly affected (Cardini and Elton 2007). Furthermore,

discriminate function analyses using geometric morphometric landmark data at the intraspecific level often leads to high dimensionality datasets, which can cause misleading group separation within CVA. The use of data augmentation in this study demonstrated the utility of this novel technique in assessing the likely effects of different sample size increases on the CV and jack-knife misidentification procedures. While this does not change interpretations of the empirical discrimination results, it does allow an exploration of the effects of sample size on the discriminant function analyses.

Interpretation of the results of data augmentation should be treated with caution due to implicit assumptions that may not always be met, *i.e. i*) that the original sample is representative of the population, *ii*) that population values are distributed normally and *iii*) variances for each variable are equal (Macleod, 2016). These analyses repeatedly gave higher support to the two-group scenario of population turnover after the Bølling over the scenario of three, discrete groupings but both scenarios were reasonably well supported. Taken in isolation without reference to the ancient DNA results, it is not possible to form a firm conclusion about the population history of *M. arvalis* at Gully Cave from these results alone, other than climate-correlated M_1 shape change can be observed. Further clarification on this would require larger sample sizes and finer resolution in the age determinations of each individual than was possible here.

The common vole was considered to have disappeared from mainland Britain in the Holocene (Corbet 1961, Yalden 1999) and the evidence presented here (and in Chapter 3) is the first confirmation that this species persisted in Britain at this time. The species has subsequently become regionally extinct throughout Britain, which raises the question as to why this should have occurred. Suitable habitat within the known latitudinal range would have been available and no ecological competition exists with conspecifics. While the field vole *Microtus agrestis* is present in Britain and occupies a similar ecological niche, these two species co-exist in sympatry throughout central Europe where they partition use of the environment (Jurdíkova *et al.* 2000; Santos *et al.* 2001). There is evidence that Holocene Britain was an unstable period for species that had recolonised from mainland Europe. A series of minor cold events

occurred across the early stages of this period that may have affected mammal populations in addition to the, potentially turbulent, process of the establishment of new mammalian communities (Alley *et al.* 1997; Crees *et al.* 2016). From the Neolithic onwards large-scale human-mediated changes in land-use have been associated with significant modification of the genetics of the common vole with Martínková *et al.* (2013) reporting large-scale cyt*b* replacement in this species in the Late Holocene. The results obtained by this investigation show that abrupt climatic changes did have an effect on common vole M_1 morphology and populations but also suggest that intriguing, and yet undetected, phylogeographic and population-level patterns may exist for this species in Holocene Britain.

Chapter 5: Resolving the history of the Eurasian beaver *Castor fiber* in Britain.

5.1. Introduction

5.1.1 Overview

The Eurasian beaver *Castor fiber* is a typical warm-stage taxon with an ecology heavily reliant on riparian woodland. Its ability to significantly modify the environment via tree-felling, dam building and localised flooding indicates that it would have been a key element of the Palaearctic ecosystem during temperate phases of the Pleistocene and throughout the Holocene. The species has high dispersal potential, being shown to rapidly colonise new habitat as it becomes available (Hartman 1996; Nolet & Baveco 1996), and survives along watersheds wherever appropriate wooded areas are present. Thus, *C. fiber* has a vast natural range throughout Eurasia with a longitudinal distribution extending from Great Britain through to Eastern Siberia, and a latitudinal distribution from the Mediterranean to the far north of Scandinavia (Halley & Rosell 2002).

Its ecological niche makes it unlikely that the species persisted in cryptic Northern refugia during the LGM and beaver populations probably followed the general pattern of isolation in refugia south of the ice sheet often observed in most medium to large bodied mammals (Hewitt 1999, 2000, 2004). *Castor fiber* was extirpated from Britain by the 16th century (Manning *et al.* 2014) and only a small number of officially recognised populations exist in Britain today, all of which are introductions from European stock. Larger-scale re-introductions of beaver from European populations are under discussion among major conservation bodies for whom a key consideration is identification of an appropriate European source population. Ancient DNA can therefore play a key and timely role in informing conservation biologists in this endeavour via identification of extant European populations most closely related phylogenetically to the ancient (and now extinct) British population.

These attributes make the species both an excellent model with which to *i*) examine the effects of rapid climate change on the phylogeography and population structure

of a Palaearctic mammal over the Pleistocene – Holocene boundary in Britain and *ii*) use this information to assist vital conservation work associated with re-establishing the Eurasian beaver in Britain.

The broad aims of this chapter are:

1. To determine the phylogenetic placement of ancient British beavers within an existing European phylogeny
2. Establish the geographic source populations from which Late and Postglacial recolonisation occurred, and
3. Investigate population continuity over the Younger Dryas cold interval

5.1.2 Phylogenetic structure and subspecific taxonomy

Castor fiber is the largest Eurasian rodent and sister species to the North American beaver, *C. canadensis*. The two species are the only extant representatives of the Castoridae family, a once highly speciose lineage with an origin *c.* 40 Mya but which is now depauperate and restricted to a single genus, *Castor*. This genus diverged from the scaly-tailed squirrel lineage *Anomalurus c.* 54 Mya (CI: 44-64 Mya) and this long evolutionary history has been suggested as the reason for their peculiar life history traits, morphology and behaviour relative to other members of the Rodentia (Horn *et al.* 2011). Despite the remarkable similarities between *C. fiber* and *C. canadensis,* the divergence date of the mitochondrial genome from a common ancestor has been estimated at > 7 Mya (Horn *et al.* 2011). Speciation between them is likely to have been initiated by ancestral dispersal over the Bering land-bridge followed by isolation of the two populations in Eurasia and North America (Horn *et al.* 2011). Overlap in the New and Old World fossil records for these species does not allow the direction of dispersal to be ascertained (Flynn & Jacobs 2007). It has been assumed by some authors that *Castor* arose in Eurasia as a close relative of *Steneofiber* and subsequently dispersed to North America via Beringia (Xu 1994; Hugueney & Escuille 1996).

The intraspecific taxonomy of this species is poorly resolved in some Eastern areas of its range but it is generally accepted that there are eight subspecific ranks. These correspond to one of eight early-20th century relict populations and are described on the basis of cranio-metric similarities and mitochondrial haplotypes (Freye 1960; Lavrov 1979; Heidecke 1986; Frahnert 2000; Durka *et al.* 2005). *Castor fiber galliae* belongs to the Lower Rhône, *C. f. albicus* – River Elbe, *C. f. fiber* – Telemark, *C. f. belorussicus* - Dnepr and Neman river basins, *C. f. orientoeuropeaus* - Voronezh, *C. f. pohlei* - Konda, *C. f. tuvinicus* – Azas, and *C. f biruli* to the Bulgan river system (Figure 5.1, Table 5.1). Savelijev (2003) argued against the classification of *C. f. pohlei* and *C. f. biruli* as separate subspecies as they may have formed one continuous population along the Irtysh river system. Conversely, the relict population from which *C. f. belorussicus* is described contains two separate river systems (Neman and Dnepr) that may have comprised two discrete populations (Savelijev 2003). The classification adopted here follows that most commonly adopted in the literature and outlined in Durka *et al.* (2005).

A lineage split exists for two reciprocally monophyletic clades, which correspond to the Rhóne, Elbe and Telemark relict populations (Western) and Voronezh, Konda, Azas and Bulgan relict populations (Eastern; Durka *et al.* 2005; Horn *et al.* 2014). These have one contact zone in eastern Europe that has apparently persisted in a similar geographic location since the early Holocene (Horn *et al.* 2014; Biedrzycka *et al.* 2014) and one in NW Scandinavia. A seminal study using partial DNA sequences of tRNAPro and the Hypervariable Region (HV-1) of the Control Region (CR) from ancient beavers uncovered two previously unknown lineages: a Scandinavian subclade nested within the Western group and an extinct and highly divergent lineage from the Danube Basin, which forms a sister group to the Eastern clade Horn *et al.* (2014). The Western and Eastern lineages have a divergence age estimated at 210,000 years ago (CI: 110,000 – 340,000; Horn *et al.* 2011), a timing which corresponds to the penultimate (Eemian) interglacial. Durka *et al.* (2005) proposed these lineages as Evolutionary Significant Units (ESU's) *sensu* Moritz (1994) and recommended that they not be mixed. However, while neutral nuclear markers broadly correspond to the split defined by mitochondrial DNA (Senn *et al.* 2014), more intensive sampling of

modern relict populations has shown that the strict East-West division is not as clear as previously thought, with new haplotypes sequenced that break the conditions of reciprocal monophyly between Eastern and Western populations (Biedrzycka *et al.* 2014; Senn *et al.* 2014). Therefore, while the Eastern and Western lineages most likely diverged following population isolation in refugia during the LGM, introgression following secondary contact of re-emergent populations may have caused mixing of divergent haplotypes via eastern European contact zones.

5.1.3 Conservation and population history of Castor fiber

The Eurasian beaver has a complex demographic history. Archaeological evidence suggests that beavers have been an important resource for humans for many millennia, being hunted mainly for meat, fur and castoreum. At Lake Onega in Northern Europe stone engravings of beavers date to *c.* 3-4 kya (Ravdonikas 1936; Brentjes 1968). In mid-late Holocene Britain skeletal remains have been uncovered at a variety of sites, most notably the Cambridgeshire fenlands, with butchery marks and modifications that suggest their use as artefacts (Coles 2006). The species was once extremely abundant and widely distributed throughout Eurasia but an intensification of hunting practices around the Medieval period and, to a lesser extent, habitat loss and fragmentation, led to severe demographic reductions and the extirpation of the species from huge areas of its historical range (Veron 1992; Nolet & Rosell 1998). It was previously thought that the beaver became extinct in England and Wales by the 12th century but recent indirect radiocarbon evidence from beaver-gnawed wood pushes this date forward to the 14th century (Manning *et al.* 2014). The species persisted in Scotland until around the 1500s although in both regions they probably became functionally extinct much earlier (Manning *et al.* 2014).

By the mid-19th century the species was on the brink of extinction with the total Eurasian population estimated at just 1200 individuals (Lavrov 1983; Nolet & Rosell 1998). Remaining individuals were distributed between scattered and highly isolated relict populations, with the largest populations in the former U.S.S.R (Lavrov 1983; Ducroz *et al.* 2005). Around eight relict populations have been identified (although

recent evidence suggests more may have persisted; Biedrzycka *et al.* 2014; Senn *et al.* 2014) and these passed through severe bottlenecks of between 30 and 300 individuals (see Table 5.1 for references). These relict populations were located along the Lower Rhône delta (France), the middle Elbe (Germany), Telemark (Norway), the Dnepr and Neman river systems running through Belarus, Ukraine, Lithuania and Russia, Voronezh (Russia), Konda River (Russia), Azas River (Russia) and the Bulgan region (China/Mongolia; summarized in Fig 5.1, Table 5.1).

The massive population reduction experienced by beavers had severe genetic consequences for the surviving individuals and their descendants (Durka *et al.* 2005; Ducroz *et al.* 2005; Frosch *et al.* 2014; Horn *et al* 2014; Senn *et al.* 2014). Durka *et al.* (2005), using data from the HV-I domain of the mitochondrial genome, found relict populations to be severely genetically impoverished with low within-population genetic diversity, low divergence among populations and strong phylogeographic structuring. No haplotypes were shared among populations and some populations in Western Europe were monomorphic. A similar signature of low genetic diversity due to a bottleneck effect was found in the major histocompatibility complex (MHC; Ellegren *et al.* 1993). More recent studies using unlinked presumed-neutral nuclear markers such as microsatellites and short nucleotide polymorphisms (SNPs) also confirm the genetically depauperate status of unmixed, relict populations (Frosch *et al.* 2014; Senn *et al.* 2014). Comparison of genetic diversity at mitochondrial loci in extant populations with that of diversity in ancient Holocene populations shows a huge loss of diversity, concordant with an associated decrease in effective population size (Horn *et al.* 2014). A large proportion of the gene pool has undoubtedly been lost in this species; among the ancient dataset in Horn *et al.* (2014), 43 haplotypes were detected from 48 individuals while in the modern dataset only 16 haplotypes were discovered in 152 individuals. No haplotypes were shared among modern and ancient populations. It is therefore likely that the subspecific division is wholly an artefact of human-induced bottleneck effects and artificial population subdivision. The partitioning into subspecies is presumably now invalid although the nomenclature is still utilised in the literature (and in this study) as it provides a convenient means of referring to individuals from relict populations.

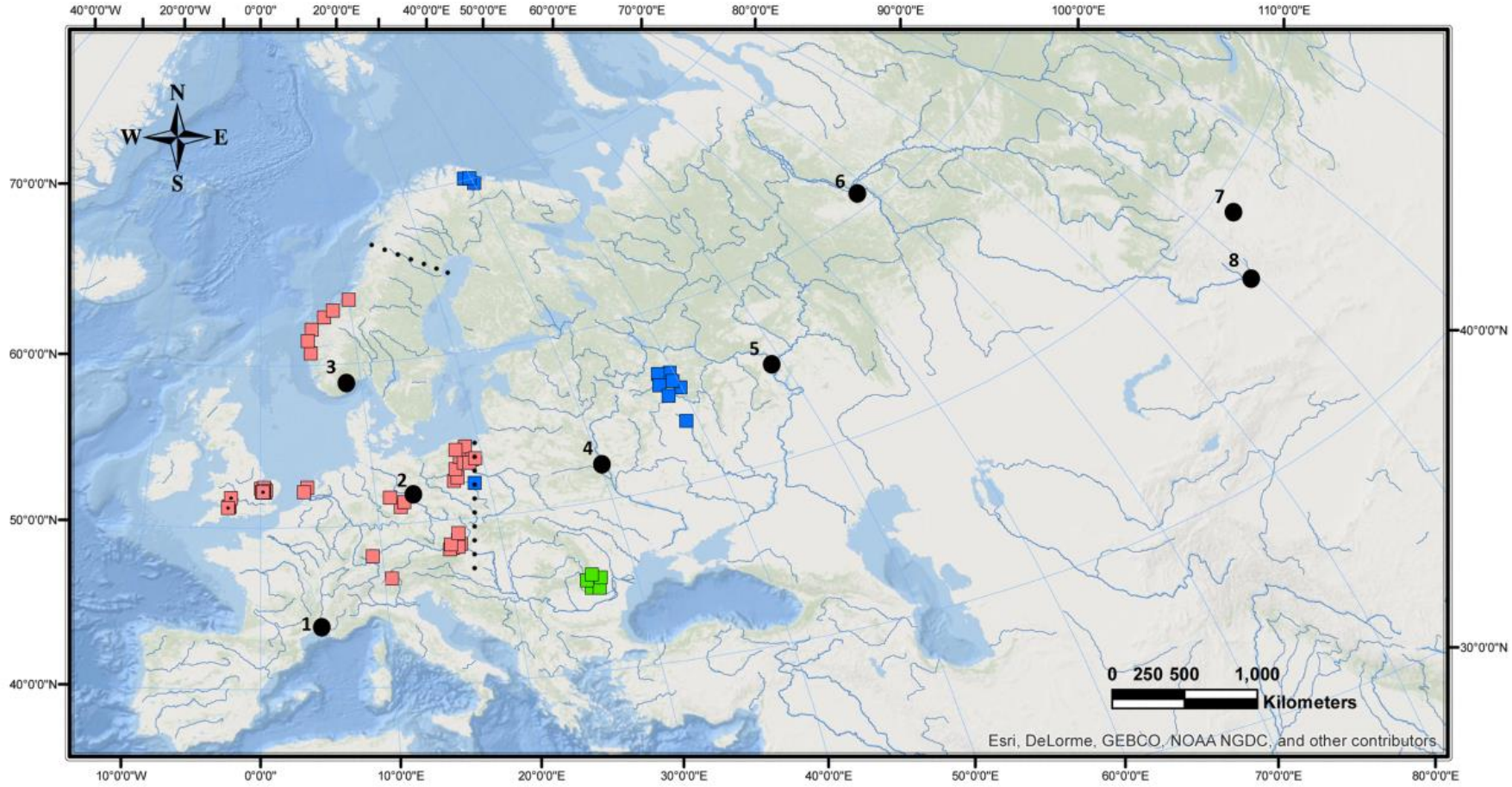


Figure 5.1. Map of sampling localities for ancient *Castor fiber* individuals and locations of relict populations from where modern *C. fiber* haplotypes derive. The Western clade (pink) and Eastern clade (blue) have two Holocene contact zones; one in Eastern Europe and one in North-West Scandinavia. A now extinct clade existed in the Danube Basin (green) until the Late Holocene (Horn *et al.* 2014). Sampling localities for the British data are indicated with dotted square icons, all other data points refer to those generated by Horn *et al.* (2014). Information on relict populations is given in Table 5.1.

Preservation measures and a ban on hunting introduced in the 20th century enabled the persistence and expansion of existing relict populations. Subsequent conservation interventions have focused heavily on translocations. Reintroductions of individuals of multiple origins to restock areas of the species' former range were carried out extensively, with over 203 translocations documented outside the former Soviet Union alone (Halley & Rosell 2002). This proved to be extraordinarily successful with some populations showing exponential increases in numbers and it is now expected that *C. fiber* may soon reoccupy most of its former range. The current Eurasian population size is now estimated at 1 million individuals (Halley *et al.* 2012) and the species is listed as Least Concern (LS) by the International Union

Table 5.1. Locations of relict populations, their subspecies designation and population bottleneck size (Durka *et al.* 2005; Senn *et al.* 2014 and references therein). All haplotypes listed here were used in the analyses.

Population	Subspecies	Country	Locality	Haplotype	Population bottleneck size	Clade
1	*C. f. galliae*	France	Lower Rhône	*ga1*	30	
2	*C. f. albicus*	Germany	Elbe	*al1, al2*	200	**Western**
3	*C. f. fiber*	Norway	Telemark	*fi1, fi3*	60-120	
4	*C. f. belorussicus*	Lithuania/Belarus/ Ukraine/Russia	Dnepr and Neman River Basins	*in1, in2, in3, nh2, nh3, nh4, nh5*	<300	**in1, in2, in3, nh2, nh5 - Eastern; nh3, nh4 - Western**
5	*C. f. orientoeuropaeus*	Russia	Voronezh	-	70	-
6	*C. f. pohlei*	Russia	Konda	*po1, po2*	300	
7	*C. f. tuvinicus*	Azas	Russia	*tu1, tu2, tu3, tu4*	30-40	**Eastern**
8	*C. f. biruli*	Bulgan	Mongolia/China	*bi1, bi2, bi3*	<100-150	

for Conservation of Nature (IUCN). Translocations were, however, fundamentally disparate with no unifying consensus or strategy as to the choice of source population for re-stocking. Three schemes concerning decision-making over this issue in Western Europe, outlined in Halley (2011), are *i*) use of the geographically closest beaver relict population *ii*) mixture of individuals from two or three Western sub-clades and *iii*) release of individuals of multiple origins regardless of ESU assignment. All three approaches have been adopted in Western Europe, often coupled with poor post-release monitoring. Moreover, the literature documenting translocation and

reintroduction projects is scattered and inconsistent, making reconstructing the post-bottleneck population history of the beaver in Europe difficult.

Recent studies using nuclear markers have added some clarity regarding fine-scale population structure. Both Senn *et al.* (2014) and Frosch *et al.* (2014) found a heterogeneous pattern of admixture across Eurasian populations. This pattern is not always congruent with documented evidence of translocations. Using SNP markers and a Bayesian genotype clustering approach, Senn *et al.* (2014) identified five population clusters: Telemark (Norway), Rhône (France), Hesse (Germany, descendants of relict Elbe population), Eastern Europe and Central Eurasia. In cases of relict populations, their distinctiveness may have been accentuated by the effect of genetic drift.

Until now no direct examination of the Late and Postglacial recolonisation of Britain by *C. fiber* has been undertaken using ancient DNA. During the LGM, glacial refugia suitable for beaver habitation were present in the Rhône and Gironde catchments in Southern France and much of the Iberian Peninsula (Coles 2006). In the Eastern part of the range, the Lower Danube and Black Sea also offered suitable temperate forest habitat. The Danube and Black Sea catchments are separated from Britain by a much larger geographic distance and mountain ranges where glacial conditions would have persisted longer than on lower ground (Coles 2006; Halley 2011). Given this geographic separation, it is more probable that during climatic amelioration at the onset of the Lateglacial Interstadial, *C. fiber* recolonised from Western European refugia. At this time, Britain was a peninsula on the western edge of mainland Europe and the Thames a tributary of the Rhine. During the Younger Dryas cold phase, beavers either became locally extinct in Britain, or survived in small southern refugia where local microclimates allowed the persistence of small areas of woodland.

The final Postglacial change to a warmer climate at the onset of the Holocene allowed beavers to disperse rapidly through Britain. The earliest directly-dated Holocene specimen is dated at 11, 068 – 10, 232 cal BP (Gough's Cave, SW England) and the species is also recorded at Star Carr in Yorkshire as early as 10, 900 – 10, 700 cal BP (Coles 2006). During the early Holocene, the Rhine and the Thames were no longer

contiguous but rivers from the east of Britain flowed across the large Doggerland plain. Beavers may therefore have colonised Britain during the Early Holocene from one or a mix of several geographical locations on the mainland, *i*) from Eastern Europe via the North Sea, *ii*) from France along the English Channel, and/or, *iii*) from Germany via the southern North Sea (Senn *et al.* 2014). Post-recolonisation, beaver populations in Britain were presumably part of a larger continental stock until Britain became an island *c.* 8, 200 kya. This indicates that there may have been large numbers of founders that were potentially very genetically diverse; therefore, it is unlikely that the early Holocene British population suffered negatively from founder effects.

In 2009, *C. fiber* was the first extirpated mammal to be officially reintroduced into the UK. Three families (16 individuals) were released into three separate freshwater lochs within Knapdale Forest, Argyll, as part of the Scottish Beaver Trial (Jones & Campbell-Palmer 2014). The success of this trial supports rolling out the reintroduction scheme on a larger scale and, as feasibility studies have been carried out for similar introductions into England and Wales (Gurnell *et al.* 2009), it seems likely that beaver reintroductions across Britain will soon be initiated. The IUCN has strict guidelines for selecting founder populations for reintroductions. The 1998 IUCN guidelines recommend that individuals from the most historically similar population be selected as founders; 'If there is a clear choice of wild populations to supply founder stock for translocations, the source population should ideally be closely related genetically to the original native stock and show similar ecological characteristics (morphology, physiology, behaviour, habitat preference) to the original subpopulation,' (IUCN 1998, p8). However, the 2013 guidelines are not as stringent; 'Founders should show characteristics based on genetic provenance and of morphology, physiology and behaviour that are assessed as appropriate through comparison with the original or any remaining wild populations…' (IUCN 2013, p8). Careful selection of founders for reintroduction to Britain is particularly important as the island is separated from the European landmass. Because of this separation no natural recolonisation, admixture of populations and/or external gene flow can occur as on mainland Europe.

There has been much debate surrounding the most appropriate source population for reintroduction to Britain (Halley 2001; Rosell *et al.* 2012; Senn *et al.* 2014). The selection of individuals with the highest adaptive potential and lowest inbreeding coefficient (outbreeding depression is not considered a major conservation concern; Senn *et al.* 2014) should be of highest importance. However, these individuals may not necessarily be the closest relatives of the original British population. Under a scenario where population size is small and immigration levels are low, reduced mating opportunities and genetic drift may lead to the accumulation of deleterious and recessive alleles and an increase in genome-wide homozygosity (including at fitness-affecting loci; Frankham 1996, Keller & Walker 2002; Szulkin *et al.* 2010). Low levels of genetic diversity in wild populations have been associated with reductions in fitness at both the individual and population level, while positive associations between higher levels of genetic diversity have been associated with several components of fitness (reviewed in Keller & Walker 2002; Reed & Frankham 2003). Loss of genetic diversity can reduce the ability of populations to cope with novel environments via the negative effects of genetic drift and reduced mating opportunities. This could severely impact the success of beaver reintroductions into Britain.

In *C. fiber,* genetic diversity at both neutral nuclear markers and mitochondrial loci is markedly higher in admixed populations relative to unmixed relict populations (Frosh *et al.* 2014; Senn *et al.* 2014). While good rates of population growth have been recorded from translocations using individuals from relict populations, there is some evidence that inbreeding depression affects beavers at both the individual and population level in genetically depauperate populations (Halley 2011, although see Rosell *et al.* 2012). In the Elbe relict population, jaw abnormalities of purportedly genetic origin have been found in *c.* 8.7% of the population (Piechocki 1977). In introduced populations of *C. canadensis* in the Russian Far East (which have also suffered bottleneck effects), dental abnormalities are extremely common (Saveljev & Milishnikov 2002; Saveljev 2011). Conversely, admixed populations tend to show increased fecundity, with higher litter sizes observed in Russian populations of mixed ancestry relative to the unmixed relict populations on the Elbe and in Telemark (Nolet *et al.* 1994; Rosell & Pedersen 1999; Saveljev & Milishnikov 2002). This has led some

authors to argue that the most important factor when selecting source populations for reintroduction in Britain is to maximise the adaptive potential of reintroduced individuals, irrespective of their historical relationship to the original British population (Senn *et al.* 2014).

The ideal solution would be to obtain founders from a source population that is both genetically diverse and shows a close phylogenetic relationship with the Holocene British beaver stock. A comparative cranio-metric analyses on fossil, subfossil and modern beaver specimens showed that the original British stock was morphologically most similar to the Norwegian Telemark population (Kitchener & Lynch 2000). However, bone shows a high level of developmental plasticity. Particularly at the intraspecific level, apparent morphological similarity can be due to a host of environmental factors and does not necessarily indicate relatedness. One means to robustly reconstruct the phylogenetic history of the post-Ice Age British beaver populations is via ancient DNA analyses.

5.1.4 Aims and research Questions

The principle aims of this study are bipartite. Firstly, the competing hypotheses of population continuity or discontinuity over the Lateglacial Interstadial – Younger Dryas – Holocene climatic phases will be tested. This will determine the influence of abrupt climate change on the presence and population structure of a temperate adapted mammal in Britain. Secondly, the origins of early Holocene British beavers, and their relationship to beaver populations of both modern and ancient age on the European mainland, will be established. The objective of this part of the study is to inform future conservation measures. Accordingly, this study will undertake a program of radiocarbon dating and genetic analysis of partial tRNAPro and HV-I loci of the mitochondrial CR. This mitochondrial locus has previously proven suitable for population-level investigations in numerous mammal genera such as *Apodemus*, *Canis*, *Lutra*, and *Ursus* (Durka *et al.* 2005 and references therein).

Specific research aims are to:

1. Reconstruct the phylogenetic history of British beavers of Lateglacial - Holocene age within an existing phylogeny of ancient and modern Eurasian beavers
2. Determine the most recent common ancestor of the Western *C. fiber* lineage
3. Examine sequence data for evidence of a rapid population expansion
4. Use phylogenetic reconstruction to infer the most likely geographic source population on the European mainland for Lateglacial and Early Holocene British beavers
5. Establish whether beavers persisted in Britain over the Younger Dryas cold phase and determine (if possible) their genetic contribution to later Holocene populations

5.2 Methods

5.2.1 Data collection

Bone and tooth samples of *C. fiber* were sourced from a range of museum collections across the UK (full list in Appendix D, Table DT1). Sampling was done *in situ* following the protocol outlined in Chapter 2, Section 2.3.1. Where teeth were available the cementum-rich tooth-root tip was sampled as this area has been shown to yield higher quantities of DNA than dentine (Adler *et al.* 2011, Higgins *et al.* 2013). Sample sites in cranial and post cranial bone material depended largely on what skeletal remains were available and also on whether permission to sample particular parts of the bone was granted from curators. Material was sourced from all UK archaeological sites that potentially had beaver material of Lateglacial to Holocene age and priority was given to specimens that had existing radiocarbon dates available.

5.2.2 Accelerator Mass Spectrometry (AMS) ^{14}C radiocarbon dating

Six samples for which no accurate age could be estimated from either stratigraphic information or dating from associated contemporaneous faunal remains were submitted for dating. Only samples that had yielded CR ancient DNA sequences were included. These were: MM004 (Gough's Cave), MM005 (Gough's Cave), MM007 (Wet Sink), MM033 (Swaffham fen), MM015 (Allermoor) and MM061 (Burwell Fen; see Table 5.3). Chemical pre-treatment and ultrafiltration were used to remove any low-molecular weight contaminants that might have been incorporated into the bone collagen. Sample MM004 had been previously dated by R. Jacobi (Hedges *et al.* 1987) prior to the introduction of ultrafiltration. A number of studies have shown that dates obtained prior to the introduction of current pre-treatment and ultrafiltration practices can have erroneous age estimates due to contamination (*e.g* Higham *et al.* 2006; Jacobi *et al.* 2006). It was therefore decided to re-date this specimen.

Two of the age estimates returned for the Gough's Cave individuals, MM004 and MM005 were very similar. Sampling from these individuals was performed on different skeletal elements and these specimens also had identical CR sequences. This raises the possibility that they may represent one individual. As such the ^{14}C age estimates returned for these were used as the basis for a *chi*-square analysis that tests whether the two dates fall outside of each other's 95 percent confidence ranges (Ward & Wilson 1978).

5.2.3 DNA extraction and PCR amplification

All DNA extractions and polymerase-chain reaction (PCR) preparations were carried out in a dedicated ancient DNA facility within the Department of Earth Sciences, Division of Palaeontology at the Natural History Museum, London. These procedures followed the protocols outlined in Chapter 2, Section 2.3.1 and detailed in Appendix A, Protocols A1 (DNA extraction) and A2 (PCR). All samples were subjected to an initial screening process via single-plex PCR to check for the presence of ancient DNA. Two primers pairs were designed on a template consensus sequence

of 80 previously published *C. fiber* CR sequences using Primer 3 within the GeneiousPro Platform v.8 (http://www.geneious.com, Kearse *et al.* 2012). These produced products of *c.* 100bp (primer sequences and properties are listed in Appendix D, Table DT2). Negative extraction controls and negative PCR controls (with H_2O) were performed with each round of PCR. Amplicons were visualised on 2% agarose gels and products of successful amplifications were purified using 0.15x Ampure Beads to remove primer dimer and residual contaminants of the PCR process. Sequencing was performed commercially by Macrogen on an AB1 Prism 310 Sequencer with primers at 5pM.

Out of a total of 49 individuals, 40 generated PCR products for both primer pairs. A sub-set of 20 for which sequencing quality was highest were chosen for further analyses. A series of overlapping primer pairs designed to amplify a 500bp region of the mitochondrial control region (CR) were either originally designed or selected from previously published sequences in Horn *et al.* 2014 (Appendix D, Table DT2). The high number of primer pairs required to amplify a small target region (500bp) meant that some pairs had sub-optimal properties such as; melting temperatures (*Tm*) outside the 55°C – 65°C range, differences in melting Tm between pairs outside a 5°C range, hairpin, self and pair dimers (Appendix D, Table DT2). All primer pairs were therefore subjected to gradient PCR's which increased annealing temperatures by 1°C over a range of 47°C to 53°C in order to determine the optimal annealing temperature while minimising dimer artefacts. Products ranged in size from 81 to 130bp and were purified and sequenced as above. Chromatograms were visualised and edited in the GeneiousPro Platform v.8 (Kearse *et al.* 2012). All primer sequences and sequencing artefacts were trimmed and base calls manually inspected for errors. Where read quality was high contiguous sequences were created between overlapping pairs. Reads, either as merged contigs or as single reads, were then mapped to a reference consensus sequence created from an alignment of all published *C. fiber* control region sequences. Where low-quality peaks rendered base calling impossible, PCR and sequencing were repeated. For 4 individuals, consistently poor quality reads were produced via Sanger Sequencing for specific regions. For these samples, shot-gun sequencing was used to fill gaps due to missing data. Library builds and NGS

sequencing were carried out as described in Chapter, 2, Section 2.3.1 and Appendix A, Protocol A3. Paired Illumina reads were imported into the CLC Workbench software v.8 (CLC Bio-Qiagen, Aarhus, Denmark) with sequence trimming and merging carried out as described in Chapter 2, Section 2.3.1. Trimmed and merged reads were mapped to a consensus *Caster fiber* partial tRNAPro – HV-1mitochondrial genome reference sequence. Alignment parameters used are described in Chapter 2, Section 2.3.1. Reads that mapped to the consensus with a minimum coverage of 5 reads per base were then exported and re-aligned to a consensus sequence along with PCR generated reads.

5.2.4 Phylogenetic analysis and genetic diversity

To search for new haplotypes, sequences generated in this study were aligned with all published *C. fiber* sequences deposited in GenBank: Accession nos. AY623632-AY623643 (Ducroz *et al.* 2005), DQ088700-DQ088703 (Durka *et al.* 2005), JF264886-JF648888 (Horn *et al.* 2011), HQ880649-HQ880656 (Horn *et al.* 2012), HF674458-HF674455 (Kropf *et al.* 2014), HG915912-HG915951 (Horn *et al.* 2014), KC693753-KC693762 (Beidrzycka *et al.* 2014), KF731635-KF731637 (Frosch *et al.* 2014), KJ670496-KJ670499 (Senn *et al.* 2014). The alignment was performed using Mafft in the GeneiousPro platform v8 before manual inspection and editing (Kearse *et al.* 2012). Sequences were between 487-492bp in length (including gaps). DNAsp (Rozas 2010) was used to generate a haplotype list with gaps and missing data considered and invariable sites not included.

This study used two network based approaches, Median-Joining (MJ) and Neighbour-Net (NN), in addition to Bayesian model-based tree-building. To determine the relative placement of British ancient beavers to ancient and modern populations on the European mainland, the data generated here was combined with previously published CR haplotypes. Due to the complex demographic history of beavers many admixed populations contain individuals from mixed ESUs, which can potentially lead to inaccurate phylogenetic placement in a tree or network and obscure true evolutionary patterns. Therefore, only modern haplotype data from individuals

known to originate from a putative relict population were included. Full details of numbers of individuals, their origins and the analyses they were used in are in Table 5.2. For construction of both networks a CR dataset containing British and European CR sequences was aligned using Mafft in the GeneiousPro Platform v.8 (Kearse *et al.* 2012). This alignment contained all haplotypes generated here in addition to the haplotypes generated by Ducroz *et al.* (2005: accession numbers AY623632-643), Durka *et al.* (2005: DQ088700-703), Horn *et al.* (2012: HQ880649-650), Horn *et al.* 2014 (HG915912-939) and Senn *et al.* (2014: KJ670496-499). The MJ network was implemented in PopART (http://popart.otago.ac.nz). Due to the small genetic distances observed in this dataset ε was set at 0. The NN network was constructed in SplitsTree v.4.14.2 (Huson and Bryant 2006).

To test whether the data is consistent with a neutral model of evolution or shows a signature of a rapid population expansion, three tests of neutrality were performed in DNAsp v.5 (Rozas 2010) and ARLEQUIN v 3.5 (Excoffier & Lischer 2010) for the Western clade only. These were Tajima's D, Fu's F (Fu 1997) and R2 (Ramos-Onsins & Rozas 2002). Tajima's D was calculated as:

$$D = \frac{d}{\sqrt{V(d)}}$$

Equation 5.1. Tajima's D statistic

where d is the difference between *theta* θ and *pi* π (nucleotide diversity). Tajima's D takes into account nucleotide sequence information as both the frequencies of different sequences and the differences between them and tests for violations of the assumptions associated with neutrality (*i.e* infinite sites, random mating, no recombination, equilibrium and constant population size). The *F* statistic of Fu (1997) is based on the probability of having a number of haplotypes greater or equal to the observed number of samples drawn from a constant-size population. Both the Tajima's D and Fu's *F* values are expected to be equal to zero under a null hypothesis of population equilibrium and selective neutrality. The significance levels of both tests were assessed via the generation of 1,000 random samples and significant negative values can be interpreted as evidence of population expansion. The R2 statistic is

based on the difference between the number of singletons and the average number of nucleotide differences and has been shown to have greater power at detecting population expansion for non-recombining regions, even under scenarios where the number of segregating regions is low (Ramos-Onsins & Rozas 2002). A significantly positive R2 value can be interpreted as a signature of population expansion. To complement these analyses a mismatch distribution, performed in ARLEQUIN v 3.5 (Excoffier & Lischer 2010) and based on the sudden demographic expansion model, was examined for evidence of a population experiencing exponential growth. If samples are drawn from populations at demographic equilibrium the mismatch distribution will be multimodel; if populations have passed through a demographic expansion or range expansion with high levels of migration this distribution will fit a unimodal curve (Slatkin & Hudson 1991, Excoffier 2004). The estimated model was tested by obtaining the sum of the squared differences (SSDs) between the observed and estimated mismatch distributions.

Inference of phylogeny was conducted using a Bayesian Markov Chain Monte Carlo (MCMC) approach in MrBayes (Huelsenback *et al.* 2001) and Bayesian Evolutionary Analysis Sampling Trees (BEAST: Drummond *et al.* 2012). The best fit model of nucleotide substitution was determined by JModelTest (Darriba *et al.* 2012). The Akaike, Akaike Corrected and Bayesian Information Criteria (AIC, AICc and BIC) as well as the Decision Theory method all calculated the HKY + G as the most likely model of nucleotide substitution in the dataset. For phylogenetic analyses in MrBayes where the *C. fiber* sequences required an outgroup, two *C. canadensis* control region sequences (Accession numbers AY623645-6; Ducroz *et al.* 2005) were aligned using the same alignment protocol as above. The General-Time Reversible model with gamma distributed rate variation (GTR + G) was calculated as the most likely model of nucleotide substitution.

In datasets where the number of taxa is large and the number of informative characters is low, such as is often typical of population level mitochondrial sequence data, the phylogenetic signal can be weak. Therefore, in the MrBayes analysis, the data was pruned so that only unique haplotypes were included. The final sequence alignment contained ancient Late Pleistocene and Holocene haplotypes generated in

this study combined with additional Holocene-age ancient haplotypes generated by Horn *et al.* (2012: accession numbers HQ880649-656), Horn *et al.* (2014: HG915912-950), and modern haplotypes that originated in relict populations with no history of admixture from Durka *et al. (*2005: DQ088700-703), Ducroz *et al.* (2005: AY623632-643) and Senn *et al. (*2014: KJ670496-670499). The final dataset for the MrBayes run contained 79 *C. fiber* individuals and two *C. canadensis* haplotypes as outgroup taxa. Two independent chains were run for 10^6 iterations each, sampling trees and model parameters from the posterior every 1000 iterations. The first 10% of each run was discarded as burn-in. TRACER v.1.5 (Rambaut *et al.* 2014) and Are We There Yet (AWTY: Wilgenbusch *et al.* 2004) were used to check for chain convergence and sufficient sampling before a 50% consensus tree was created and edited in FigTree v.1.3 (Rambaut & Drummond 2010).

To take advantage of the available time-stamped data, and to see whether this improved tree topology, data that had accurate age determinations was subjected to BEAST analyses. As intraspecific sequence variation is driven by stochastic rather than evolutionary factors, a strict molecular clock (which assumes a constant rate of evolution across lineages) is generally considered the most appropriate clock model at the population level (Drummond *et al.* 2007). A full exploration of clock rates, demography and divergence dates for *C. fiber* is detailed in Horn *et al.* (2014) and is therefore not undertaken here. The Horn *et al.* (2014) investigation was not able to ascertain accurate estimates of the tMRCA for the main *C. fiber* lineages and thus these could not be incorporated in this analysis. For CR sequences generated by this study, ages were determined by AMS radiocarbon dating. In the ancient data from Horn *et al.* (2012: Acc. Nos HQ880649-565), Horn *et al.* (2014: HG915912-950), sample ages were inferred by both radiocarbon ages and by zoo-archaeological information. All other samples included were of modern age (0); Ducroz *et al.* (2005: AY623632-643), Durka *et al.* (DQ088700-703) and Senn e*t al.* (KJ670496-499). All ages were input as tip dates in median cal BP format. Identical sequences were included in the BEAST run if they possessed accurate age information. BEAST assumes that a population-level dataset has been sampled at random and, if a dataset includes only unique haplotypes, the algorithm will erroneously assume a large population size. For all

BEAST analyses, the tMRCA for all lineages was set to 210, 000 as determined by Horn *et al.* (2011) and the HKY + G model of nucleotide substitution was used. To calibrate the molecular clock a variety of substitution rates were used: 1.0 x 10-7, 3.0 x 10-7 and 5.0 x 10-7 (as in Horn *et al.* 2014) with a uniform prior, the upper and lower bounds of which were set at ± 0.25 of the initial value. These clock rate values were obtained as an estimate for the *C. fiber* mitogenome as a whole (Horn *et al.* 2011) and, as the CR is known to have a particularly high mutation rate, these may be an underestimate. Two demographic models were compared, *i*) constant population size and *ii*) coalescent population size. For each separate BEAST run, a 'null' simulation was carried out where sampling is performed only from the prior and not the data.

Two independent MCMC chains were run for 1 x 10^8 iterations each, sampling trees and model parameters from the posterior every 10000 iterations. The first 10% was discarded as burn-in and TRACER v.1.5 used to inspect the data for convergence and sufficient sampling. To select the most appropriate model, the integrated BEAST option of path sampling and stepping stone sampling was used in order to estimate marginal likelihoods (MLEs) for each model. These were run on the default parameters of four independent MCMC chains comprised of 1000 steps of 100,000 generations following a burn-in period of 10 million generations. Bayes factors were

Table 5.2. All *Castor fiber* tRNAPro/Control Region haplotypes used in this study. Also indicated is which analyses each sample was used in. B – MrBayes, Be –BEAST, N - network. In cases of identical haplotypes some of these were collapsed for the BEAST run. In these instances, tip labels in Fig 5.5 indicate which sequences are represented.

Accession	Reference	Modern/ Ancient	Haplotype/ID	Clade	Subspecies	Locality	Country	Median Date (cal BP)	Analyses
DQ088700	Durka et al. 2005	M	al1	Western	*albicus*	Elbe	Germany	0	B/Be/N
DQ088701	Durka et al. 2005	M	al2	Western	*albicus*	Elbe	Germany	0	B/Be/N
HG915944	Horn et al. 2014	A	AUT1	Western	*N/A*	Bernhardsthal, "W/41 Verf.A"	Austria	1575	B/Be/N
HG915945	Horn et al. 2014	A	AUT2	Western	*N/A*	Traismauer	Austria	1575	B/Be/N
HG915946	Horn et al. 2014	A	AUT3	Western	*N/A*	Traismauer	Austria	1575	B/Be/N
HG915947	Horn et al. 2014	A	AUT4	Western	*N/A*	Stillfried	Austria	3450	B/Be/N
HG915948	Horn et al. 2014	A	AUT5	Western	*N/A*	Melk-Winden	Austria	6450	B/Be/N
HG915949	Horn et al. 2014	A	AUT6	Western	*N/A*	Kamegg	Austria	6450	B/N
AY623632	Ducroz et al. 2005	M	bi1	Eastern	*biruli*	Bulgan-gol, Chovd Ajmak	Mongolia	0	B/Be/N
AY623633	Ducroz et al. 2005	M	bi2	Eastern	*biruli*	Bulgan-gol, Chovd Ajmak	Mongolia	0	B/Be/N
AY623634	Ducroz et al. 2005	M	bi3	Eastern	*biruli*	Bulgan-gol, Chovd Ajmak	Mongolia	0	B/Be/N
HG915951	Horn et al. 2014	A	CZE1	Western	N/A	Krtiny, Byciskala Cave	Czech Republic	10000	B/N
DQ088702	Durka et al. 2005	M	fi1	Western	*fiber*	Telemark	Norway	0	B/Be/N
DQ088703	Durka et al. 2005	M	ga1	Western	*galliae*	Seine/Rhone	France	0	B/Be/N
-	This study	A	GB1/MM002	Western	*N/A*	Burwell Fen	GB	4892.5	B/Be/N
-	This study	A	GB1/MM049	Western	*N/A*	Swaffham Fen	GB	Unknown	B/N
-	This study	A	GB1/MM051	Western	*N/A*	Burwell Fen	GB	4784	B/Be/N
-	This study	A	GB10/MM028	Western	*N/A*	Burwell Fen	GB	4775	B/Be/N
-	This study	A	GB11/MM030	Western	*N/A*	Reach Fen	GB	Unknown	B/N
-	This study	A	GB12/MM031	Western	*N/A*	Reach Fen	GB	4883.5	B/Be/N
-	This study	A	GB13/MM033	Western	*N/A*	Swaffham Fen	GB	2338.5	B/N
-	This study	A	GB14/MM061	Western	*N/A*	Burwell Fen	GB	4332.5	B/Be/N
-	This study	A	GB2/MM003	Western	*N/A*	Fen	GB	Unknown	B/N
-	This study	A	GB3/MM004	Western	*N/A*	Gough's Cave	GB	11697	B/Be/N
-	This study	A	GB3/MM005	Western	*N/A*	Gough's cave	GB	12111	B/Be/N
-	This study	A	GB4/MM006	Western	*N/A*	Sunhole Cave	GB	Unknown	B/N
-	This study	A	GB5/MM007	Western	*N/A*	Wet Sink	GB	Unknown	B/N
-	This study	A	GB6/MM008	Western	*N/A*	Somersham Fen	GB	4699	B/Be/N
-	This study	A	GB7/MM015	Western	*N/A*	Allermoor, Somerset Levels	GB	5454.5	B/Be/N

Table 5.2 cont. All *Castor fiber* tRNAPro/Control Region haplotypes used in this study. Also indicated is which analyses each sample was used in. B – MrBayes, Be –BEAST, N - network. In cases of identical haplotypes some of these were collapsed for the BEAST run. In these instances, tip labels in figure 5.5 indicate which sequences are represented.

Accession	Reference	Modern/ Ancient	Haplotype/ID	Clade	Subspecies	Locality	Country	Median Date (cal BP)	Analyses
-	This study	A	GB8/MM025	Western	*N/A*	Fen	GB	Unknown	B/N
-	This study	A	GB9/MM026	Western	*N/A*	Burwell Fen	GB	5207.5	B/Be/N
HG915926	Horn et al. 2014	A	GER1	Western	*N/A*	Zigeunerfels, Sigmaringen	Germany	10950	B/Be/N
HG915927	Horn et al. 2014	A	GER2	Western	*N/A*	Leipzig-Connewitz	Germany	2050	B/Be/N
HG915928	Horn et al. 2014	A	GER3	Western	*N/A*	Leipzig-Connewitz	Germany	2050	B/Be/N
AY623641	Ducroz et al. 2005	M	in1	Eastern	*belorussicus/orientoeuropaeus*	Orël Refion	Russia	0	B/Be/N
AY623642	Ducroz et al. 2005	M	in2	Eastern	*belorussicus/orientoeuropaeus*	Various	Lithuania	0	B/Be/N
AY623643	Ducroz et al. 2005	M	in3	Eastern	*belorussicus/orientoeuropaeus*	Kirsna, Kaunas	Lithuania	0	B/Be/N
HG915950	Horn et al. 2014	A	ITA1	Western	*N/A*	Ganglegg, Sluderno	Italy	2050	B/Be/N
KJ670496	Senn et al. 2014	M	nh2	Eastern	*belorussicus*	Dnepr and Neman River Basins	Belarus	0	B/Be/N
KJ670497	Senn et al. 2014	M	nh3	Western	*belorussicus*	Dnepr and Neman River Basins	Belarus	0	B/Be/N
KJ670498	Senn et al. 2014	M	nh4	Western	*belorussicus*	Dnepr and Neman River Basins	Belarus	0	B/Be/N
KJ670499	Senn et al. 2014	M	nh5	Eastern	*belorussicus*	Dnepr and Neman River Basins	Belarus	0	B/Be/N
HG915912	Horn et al. 2014	A	NOR1	South_Scandinavia	*N/A*	Ruskeneset, Bergen, Hordaland	Norway	3387	B/Be/N
HG915913	Horn et al. 2014	A	NOR2	South_Scandinavia	*N/A*	Sauehelleren, Aukra, Møre og Romsdal	Norway	1510	B/Be/N
HG915914	Horn et al. 2014	A	NOR3	South_Scandinavia	*N/A*	Dalen, Rissa, Sør-Trøndelag	Norway	3750	B/Be/N
HG915915	Horn et al. 2014	A	NOR4	South_Scandinavia	*N/A*	Skipshelleren, Vaksdal, Hordaland	Norway	3481	B/Be/N
HG915916	Horn et al. 2014	A	NOR5	Eastern	*N/A*	Mortensnes, Nesseby, Finnmark	Norway	5950	B/Be/N
HG915917	Horn et al. 2014	A	NOR6	South_Scandinavia	*N/A*	Skylehammeren, Ålesund	Norway	2950	B/Be/N
HG915918	Horn et al. 2014	A	NOR7	South_Scandinavia	*N/A*	Ullshelleren, Odda, Hordaland	Norway	1950	B/Be/N
HG915919	Horn et al. 2014	A	NOR8	Eastern	*N/A*	Mestersanden, Sør-Varanger, Finnmark	Norway	1800	B/Be/N
HG915920	Horn et al. 2014	A	NOR9	Eastern	*N/A*	Kjelmøy, Sør-Varanger, Finnmark	Norway	1750	B/Be/N
HQ880650	Horn et al. 2012	A	Nsea1	Western	*N/A*	Nordsee, Eurogeul/Maasvlakte	North Sea	15000	B/Be/N
HQ880649	Horn et al. 2012	A	Nsea2	Western	*N/A*	Northsea, West of Bruine Bank	North Sea	10400	B/N
AY623635	Ducroz et al. 2005	M	po1	Eastern	*pohlei*	Konda River, Kondiskyj Zakaznik	Russia	0	B/Be/N
AY623636	Ducroz et al. 2005	M	po2	Eastern	*pohlei*	Konda River, Kondiskyj Zakaznik	Russia	0	B/Be/N
HQ880652	Horn et al. 2012	A	POL1	Western	*N/A*	Gluchowo	Poland	2450	B/Be/N
HG915936	Horn et al. 2014	A	POL10	Western	*N/A*	Mniszek	Poland	350	B/Be/N
HG915937	Horn et al. 2014	A	POL11	Western	*N/A*	Starorypin	Poland	650	B/Be/N

Table 5.2 cont. All *Castor fiber* tRNAPro/Control Region haplotypes used in this study. All haplotypes used in this study. Also indicated is which analyses each sample was used in. B – MrBayes, Be –BEAST, N - network. In cases of identical haplotypes some of these were collapsed for the BEAST run. In these instances, tip labels in figure 5.5 indicate which sequences are represented.

Accession	Reference	Modern/ Ancient	Haplotype/ID	Clade	Subspecies	Locality	Country	Median Date (cal BP)	Analyses
HG915938	Horn et al. 2014	A	POL12	Western	*N/A*	Zolte	Poland	650	B/Be/N
HG915929	Horn et al. 2014	A	POL2	Western	*N/A*	Gluchowo	Poland	2450	B/Be/N
HG915930	Horn et al. 2014	A	POL3	Eastern	*N/A*	Gluchowo	Poland	650	B/Be/N
HG915931	Horn et al. 2014	A	POL4	Eastern	*N/A*	Gluchowo	Poland	650	B/N
HQ880653	Horn et al. 2012	A	POL5	Western	*N/A*	Ostrow Lednicki	Poland	650	B/Be/N
HG915932	Horn et al. 2014	A	POL6	Western	*N/A*	Bydgoszcz	Poland	650	B/Be/N
HG915933	Horn et al. 2014	A	POL7	Western	*N/A*	Bydgoszcz	Poland	650	B/Be/N
HG915934	Horn et al. 2014	A	POL8	Western	*N/A*	Gdansk	Poland	650	B/Be/N
HG915935	Horn et al. 2014	A	POL9	Western	*N/A*	Mniszek	Poland	350	B/Be/N
HG915921	Horn et al. 2014	A	ROU1	Danube	*N/A*	Pietrele	Romania	6320	B/Be/N
HG915922	Horn et al. 2014	A	ROU2	Danube	*N/A*	Pietrele	Romania	6320	B/Be/N
HG915923	Horn et al. 2014	A	ROU3	Danube	*N/A*	Pietrele	Romania	6320	B/Be/N
HG915924	Horn et al. 2014	A	ROU4	Danube	*N/A*	Pietrele	Romania	6320	B/Be/N
HG915925	Horn et al. 2014	A	ROU5	Danube	*N/A*	Pietrele	Romania	6320	B/Be/N
HG915939	Horn et al. 2014	A	RUS1	Eastern	*N/A*	Rostislavl	Russia	10000	B/N
HG915940	Horn et al. 2014	A	RUS2	Eastern	*N/A*	Rostislavl	Russia	10000	B/N
HQ880655	Horn et al. 2012	A	RUS3	Eastern	*N/A*	Ivanovskoe, cloij 1	Russia	5000	B/Be/N
HG915941	Horn et al. 2014	A	RUS4	Eastern	*N/A*	Ivanovskoe, cloij 2	Russia	6260	B/Be/N
HQ880654	Horn et al. 2012	A	RUS5	Eastern	*N/A*	Ivanovskoe, cloij 3	Russia	8500	B/Be/N
HG915942	Horn et al. 2014	A	RUS6	Eastern	*N/A*	Ivanovskoe, cloij 4	Russia	9600	B/Be/N
HQ880656	Horn et al. 2012	A	RUS7	Eastern	*N/A*	Ivanovskoe, cloij 4	Russia	9600	B/Be/N
HG915943	Horn et al. 2014	A	RUS8	Eastern	*N/A*	4759/10740, Stanovoe-4	Russia	10000	B/N
AY623637	Ducroz et al. 2005	M	tu1	Eastern	*tuvinicus*	Azas River, Tyva Republic	Russia	0	B/Be/N
AY623638	Ducroz et al. 2005	M	tu2	Eastern	*tuvinicus*	Azas River, Tyva Republic	Russia	0	B/Be/N
AY623639	Ducroz et al. 2005	M	tu3	Eastern	*tuvinicus*	Azas River, Tyva Republic	Russia	0	B/Be/N
AY623640	Ducroz et al. 2005	M	tu4	Eastern	*tuvinicus*	Azas River, Tyva Republic	Russia	0	B/Be/N
HQ880651	Horn et al. 2012	A	Nsea3	Western	None	Northsea, West of Bruine Bank	North Sea	15000	B/Be/N

then calculated from these data and model preference determined using the criteria of Kass & Raftery (1995). A maximum clade credibility (MCC) tree, which estimates the maximum *a posteriori* tree topology by outputting a tree with the maximum product of the posterior clade probabilities, was generated in TreeAnnotator (Rambaut 2008).

5.3 Results

5.3.1 Accelerator Mass Spectrometry (AMS) radiocarbon dating

Of the six samples submitted to the ORAU, five were dated successfully and one sample (MM007) failed due to low collagen yield. Full results are in Table 5.3, Fig 5.2. Surprisingly two samples from Gough's Cave, MM004 (11,989-11,405 cal BP) and MM005 (12,386-11,836 cal BP), date to the Younger Dryas cold phase during the Terminal Pleistocene, representing the oldest post-LGM directly dated *C. fiber* samples in Europe. The result obtained from the re-date of sample MM004 returned a much earlier date than that obtained by R. Jacobi in 1987 (11,068-10,957 cal BP; Hedges *et al.* 1987). This is significant because it places this individual in the closing phase of the Younger Dryas with a probability limit extending into the very Early Holocene rather that firmly in the Early Holocene. There is a small amount of overlap in the ^{14}C probability distribution for MM004 and MM005. The chi-square test showed that these are significantly different from each other on the basis of the probability distributions of ^{14}C evidence ($d.f = 1$, $T = 8.0$, $P < 0.05$). While this is not conclusive, it does strongly suggest that, on the basis of ^{14}C evidence, these are discrete individuals. The full date-range for the Cambridgeshire fenland beaver estimated directly from radiocarbon dated bone is 5450 - 2380 cal BP (Mayhew 1978, Coles 2006). All the fenland beaver material in this study dated to within this range, apart from MM033, which has an age of 2466-2211 cal BP and thus extends the known range of fenland beaver (Table 5.3). Sample MM015 from Allermoor, Somerset, returned a mid-Holocene date range of 5583-5326 cal BP.

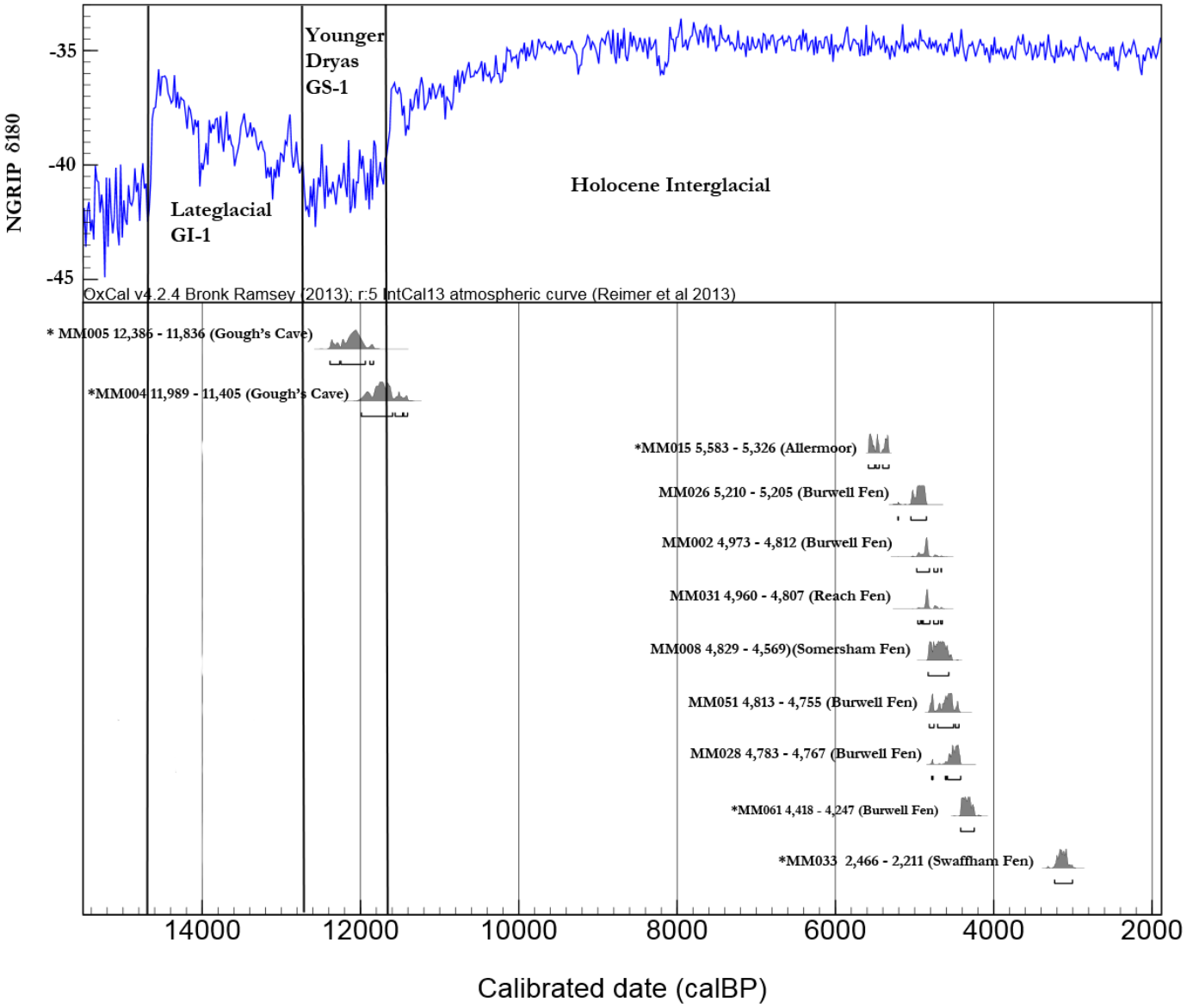


Figure 5.2. AMS radiocarbon dates for all *Castor fiber* samples used in the phylogenetic analyses in this study. For samples dated specifically for this investigation (denoted by *) the procedure was carried out at the ORAU. Two samples have ages which place them in the Younger Dryas cold phase while the remainder of samples date to the mid - late Holocene. Full details in Table 5.3.

Table 5.3. AMS radiocarbon age determinations for *Castor fiber* samples used in the phylogenetic analysis in this study.

Lab ID	Collection ID	Museum	Site 1	County	Sample Site (aDNA)	Year dated	14C date	Error	14C cal BP, 95.4% Prob	$\delta^{13}C$	$\delta^{15}N$	%Yld	%C	C/N	OxA ID	Reference
MM002	E1794	UMZC	Burwell fen	Cambridgeshire	Maxilla	2002	4285	45	4973-4812	-	-	-	-	-	11, 093	Coles (2006)
MM003	E1802	UMZC	Unknown	Cambridgeshire	Mandible	-	-	-		-	-	-	-	-	-	-
MM004	M17.2.9	UBSSM	Gough's Old Cave	Mendips	Mandible	2016	10110	45	11989-11405	-21.4	2.8	7.1	45.7	3.2	33,525	This study
-	M17.2.9	UBSSM	Gough's Old Cave	Mendips	Mandible	1987	9320	120	11068-10957	-	-	-	-	-	1,119	Hedges *et al.* (1987)
MM005	M17.2.7	UBSSM	Gough's Old Cave	Mendips	Maxilla	2016	10300	50	12386-11836	-21.7	3.2	7.4	46	3.2	33,526	This study
MM006	M5.2.3	UBSSM	Sunhole cave	Mendips	Humerus	-	-	-		-	-	-	-	-	-	-
MM007	2.792	GCMA	Wet Sink	Gloucester	Maxilla	2016	Failed Low Yield	-		-	-	-	-	-	-	This Study
MM008	E1798.2	UMZC	Somersham Fen	Cambridgeshire	Mandible	2002	4150	40	4829-4569	-	-	-	-	-	11,094	Coles (2006)
MM015	46/1995/421	SCCM	Aller Moor	Somerset	Maxilla	2016	4726	29	5583-5326	-22.8	3.4	3.4	43.2	3.3	33,524	This study
MM025	D28651	SM	Unknown	Cambridgeshire	Mandible	-	-	-		-	-	-	-	-	-	-
MM026	D28705	SM	Burwell fen	Cambridgeshire	Mandible	2002	4375	40	5210-5205	-	-	-	-	-	11,090	Coles (2006)
MM028	D28710	SM	Burwell fen	Cambridgeshire	Mandible	2002	4032	37	4783-4767	-	-	-	-	-	11,218	Coles (2006)
MM030	D28712	SM	Reach Fen	Cambridgeshire	Mandible	-	-	-		-	-	-	-	-	-	-
MM031	D28713	SM	Burwell fen	Cambridgeshire	Maxilla	2002	4265	40	4960-4807	-	-	-	-	-	11,092	Coles (2006)
MM033	D28719	SM	Swaffham Fen	Cambridgeshire	Mandible	2016	3870	33	2466-2211	-22.7	4.3	11.9	44.3	3.2	33685	This Study
MM049	D28708	SM	Swaffham Fen	Cambridgeshire	Mandible	-	-	-		-	-	-	-	-	-	-
MM051	E1796	SM	Burwell Fen	Cambridgeshire	Maxilla	2002	4085	40	4813-4755	-	-	-	-	-	11,088	Coles (2006)
MM061	D28716	SM	Burwell Fen	Cambridgeshire	Mandible	2016	3899	29	4418-4247	-22.5	5.3	5.2	45	3.2	33,529	This study

5.3.2 Phylogenetic analysis and genetic diversity

In total 49 samples of *C. fiber* bone and tooth material from UK Lateglacial-Early Holocene sites were collected over the duration of this study, of which 40 successfully generated amplicons during a PCR screening process. This dataset was reduced to a subset of 20, of which 17 generated full-length tRNAPro/HV-1 CR mitochondrial DNA sequences (Table 5.4) that were 492 bp in length after trimming. Only two of these, MM004 and MM005, were of Late Pleistocene age, dating to the Younger Dryas. No Lateglacial specimens yielded full-length sequences. All sequences originate from individuals in the SW of England covering five sites (Fig 5.1, Table 5.4): Gough's Cave, Cheddar (n = 2); Wet Sink, Gloucestershire (n = 1), Sunhole Cave, Mendips (n = 1); Allermoor, Somerset Levels (n = 1) and the Cambridgeshire fenlands (n = 12). Despite intensive efforts to obtain full-length sequence data from samples originating in Scotland and the North of England, aDNA yields were consistently too low and of poor quality.

DNAsp (Rozas 2010) was used to generate a list of unique and identical haplotypes with sites with gaps, ambiguities and missing data considered. Of the 17 sequences generated here, 13 were unique. Specimens MM004 and MM005 from Gough's Cave were identical to each other as were three haplotypes from the Cambridgeshire Fens, MM002 (Burwell Fen), MM049 (Somersham Fen), MM051 (Burwell Fen). MM006 was identical to the ancient POL2 haplotype discovered by Horn *et al.* (2014). No haplotypes generated here were shared with sequences from modern populations. Haplotypes were renamed according to their location (Table 5.4).

Table 5.4. *Castor fiber* tRNAPro-Hypervariable Region 1, Control Region (HV-1, CR) mitochondrial haplotypes from 17 British specimens.

Haplotype	Sample	Locality
GB1a, b, c	MM002, MM049, MM051	Burwell Fen; Swaffham Fen; Burwell Fen (Cambridgeshire Fens)
GB2	MM003	Cambridgeshire Fens
GB3a, b	MM004, MM005	Gough's Cave
POL2/GB4	MM006	Sunhole Cave
GB5	MM007	Wet Sink
GB6	MM008	Somersham Fen
GB7	MM015	Allermoor, Somerset Levels
GB8	MM025	Cambridgeshire Fens
GB9	MM026	Burwell Fen
GB10	MM028	Burwell Fen
GB11	MM030	Reach Fen
GB12	MM031	Burwell Fen
GB13	MM033	Swaffham Fen
GB14	MM061	Burwell Fen

A database of 85 haplotypes was used to create both MJ and NN networks (Figures 5.3 and 5.4). Calculation of measures of genetic diversity were performed in DNAsp (Rozas 2010) using this dataset rather than the pruned alignments used for tree-building (see Methods). Excluding the 108 sites with gaps and missing data, there were 384 sites considered in the analysis. Of these 329 were invariable (monomorphic), leaving 55 polymorphic sites of which 33 were parsimony informative. Total number of mutations was 57. There were 51 unique haplotypes with a haplotype diversity index value (*Hd*) of 0.971 and a nucleotide diversity index value (π) of 0.021. When sites with gaps and missing data were considered there were 79 unique haplotypes with 89 variable sites and *Hd*=0.998.

Pairwise genetic distances were computed for the following partitions in the ancient dataset with gaps coded as a fifth character state: *i*) Western (including South Scandinavia) and Eastern clades (excluding Danube) and *ii*) South Scandinavia and the remaining Western haplotypes. No specific comparison was made between measures of genetic diversity in the British data versus the other clades as there is no evidence that the British data forms a distinct, divergent lineage (see below). The addition of the British haplotypes increased *Fst* values between the Western and Eastern clades from 0.61 in Horn *et al.* (2014) to 0.7 in this study. Genetic distance

between the South Scandinavian sub-clade and the rest of the Western lineage was very small at *Fst* =0.054 but may have been affected by the small sample size in the South Scandinavian lineage (*n*=6).

The patterns of lineage division and placement of the British *C. fiber* CR mitochondrial haplotypes were highly congruent in both the Median-Joining MJ and Neighbour-Net NN networks (Figs 5.3 and 5.4). Both networks recovered the splits between the three major clades: Western, Eastern and the Danube Basin with the South Scandinavian Western lineage shown as a moderately divergent subclade within the Western ESU. The British data all grouped unambiguously within the Western clade in both analyses and show no detectable divergence from other Western haplotypes. Like DNAsp (Rozas 2010), the MJ network analysis excluded sites with gaps and missing data, which resulted in a reduction in the number of unique haplotypes and had the effect of collapsing a number of haplotype sequences in the network into one unique haplotype circle, shown with size proportional to number of haplotypes (Fig 5.3). The topology of the Western clade in this network is star-shaped with a central circle (AUT5 in Fig 5.3) representing British haplotypes GB2, GB3a-b, GB4, GB8 and GB9, GB11, GB13, GB14 along with al1, AUT5 and POL2. From this central circle one mutational step leads to British haplotypes GB1a-c and GB12. Haplotype GB7 (MM015, Allermoor, Somerset) is the most divergent with three mutational steps from the central circle (Fig 5.3). The minimum and maximum number of mutational steps from the British data and the closest South Scandinavian haplotype was two (haplotypes within 'AUT5' circle to NOR7) and four (haplotypes within GB1a circle to NOR7) indicating that while the Southern Scandinavian sub-group is sufficiently divergent to form its own sub-clade, only a small number of mutational steps separate it from the British data.

The NN network does not exclude alignment sites with missing data or gaps and thus gives a better representation of true haplotype diversity within the network (Fig 5.4). In this analysis branch lengths are proportional to genetic divergence and parallelograms show alternative pathways between haplotypes. Here, the British data typically exhibit short branch lengths indicating a lack of substructure between them

and the rest of the Western clade, excluding South Scandinavia, which appears more divergent in this network (although not excessively so). Haplotype GB7 (MM015) is again the most divergent in the British dataset with a long branch length. The large number of parallel boxes between haplotypes indicate that the phylogenetic relationships between the British data and the rest of the Western clade, and within the Western clade as a whole, are highly unresolved with the exception of the South Scandinavian lineage (Fig 5.4).

Both the MrBayes and the BEAST analyses returned tree topologies that were broadly congruent both with each other and the topology returned in the network analyses. All estimated samples sizes (ESS) for parameters in both analyses were well over the recommended value of 200 and the Potential Scale Reduction Factor (PSRF) for all parameters was > 1. Chain convergence, analysis of splits and all other MCMC outputs were examined in TRACER v.1.5 (Rambout *et al.* 2014) and AWTY (Wilgenbusch *et al.* 2004). The North American beaver *C. canadensis* was used as an outgroup in the MrBayes analysis and the high level of sequence divergence between this species and *C. fiber* at the CR locus (Horn *et al.* 2011) resulted in long branch lengths between the two species (Fig 5.5). In the MrBayes analysis, the posterior probability values in the 50% consensus tree for branches leading to the Danube, Eastern and South Scandinavian clades were all returned as 1 but support for the branch leading to the Western clade was low at 0.54. Within the Western clade (excluding South Scandinavia), most of the haplotypes, including the British data, are collapsed into a large polytomy indicating complete absence of phylogenetic structure.

Of all the BEAST analyses, Bayes factors calculated from 1000 bootstraps on the marginal likelihood estimates (MLEs) of path and stepping stone sampling supported the scenario of a strict molecular clock, clock rate of 5×10^{-7} and a coalescent exponential growth model of demography (Fig 5.5). However, it should be noted that there were no differences in either tree topology or nodal and branch support between runs. Estimated sample sizes (ESSs) in the BEAST runs which sampled only from the prior were all > 200 apart from in the scenario favoured by the Bayes factors

(Appendix D, Table DT3). This indicates that in all simulations (bar one), the priors were contributing to the results and thus those results were not driven by the data alone. The maximum clade credibility (MCC) tree topology, in contrast to the MrBayes phylogeny, showed high support (0.99) for the branch leading to the Western clade and high support (1) for branches leading to all other major clades (Fig 5.6), although nodal support for the split between the inner Eastern lineage and the Danube lineage was low (0.58) The British data all grouped together with haplotype ga1. While British individuals, as in the MrBayes and network analyses, appear more similar to each other than to other individuals from different regions there was no support for the British data forming a distinct lineage. Overall, there was a general pattern of unresolved phylogenetic division within the Western clade as a whole. The haplotypes used in the BEAST simulations all had accurate time-stamped data accompanying them and it is likely that the tree calibration is well resolved, particularly towards the tips of this tree. The unresolved placement of the British data within the Western clade meant that obtaining an estimated divergence date for the British population was not possible. A range of 30,651 – 73,056 cal BP was returned for the tMRCA of the western clade as a whole (mean = 50,733 cal BP; ESS = 5503.53) and 22, 686 – 56,402 cal BP for the tMRCA of the Eastern lineage (mean = 38,545; ESS 5882.36).

Tajima's D statistic and Fu's F returned negative and significant values for the Western clade (D = -1.66, p = <0.05; F = -17.41, p = < 0.001) and R2 returned a significant and positive value (R2 = 0.043, p = < 0.005). The mismatch distribution found that this clade also fitted a unimodal pattern of demographic expansion (*i.e* no significant deviation from this model was observed; SSD p = 0.97) while deviating from a pattern of constant population size (Appendix D, Figure DF1). The confidence intervals for the node age estimates on the BEAST phylogeny are too large to allow a narrow age estimate for the timing of population expansion in the Western clade. However, a pre-LGM tMRCA for the Western clade together with highly congruent tests which suggest sudden demographic expansion are strongly indicative of a scenario of a glacial bottleneck followed by a rapid, post-LGM population expansion.

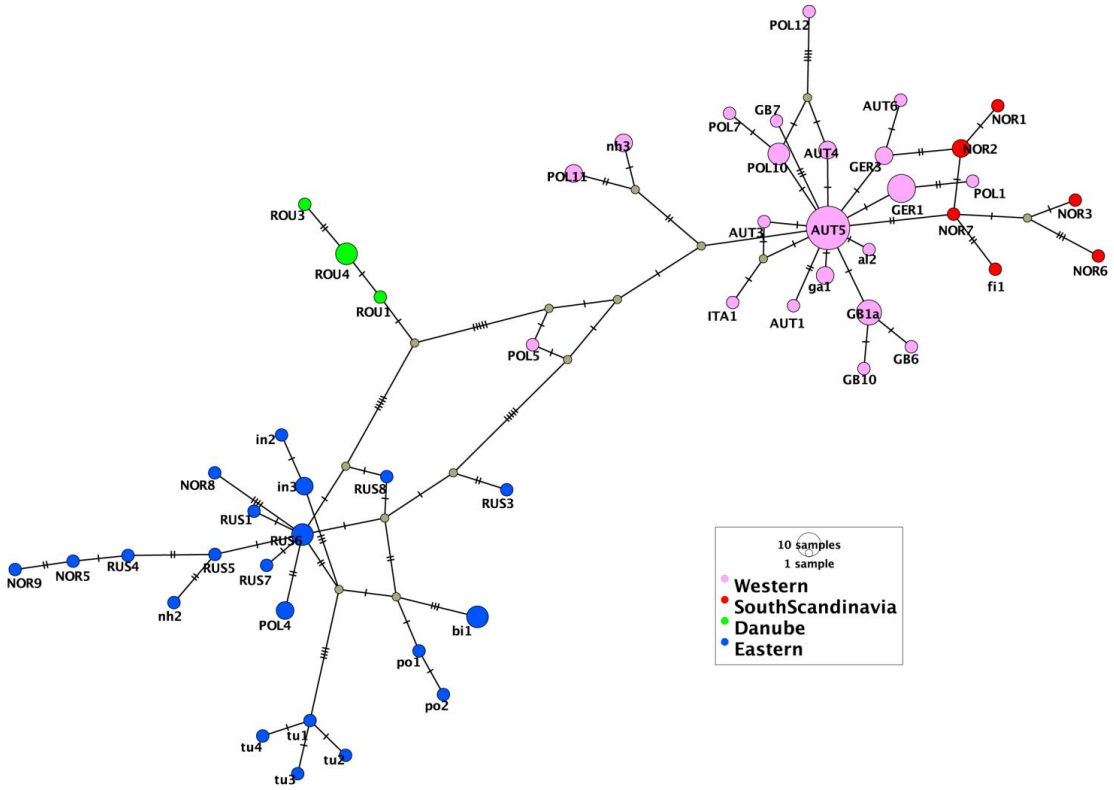


Figure 5.3. **Median-joining network showing relationships among ancient British and modern and ancient Eurasian beaver, *Castor fiber*, mitochondrial DNA tRNAPro/CR haplotypes (492bp)**. Data comprise haplotypes generated here (prefixed 'GB'), ancient haplotypes generated by Horn *et al.* (2012, 2014) and modern relict population haplotypes previously described by Ducroz *et al.* (2005), Durka *et al.* (2005) and Senn *et al.* (2014). Mutational steps are represented by hatched lines across branches and grey dots represent ancestral or missing haplotypes. The two main lineages, the Western and Eastern ESU's, are clearly separated with the South Scandinavian lineage (Western) and Danube lineage (Eastern) both diverging from their respective clades. Network generated using PopART (http://popart.otago.ac.nz) which excludes sites with missing data, gaps and ambiguities. As a result haplotype diversity appears reduced and many haplotypes are collapsed. British haplotypes GB2, GB3a-b, GB4, GB8 and GB9, GB11, GB13, GB14, al1 and POL2 are contained within the AUT5 node GB1a-c and GB 12 are contained in node GB1a.

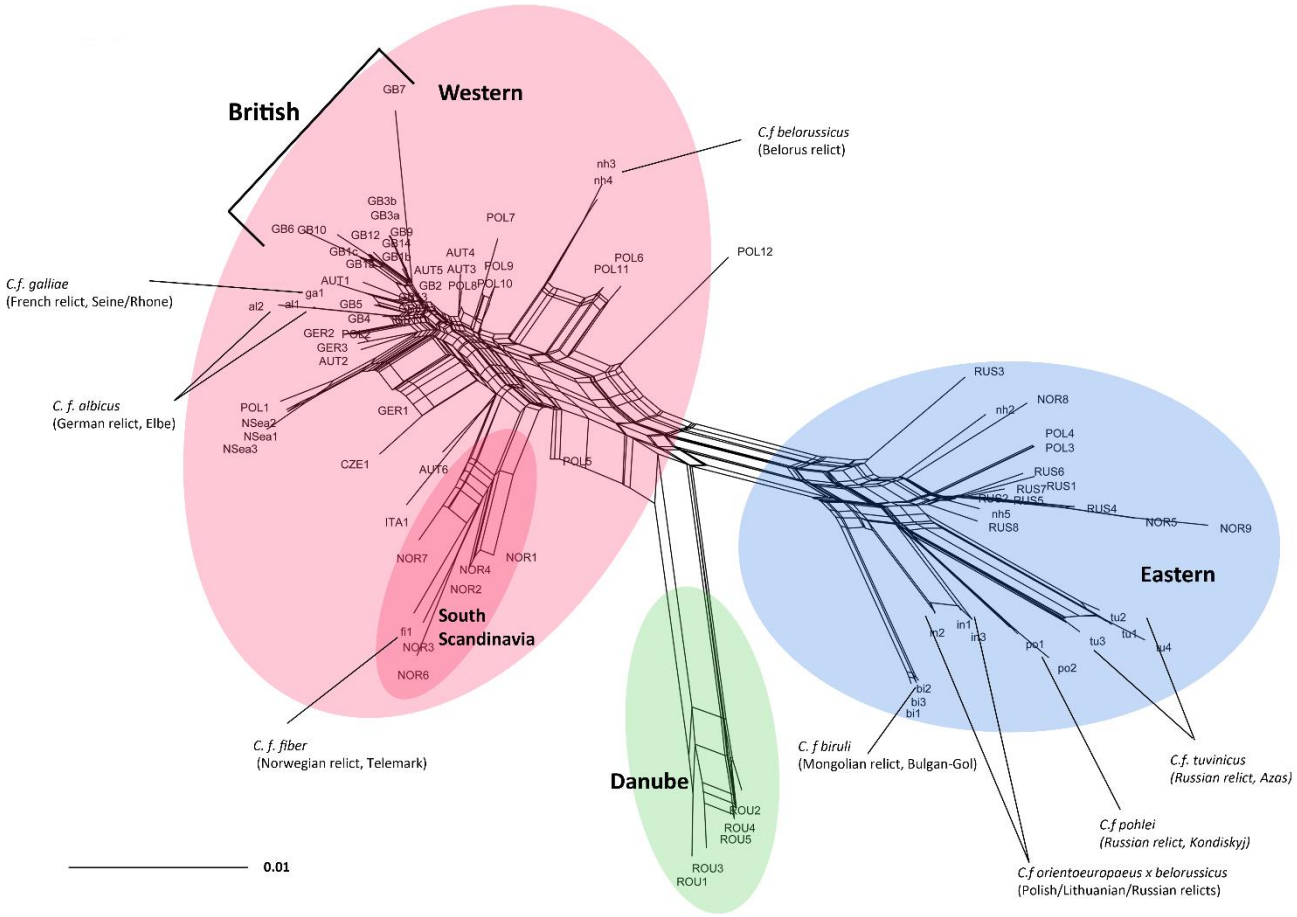


Figure 5.4. Neighbour-Net network showing relationships among ancient British and modern and ancient Eurasian beaver, *Castor fiber*, mitochondrial DNA control region haplotypes (492bp). Data comprise haplotypes generated here (prefixed 'GB'), ancient haplotypes generated by Horn *et al.* (2012, 2014) and modern relict population haplotypes previously described by Ducroz *et al. (*2005), Durka *et al.* (2005) and Senn *et al.* (2014). British data are clustered closely together within the Western ESU.

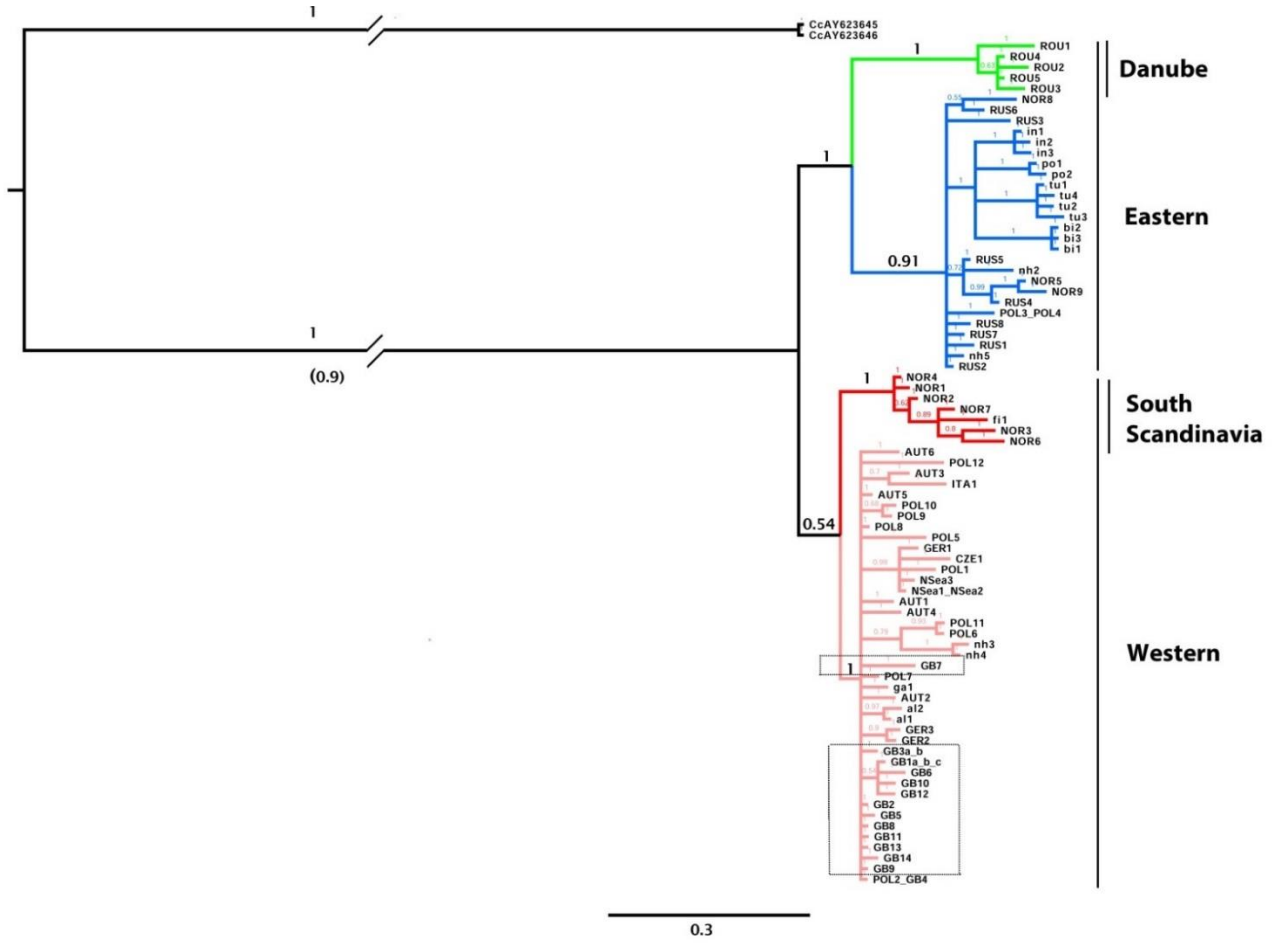


Figure 5.5. Phylogenetic relationships generated in MrBayes showing relationships among ancient British and modern and ancient Eurasian beaver, *Castor fiber*, mitochondrial DNA tRNAPro/CR haplotypes (492bp). A 50% consensus tree with posterior support values for major lineages is shown. The three major clades (Western, Eastern and Danube) and the South Scandinavian sub-clade are all well supported. British data (dashed box) group within the Western clade where they form a large polytomy with samples from both ancient and modern populations (modern haplotypes - *al1, al2, ga1*; Durka. *et al.* 2005). *Caster canadensis* was used as outgroup taxa.

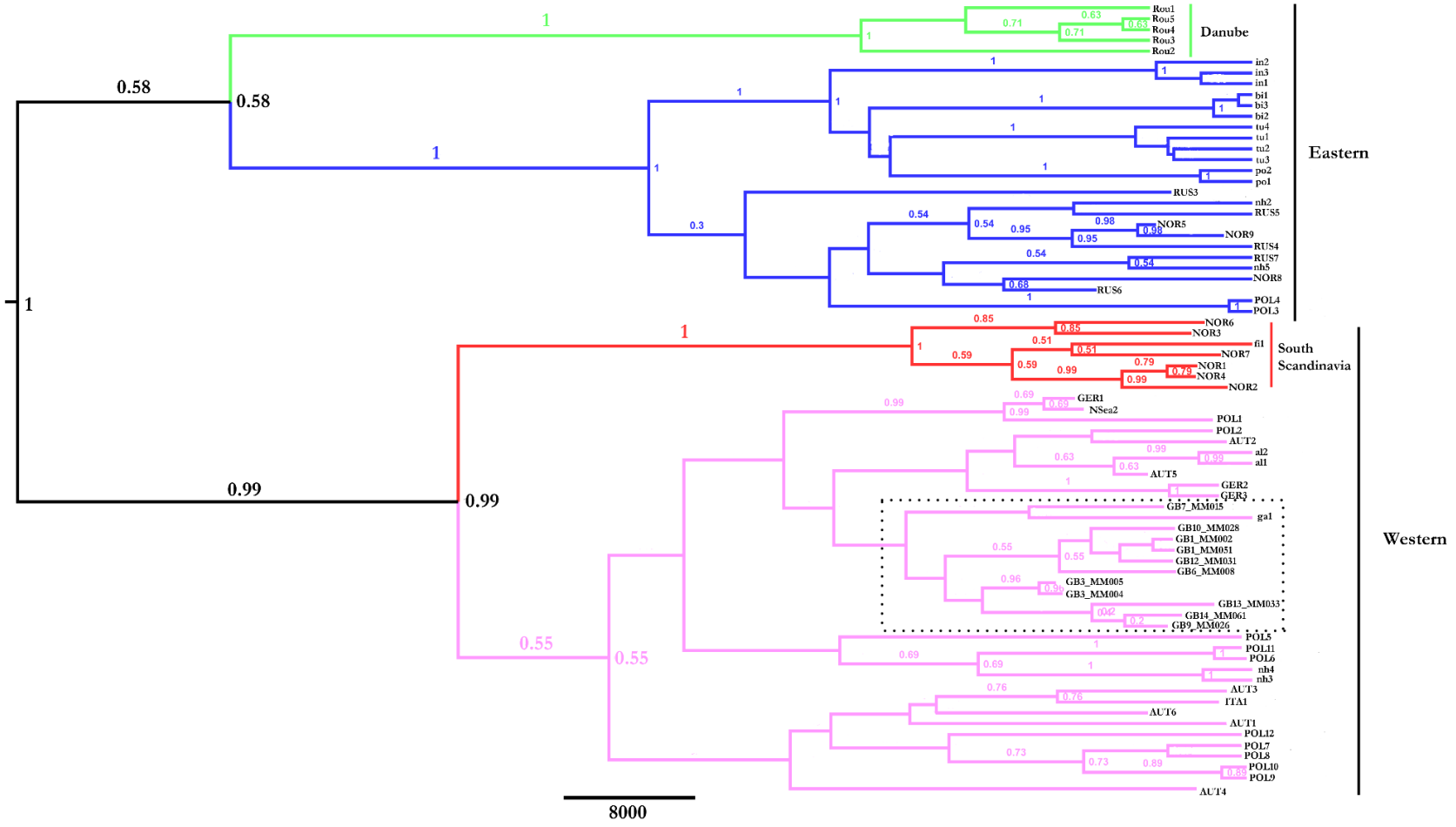


Figure 5.6. Phylogenetic relationships generated in BEAST showing relationships among ancient British and modern and ancient Eurasian beaver, *Castor fiber*, mitochondrial DNA tRNAPro/CR haplotypes (492bp). Shown is a maximum clade credibility (MCC) tree with British samples highlighted in dashed box. The genealogy was generated using a Strick Molecular Clock with a fixed rate of nucleotide substitution set to 5.0 x 10^{-7} and a coalescent expansion model of demography.

5.4 Discussion

This study used ^{14}C radiocarbon dating and ancient DNA to investigate recolonisation of Britain by the Eurasian beaver *Castor fiber* over the Pleistocene-Holocene boundary. Specific questions to be addressed were: *i*) what were the geographical origins and recolonisation routes of Lateglacial (*c.* 14 500 – 12 900 cal BP) beavers into Britain after the Last Glacial Maximum? *ii*) what is the age of the tMRCA of the Western lineage and what is the most likely expansion scenario post colonisation during the LGM? *iii*) did beavers survive the Younger Dryas (*c.* 12,900 – 11,700 cal BP) period of glacial re-advance in Britain? and *iv*) what were the geographical origins of British Holocene beaver populations? With regards to the first question, no full length HV-1 CR mitochondrial DNA sequences were able to be sequenced from Lateglacial beavers so this question remains unaddressed. However, six radiocarbon age determinations (Table 5.3, Fig 5.2) and 17 tRNAPro/HV-1 CR sequences (Table 5.4) were successfully retrieved from beavers dating to the Younger Dryas through to the mid-Holocene.

The main findings were:

- British beavers dating from the very Late Pleistocene (Younger Dryas interval) to the Mid-Holocene form a large, unresolved clade with ancient and modern beavers from Western Europe
- Age divergence estimates from BEAST analyses indicate a time to most recent common ancestor (tMRCA) for the Western clade at mean 50, 733 cal BP with a 95% range of 30, 651 – 73, 056 cal BP. This suggests a tMRCA for this lineage which predates the LGM
- Tests for demographic expansion (Tajima's D, Fu's F, R2 and a mismatch distribution) all indicated a rapid population expansion in the Western clade
- AMS ^{14}C radiocarbon dating of two specimens from Gough's Cave confirmed the presence of beaver in Britain during the Younger Dryas cold phase
- Genetic diversity in Late Pleistocene –Holocene British beavers appears higher than in extant European beaver populations

5.4.1 Accelerator Mass Spectrometry (AMS) radiocarbon dating

Direct Accelerator Mass Spectrometry (AMS) ^{14}C age estimates gave a probability distribution of two *C. fiber* samples from Gough's Cave to the Younger Dryas cold phase (MM004, 11,989-11,405 cal BP; MM005, 12,386-11,836 cal BP: Table 5.3, Figure 5.2). The *chi*-square test indicated that on the basis on those dates these specimens were likely discrete individuals. This result firmly places at least one, probably two, *C. fiber* individuals in SW England during this cold period. Beavers presumably recolonised Britain at the commencement of climatic amelioration during the onset of the Lateglacial period but it was unknown whether or not they fully retreated to the European mainland when the climate abruptly deteriorated at the onset of the Younger Dryas. This result strongly suggests that a population of beavers persisted in SW England well into the Younger Dryas cold phase and possibly into the Holocene. This population is likely to have been a relict from a larger, Lateglacial, population that suffered contractions in size and distribution with the rapid change to a colder climate. There are two alternative scenarios regarding the fate of this population which are difficult to determine with the phylogenetic results in this study, *i*) the population persisted into the Holocene (*i.e* continuity over the Younger Dryas – Holocene boundary) where it kept pace with suitable habitat as it became available therefore expanding and contributing, to an unknown extent, to later British Holocene beaver populations or *ii*) this population became extirpated (*i.e* discontinuity over the Younger Dryas – Holocene boundary) prior to the abrupt transition to warmer, Holocene, conditions. The 95.4% probability limits on the radiocarbon dates place these two individuals in the mid-closing phases of the Younger Dryas with a probability limit extending into the very Early Holocene. This suggests that a beaver population had persisted in this region for many centuries, surviving the quick transition to a colder climate. Together with the extremely low level of genetic divergence between these two specimens and later Holocene individuals (Figures 5.3 – 5.6), this would tend to support a hypothesis of population continuity over the Younger Dryas cold phase.

Despite their reliance on riparian woodland ecosystems, beavers have been observed surviving in the far North of Scandinavia in suboptimal habitat (Halley 2011). In

Norway, beaver populations inhabit areas of the montane willow scrub zone at 900m meters above sea level on the extreme tree line where ice cover often lasts for 7 months of the year (Rosell & Pederson 1999). This suggests that it is possible for this site to have supported a (presumably small) population of *C. fiber* if at least some woody scrub vegetation persisted. Based on a range of ecological and genetic parameters, the effective reproductive size (N_e) for beavers has been estimated to be as low as 3 animals (Milishnikov 2004). This is an extremely low value of effective reproductive size for a medium-sized mammal and implies that this species has an extremely high tolerance to the deleterious genetic consequences associated with small, isolated populations such as genetic drift and inbreeding. Some modern European relict populations have persisted for many decades at very low population densities before conservation measures allowed expansion (Durka *et al.* 2005; Ducroz *et al.* 2005; Halley *et al.* 2011). All Western relict populations survived the 20th century with <150 individuals, with the French relict population size estimated at just 30 individuals (Richard 1985). Presumably the effective female population was much lower. None of the relict populations had over 300 individuals (Senn *et al.* 2014 and references therein). Therefore, it is reasonable to assume that beaver populations are capable of persisting in low numbers in isolated relict populations for some (unknown) length of time, from which they can rapidly expand given favourable environmental conditions. For example, the introduced Swedish population is derived exclusively from the small, relict population in Telemark, Norway, which at one point, had a population size of just *c.* 60-120 individuals (Collett 1897). From a founder size of *c.* 100, the introduced Swedish population increased to over 170,000 individuals within a century (Halley & Rosell 2003). Therefore, while identifying beavers of Younger Dryas age in Gough's Cave was an unexpected result, it is wholly possible for this species to have persisted in this region over the Younger Dryas cold phase.

The radiocarbon dates obtained here also have implications for the interpretation of the faunal record at Gough's Cave. The dominant bulk of the faunal assemblage is predominantly an Upper Palaeolithic, Lateglacial, human-predation assemblage dominated by temperate-adapted mammals such as *Equus ferus*, (wild horse) and *Cervus elaphus* (red deer; Currant 2001). The faunal record is extensive and directly dated

faunal elements date from 15,238 – 14,462 cal BP (unidentified *herbivore*, OxA-18035: Jacobi & Higham 2009) to 12386-11358 cal BP (*Rangifer tarandus*, OxA-1120: Hedges *et al. 1987*) and show a general (but not strict) pattern of faunal turnover across the abrupt climatic changes between the Lateglacial Interstadial– Younger Dryas – Holocene transitions. However, the ^{14}C age determinations from the two Gough's Cave beavers in this study are the first to place a woodland specialist definitively in the Younger Dryas faunal assemblage at this site. Formerly, the youngest dated specimen from Gough's Cave was the *C. fiber* specimen, MM004 (11068-10957 cal BP, OxA-1119: Hedges *et al.* 1987), originally dated in 1987 before the introduction of ultrafiltration. The re-date on this (using chemical pre-treatment and ultrafiltration) ages this specimen at considerably older (11,989-11,405 cal BP). Jacobi & Higham (2009) carried out an extensive program of repeat dating on Gough's Cave specimens and in some instances, found that significant differences in age estimates exist between samples dated pre and post introduction of current filtration methods. This suggests that caution should be used when indirectly estimating the age of other woodland specialists excavated from this site that also have the ecological potential to survive in small, micro-refugia *e.g* wild cat *Felis silvestris silvestris*.

5.4.2 Phylogenetics and genetic diversity

Both network and Bayesian MCMC methods uncovered highly congruent results (Figures 5.3 – 5.6). The phylogenetic analyses all recovered tRNAPro/CR divergence in three main clades, Western Europe, Eastern Europe and Danube with a South Scandinavian subclade showing a low level of divergence from the Western clade. The British data all grouped firmly within the Western clade and showed no supported spatio-temporal substructure, either internally within British populations or between the British haplotypes and the Western European data. The two Late Pleistocene specimens MM004 and MM005 did not show any notable degree of divergence from the later Holocene British haplotypes, which suggests that this population contributed genetically to the establishment of British Holocene populations. These were the oldest specimens in the entire database of British and Eurasian ancient beaver haplotypes and the only two individuals dating to the Late Pleistocene (Table 5.2).

More sampling from both British and Eurasian ancient beaver populations is needed to examine fully the phylogenetic relationships between beaver populations of Holocene and Late Pleistocene age. Water voles, *Arvicola terrestris*, show two temporally separated 'waves' of recolonisation and population replacement into Britain which are concordant with climatic episodes over the Pleistocene-Holocene transition (Brace *et al.* 2016). While the data presented here does not suggest such a pattern for Eurasian beavers, it cannot be discounted that hidden phylogenetic structure exists until further sampling improves the spatial and temporal resolution of the haplotype database used here.

The over-all pattern observed here was one of a remarkable lack of phylogeographic structure and low genetic divergence between pre-bottleneck British and Western European mainland ancient beaver populations. The analyses all recovered the split between Western and Eastern ESUs with an *Fst* value between the clades of 0.7. British samples did not show any phylogenetic affinity with Eastern ESU haplotypes. Therefore, there was almost certainly no Postglacial recolonisation of Britain from Eastern Europe and in the (unlikely) scenario that Eastern European beavers contributed to Lateglacial recolonisation of Britain, these populations are likely to have been replaced by a second of wave of Early Holocene recolonisation from Western Europe. Again, more geographically and temporally wide-spread sampling of Late Pleistocene beavers would be required to resolve this fully.

The pre-LGM node age estimate for the Western clade (tMRCA mean = 50, 723 cal BP, 95% HPD range = 30, 651 – 73, 056 cal BP), significant and negative Tajima's D statistic and lack of phylogenetic substructure are all strong indicators of rapid population expansion from glacial refugia. Given the evidence presented here, it seems clear that the Holocene British beaver population formed part of a large, panmictic, Pan-Western European population. This population likely expanded rapidly from glacial refugia after the LGM, had high levels of gene flow and low levels of regional divergence within the control region loci. Moreover, out of 17 haplotypes described here, 13 were unique and none were shared with extant relict populations. This adds further support to the conclusions drawn by Horn *et al.* (2014) that ancient

beaver populations harboured much greater levels of genetic diversity than is seen today in post-bottleneck populations.

The coastline of NW Europe in the early Holocene extended from NW France, throughout Britain and, in the east, the landscape of Doggerland extended from the east coast of Britain to Scandinavia, where a landbridge allowed fauna access from modern-day Denmark to Sweden (Coles 2006). Therefore, a huge potential area of suitable habitat was available for woodland mammals such as *C. fiber* with little geographical barriers to gene flow. Notably, geophysical changes brought about the inundation of the British and Scandinavian land-bridges within 1000 years of each other. The Storegga Slide tsunami is hypothesised to have caused a catastrophic flood that completely severed the remaining land links between Britain and Europe *c.* 8,200 cal BP (Weninger *et al.* 2008). Processes associated with deglaciation of the Fennoscandian ice sheet initiated the formation of the Baltic Sea and Dana River *c.* 9,200 cal BP, which fully separated the Scandinavian Peninsula from mainland Europe (Björk 1995). Beavers likely colonized Scandinavia both from the south and east as a contact zone can be observed in Northern Sweden and Norway (Figure 5.1; Horn *et al.* 2014). South Scandinavian beavers have since become divergent from the rest of the Western Europe clade whereas British beavers have not, despite spending a similar amount of time in isolation from populations on the European mainland. The reasons for this are unclear but could involve genetic factors such as founder effects. The range of sampled ages for the South Scandinavian haplotypes used here is 5950 to 1650 cal BP (Horn *et al.* 2014) while the British Holocene data range from 4973-2211 cal BP (Table 5.3), so both populations have spent a similar amount of time in isolation. However, the *Fst* values between the Scandinavian sub-clade and the rest of the Western clade were extremely small (*Fst*=0.054).

5.4.3 Conservation implications

The Scottish Beaver Trial was the first official re-introduction of an extirpated mammal into the UK when three families of beavers were introduced into Knapdale, Western Scotland in 2009 (Jones & Campbell-Palmer 2014). Following post-release monitoring, it is likely that this project will be expanded and reintroductions into

England and Wales will follow (Gurnell *et al.* 2009). Guidelines for reintroductions must follow both national and international criteria and identifying suitable source populations from which to remove and translocate individuals is a fundamental consideration in such an undertaking (Batson *et al.* 2015). The IUCN (2013) guidelines indicate that efforts should focus on identifying those populations that are the most historically similar to the original extirpated population; the assumption being that these populations have adaptations already in place that will increase the likelihood of obtaining a successful breeding population in the new environment. However, this assumption does not always hold true. Closely related populations may not show the same genome-wide adaptive traits, particularly if they inhabit different environmental conditions, and distantly related populations may evolve similar adaptive traits via parallel evolution. There has been much debate in the literature as to whether phylogenetic affinity is a consideration that should be undertaken with regards to the British reintroductions (Halley 2001; Rosell *et al.* 2012; Senn *et al.* 2014) with some authors arguing that other genetic factors such as the level of genetic adaptive potential should have higher priority (Senn *et al.* 2014; Frosch *et al.* 2014).

Halley (2011) set out three alternative scenarios for deciding on the choice of source population for British reintroductions, *i*) use of the geographically closest beaver relict population *ii*) mixture of individuals from two or more Western populations and *iii*) release of individuals of multiple origins, regardless of ESU assignment. The complete lack of phylogeographic structure observed in this study within the Western Europe clade suggests that extinct ancient Late Pleistocene and Holocene British beavers did not possess any significant genetic divergence from mainland European ancient beavers (Figures 5.3 – 5.6). Nor did they show any difference in the degree of phylogenetic relatedness to any of the mainland Western extant relict populations in Germany (*al* haplotypes) or France (*ga* haplotypes). Within the Western clade, the most divergent population from the British (and other Western Europe) haplotypes was the South Scandinavian subclade, the extant population of which has been the source of individuals for the successful Knapdale reintroduction (Jones & Campbell-Palmer 2014). Also, given the extremely low levels of divergence in mitochondrial DNA among extant Western Europe populations, it is likely that all of these individual

populations will show a high level of genetic similarity to the ancient British population. Evidence from Senn *et. al.* (2014) highlighted that the division between Western and Eastern ESUs is not as well-defined as originally thought and argued that this division should be abandoned. The British data does not show any phylogenetic affinity with the Eastern clade. Pending the adoption of the recommendation from Senn *et al.* (2014), or not, the fact that the British ancient beaver population does not show any phylogenetic affinity with the Eastern clade leads to a recommendation to source individuals from inside Western Europe for British reintroduction, supporting scenarios *i*) or *ii*) from Halley *et al.* (2011).

It should be kept in mind that there are other genetic considerations involved in choosing source populations for translocation purposes, which may be given more weight than phylogenetic history. Populations of *C. fiber* are already present in Britain on the Tay and Earn catchments in Perthshire, Scotland (Campbell *et al.* 2012) and the River Otter in Devon as the result either of deliberate illegal releases or as accidental escapes from captive populations (Campbell *et al.* 2012; Carter *et al.* 2016). While the Devon population has been removed from the wild, health screened, re-released and is now heavily managed (Carter *et al.* 2016); the Tay beaver population has largely been allowed to expand naturally with minimum intervention thus far. From an unknown founder number, the population now stands at around 38 family groups (*c.* 146 individuals) that appear to be thriving (Campbell *et al.* 2012). Genetic screening has established that these individuals descend from the Bavarian region in Germany (McEwing *et al.* (2015), a highly admixed population with individuals of Scandinavian, Russian and French origin (Frosch *et al.* 2014), *i.e* not only of mixed Western European ancestry but also originating from different ESUs. European beaver populations with admixed ancestry generally tend to have elevated levels of genetic diversity relative to unmixed relict populations and display no obvious signs of outbreeding depression (Frosch *et al.* 2014; Senn *et al.* 2014). Therefore, it may be advantageous for future reintroduction projects to consider selecting founder stock that maximise genetic diversity and levels of adaptive potential above achieving close phylogenetic relatedness to the original historical population.

Chapter 6: Phylogeny and genetic diversity in wildcat *Felis silvestris silvestris* from the British Lateglacial to the Holocene.

6.1. Introduction

6.1.1 Overview

The European wildcat *Felis silvestris silvestris* is a mesocarnivore and member of the polytypic *Felis silvestris* species group that inhabits temperate woodland areas of southern and central Europe, from Britain to European Russia (but is absent from Scandinavia; Lozano *et al.* 2013). Despite being a flagship mammal of conservation concern due to habitat loss, persecution and introgression with the domestic cat, *F. s. catus* (Hubbard *et al.* 1992; Pierpaoli *et al.* 2003; Lozano *et al.* 2003; Yamaguchi *et al.* 2004; Kitchener *et al.* 2005; Oliveira *et al.* 2008a; Oliveira *et al.* 2008b; Kitchener & O'Connor 2010; Mattucchi *et al.* 2014), surprisingly little is known about the phylogeographic and population history of this subspecies, particularly in Britain. Carnivores are generally considered to be fairly resilient to abrupt changes in climate as they do not usually specialise on a narrow climatic niche (Koenigswald 2002, 2003). Sommer & Benecke (2006) have suggested that this may lead to a less drastic response in faunal turnover of felid carnivores in response to climatic changes than is seen in taxa that show clear preferences to warm or cold climes and vegetation.

Unlike many felid species, *F. s. silvestris* shows a strong preference for temperate woodland and population densities are highest in broad-leaved or mixed species forests (Yamaguchi *et al.* 2015). This suggests that they may have followed a pattern of Late and Postglacial recolonisation of Britain associated with the spread and retreat of woodland habitats, namely *i*) recolonisation and expansion from mainland Europe following the spread of woodland during the Lateglacial, *ii*) habitat tracking southwards or local extinction during the Younger Dryas cold phase and *iii*) population expansion or recolonisation during climatic ameroilation of the Holocene. However, carnivores have been shown to have been adept at adapting to many different environments over the course the Pleistocene climatic fluctuations (Koenigswald 2002, 2003). Moreover, the European wildcat now has extant

populations in Scotland that survive in treeless environments, albeit prevented from moving southwards by the Scottish industrial central-belt (Harris *et al.* 1995), and ^{14}C dates from subfossil bone places *F. s. silvestris* in central Europe during the Younger Dryas cold phase (Germany, France and Denmark; Degerbøl 1933; Altuna *et al.* 1991; Staesche 1994). While it must be kept in mind that these are all ^{14}C dates performed before the introduction of ultrafiltration (See Chapter 2, Section 2.2.1) this implies that the subspecies has some potential for adaptation to (assumed) suboptimal environments or, alternatively, *F. s. silvestris* may have a broader climatic envelope than is currently hypothesised. British wildcat populations almost certainly experienced demographic declines and range contractions during the Younger Dryas associated with the decline of woodland and development of open landscapes. However, it remains unknown and difficult to predict whether populations were able to persist in Britain in sheltered microclimates over this cold phase and, if so, whether these populations contributed to descendant Holocene populations.

Demographic declines and genetic admixture with the domestic cat have been experienced to varying degrees throughout many European wildcat populations (Beaumont *et al.* 2001; Pierpaoli *et al.* 2003; Lecis *et al.* 2006; Oliveira *et al.* 2008a O'Brian *et al.* 2009; Driscoll *et al.* 2011; Oliveira *et al.* 2015) but have been particularly severe in Britain where *F. s. silvestris* is now restricted to northern areas of Scotland (Nowell & Jackson 1996; Balharry & Daniels 1998; Kitchener 1998; Kitchener *et al.* 2005; Mattucchi *et al.* 2014; Oliveira *et al.* 2015). The genetic consequences of these events will be superimposed on the genomes of extant wildcats at both the individual and population level (Amato *et al.* 2009; Xue *et al.* 2015). The signatures of recent and historical admixture and population reductions may obscure ancient signals of demographic processes, particularly in mitochondrial DNA (Ballard & Whitlock 2003). This creates difficulties in reconstructing phylogeography, elucidating past population histories and quantifying ancient levels of genetic diversity from sequence data obtained from extant individuals. Ancient DNA is therefore an ideal means with which to circumvent these problems, allowing direct inference of population processes prior to population declines and introduction of the domestic cat. Ancient DNA analyses have previously uncovered hidden and unanticipated phylogenetic and

demographic information in large felids (*e.g* lions *Panthera leo* ssp. Burger *et al.* 2004, Barnett *et al.* 2009). For example, sequencing of the Hyper Variable I (HV-I) region of the Control Region (CR) in the Iberian lynx *Lynx pardinus* has shown this species to have had inherent low genetic diversity over the last 50, 000 years (Rodríguez *et al.* 2011). This implies that small effective female population size is typical for this species and that conservation concerns about genetic viability due to low levels of mtDNA variation in extant populations of the Iberian Lynx are misplaced.

This chapter aims to gain similar insights, if possible, from ancient DNA analyses of both mitochondrial and nuclear DNA extracted from British subfossil material of the European wildcat *F. silvestris silvestris*. Availability of material from this species in Britain is limited and confounded by potential chronological overlap with the domestic cat *F. s. catus* during the Holocene. Sample sizes obtained here are consequently much lower than would have been preferred. Nevertheless, the evidence gained from this material is used to achieve the broad objectives of:

1. Quantifying levels of genetic diversity in the European wildcat *F. s. silvestris* from the Late Pleistocene to the Late Holocene in Britain
2. Determining if European wildcats survived the Younger Dryas cold phase in sheltered microhabitats in Britain

   and

3. Attempting to infer either population continuity or replacement over the Younger Dryas cold phase

6.1.2 Phylogeny and distribution

Felis silvestris silvestris is a member of the geographically widespread species group *Felis silvestris,* which is distributed throughout central and southern Europe, Africa and central Asia (Fig 6.1; Driscoll *et al.* 2007). Despite several studies being undertaken looking at geographical variation in both morphology and genetics, there is still no clear consensus as to the number of subspecific divisions within this group (Driscoll

et al. 2007; Kitchener and Rees 2009; Yamaguchi *et al.* 2015). As this study is concerned primarily with mitochondrial DNA evidence, the subspecies ranks derived from mtDNA analysis described by Driscoll *et al.* (2007, 2011) are adopted here. This classification suggests five matrilineages representing five distinct subspecies including the domestic cat *F. s. catus,* which is subsumed into the lineage represented by the Near Eastern wildcat, *F. s. lybica*. It should be acknowledged that other authors suggest alternative subspecies arrangements (*e.g* Kitchener & Rees 2009 who suggest three lineages). Differences in taxonomic classifications of the *Felis silvestris* group are unlikely to affect interpretations made in this study as pure bred European wildcats and domestic cats fall into discrete mtDNA genetic lineages that are clearly distinct from one another in the absence of admixture (Driscoll *et al.* 2007, Fig 6.1).

The five (roughly) allopatric subspecies within *F. silvestris* are fully inter-fertile (Sunquist & Sunquist 2002). Mitochondrial DNA evidence has identified five major biogeographic matrilineages – Clade I, European wildcat *F. s. silvestris*, provenanced to Europe; Clade II Southern African wildcat *F. s. cafra*, provenanced to southern Africa; Clade III, Asian wildcat *F. s. ornata*, provenanced to central Asia, Clade IV, *F. s. lybica* Near Eastern wild cat (including the domestic cat *F. s. catus*), provenanced to Africa and the Arabian Peninsula and Clade V, *F. s. bieti*, Chinese Desert cat (Fig 6.1; Driscoll *et al.* 2007). Based on coalescent analyses of the NADH dehydrogenase ND5 and ND6 genes, and assuming a divergence rate of 2.24bp/MY, diversification of the *F. silvestris* subspecies is estimated to have been initiated *c.* 230 kyr (Driscoll *et al.* 2007; Fig 6.1). The European wildcat represents one of the earliest subspecies to diverge, although no robust date for the timing of this split has yet been suggested.

Early palaeontological evidence for the species in Europe is limited to the regions of the multiple Mediterranean peninsulae (Sommer & Beneke 2006). This places the first ancestors of the European wildcat in Britain and the rest of Europe during MIS 11 (424-374 ka BP; Schreve 1996). The subspecies is also known to have been a member of European temperate faunal assemblages during the later Eemian/Ipswichian Interglacial and Pleniglacial (115,000 to 130,000 years ago) where it is recorded from sites throughout Europe (Spain, Zeuner 1953; Italy, Britain,

Sutcliffe 1960; Sutcliffe & Zeuner 1962; Pitti & Tozzi 1971; SW France, Castanos 1984 and Moldova David 1999). Subfossil records from the Late and Postglacial are more numerous and equally geographically widespread with wildcat recorded in many north-west European regions such as Britain (Collcutt *et al.* 1981), Denmark (Degerbøl 1993) and Germany (Reisch 1976; Jullien 1981) in addition to the peninsulae of Southern Europe.

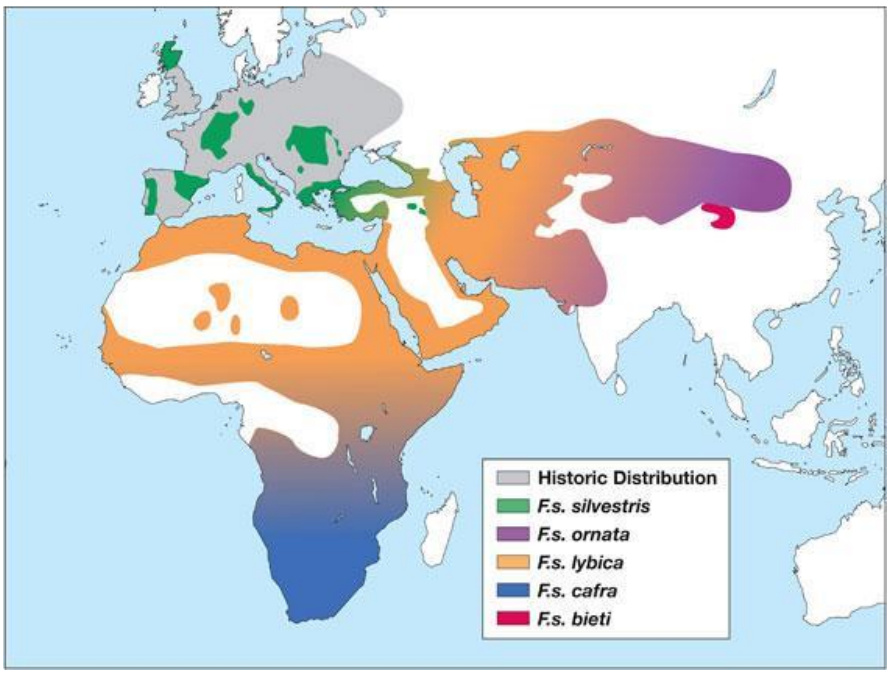


Figure 6.1. African and Eurasian distribution of the *Felis silvestris* subspecies complex. Adapted from Driscoll *et al.* (2007).

There is strong evidence that the European wildcat retreated to the Iberian, Italian and Balkan peninsulae during periods of major glacial re-advance from which they then re-colonised central Europe. Radiocarbon dating of stratigraphic layers within which wildcat remains have been found place it in Italy during the height of the LGM (Boscato & Palma di Cesnola 2000) and in Spain immediately after the LGM (Altuna 1972). In addition, genetic evidence from microsatellite markers genotyped from extant wildcat populations in central and southern Europe and analysed with Approximate Bayesian Computation simulations (ABC) reveals a genetic bottleneck dating from *c.* 60,000 to 10,000 years ago (Mattucci *et al.* 2016). The lack of support for models simulating recent, anthropogenically-driven, bottlenecks implies that the most likely scenario for the observed partitioning of genetic information among

extant populations is therefore a bottleneck creating by retreat to glacial refugia during the LGM followed by re-expansion during climatic amelioration (Mattucci *et al.* 2016). If, and to what extent, this species utilised cryptic Northern refugia during the LGM is unknown.

The European wildcat was once distributed widely throughout central and Southern Europe but endemic populations have suffered declines, range contractions and local extinctions due to the destruction and fragmentation of woodland habitat coupled with varying degrees of persecution (Yamaguchi *et al.* 2015; Fig 6.1). Formerly contiguous populations in the European mainland are now fragmented across the Iberian Peninsula, NE France, central Germany, Italy, Greece, Eastern Europe, Turkey and the Caucasus (Stahl & Artois 1991; Nowell & Jackson 1996; Mitchell-Jones *et al.* 1999; Fig 6.1), although significant heterogeneity exists in the size and degree of isolation of populations. In Britain *F. s. silvestris* has been extirpated from England and Wales and is now restricted to Scotland (Yamaguchi *et al.* 2004; Davis & Gray 2010). While small population increases have been reported for a few populations (*e.g.,* Switzerland and Belgium), many continue to decline (IUCN 2007). In Poland and Scotland populations are known to be critically decreasing (IUCN 2007; Davis & Gray 2010) and rates of decline in the Iberian population were estimated at *c.* 30 percent over the last three generations at the time of reporting (IUCN 2007).

As in the case of the European beaver *Castor fiber*, the effects of historic and ongoing anthropogenic activity on wildcat populations have created difficulties in reconstructing an accurate phylogeographic history for this species from modern populations. The genetic effects of population reduction, isolation, local extinctions and widespread genomic admixture with domestic cats are incorporated into the genomes of extant individuals. This means that extensively hybridised extant individuals cannot be included in phylogenetic analyses (such as the remnant British population in Scotland and those in Hungary; Mattucci *et al.* 2016). Furthermore, the contemporary range of extant wildcats does not include many geographic areas that are key for answering phylogeographic questions (Fig 6.2). However, data from

neutral microsatellite markers typed from (presumptive) purebred wildcats at European localities with low or absent admixture shows genetic subdivision into five main biogeographic groups; 1 - NE Italy and the Northern Balkan regions, 2 - Italy and Sicily, 3 - central Germany, 4 - central Europe, 5 - Iberia (Mattucci *et al.* 2016; Fig 6.2). Intensive sampling focused on Germany and Luxembourg by Steyer *et al.* (2016; Fig 6.3) using mtDNA and microsatellite data shows a split between a western wildcat cluster, largely provenanced to Luxembourg and West Germany and a central German cluster. The west cluster corresponds to lineage 4 in Mattucci *et al.* 2016 and includes France, Belgium and Luxembourg while the central German cluster corresponds to lineage 3 of Mattucchi *et al.* (2016), representing central Germany in both studies. This suggests a potential contact zone for the wildcat in east-central Europe roughly located between Luxembourg/western Germany and central Germany (Fig 6.3).

6.1.3 Introgression with the domestic cat Felis silvestris catus

The fate of the European wildcat is intricately linked with that of the domestic cat, *F. s. catus*. This subspecies has a long association with humans with the first evidence of domestication uncovered in Cypriot deposits from *c.* 9.5 kya (Vigne *et al.* 2004). The evolutionary process of domestication is complex and genetic signatures of domestication may be obscured by admixture events with wild progenitors, particularly in the mitochondrial genome, which can be rapidly replaced (Ballard & Whitlock 2003). In general, domestication can take three alternate pathways (Larson & Burger 2013): *i*) Directed – where wild species were deliberately selected for a specific purpose (*e.g* equids for transport) *ii*) Prey – where wild-living species that were predated by early humans eventually became managed and subsequently artificially selected for traits useful in agriculture (*e.g* cattle, goat, sheep) or *iii*) Commensal - where a wild-living species initiates the first stage of domestication by becoming habituated to human-created environments and management and/or artificial selection by humans occurs at a later stage (*e.g* canids; Larson & Burger 2013; Frantz *et al.* 2016). The process of wildcat domestication is hypothesised to have begun via the third, commensal, pathway where the wild progenitors of domestic cats fed on small mammal pests associated with the grain stores of early farmers (Clutton-

Brock 1999; Driscoll *et al.* 2009; Vigne *et al.* 2011; Hu *et al.* 2014). An increase in anthropophillic affinities in these individuals then led to domestication via 'self-selection', where behavioural changes related to exploitation of novel 'urban' human environments created reproductive isolation as a direct consequence of assortative mating (Vigne *et al.* 2011a; Larson & Burger 2013; Hu *et al.* 2014).

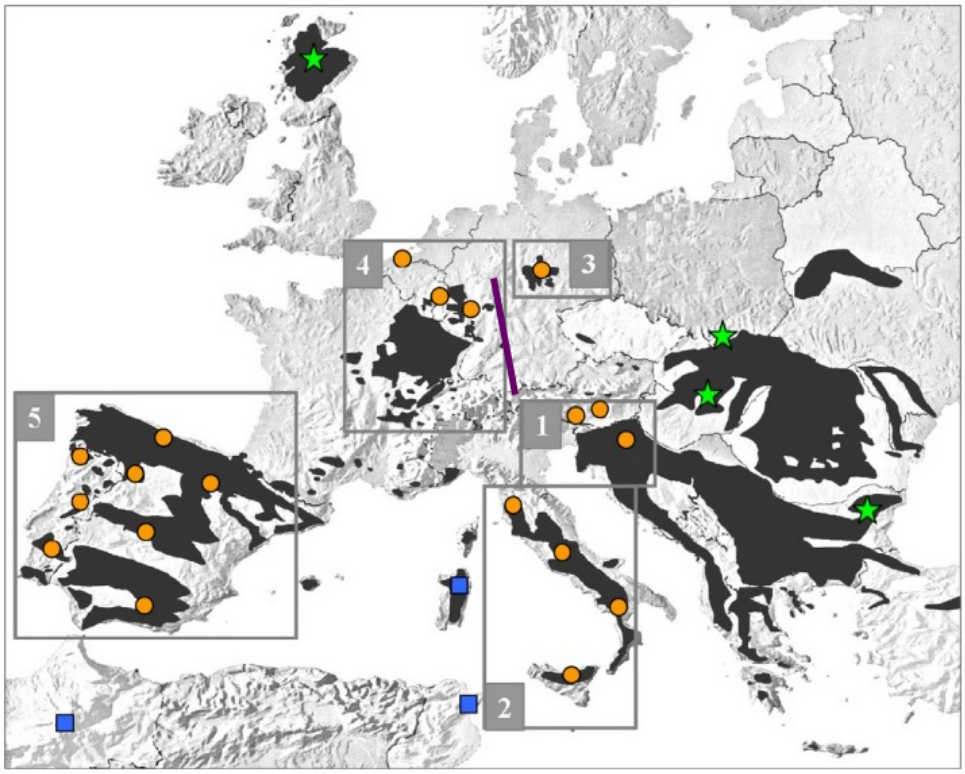


Figure 6.2. European distribution of *Felis silvestris silvestris.* Green stars indicate heavily admixed populations that were not considered in the phylogenetic reconstruction by Matucci *et al.* (2016). The five boxes indicate the five main biogeographic regions and blue boxes indicate location of *F. s. lybica.* Purple division in Germany shows region of potential contact zone.

The domestic cat is descended from at least five matrilineal haplotypes of the Old World wildcats, specifically the Near Eastern wildcat, *F. s. lybica* (Driscoll *et al.* 2007). Initiation of the domestication process is therefore most likely to have begun in the Near East in the region of the Fertile Crescent, an area extending from the Mesopotamian plains through the Taurus Mountains, across the Mediterranean coast to the Levant and towards the Indus valley (Driscoll *et al.* 2007; Lipinski *et al.* 2008). This area is one of three regions identified as a principal hub of domestication for many species (Bruford *et al.* 2003) including sheep (*c.* 11,000 cal

BP), pigs (*c.* 10,500 cal BP), cattle (*c.* 10,000 cal BP) and goats (*c.* 11,000 cal BP; Vigne *et al* 2001b).

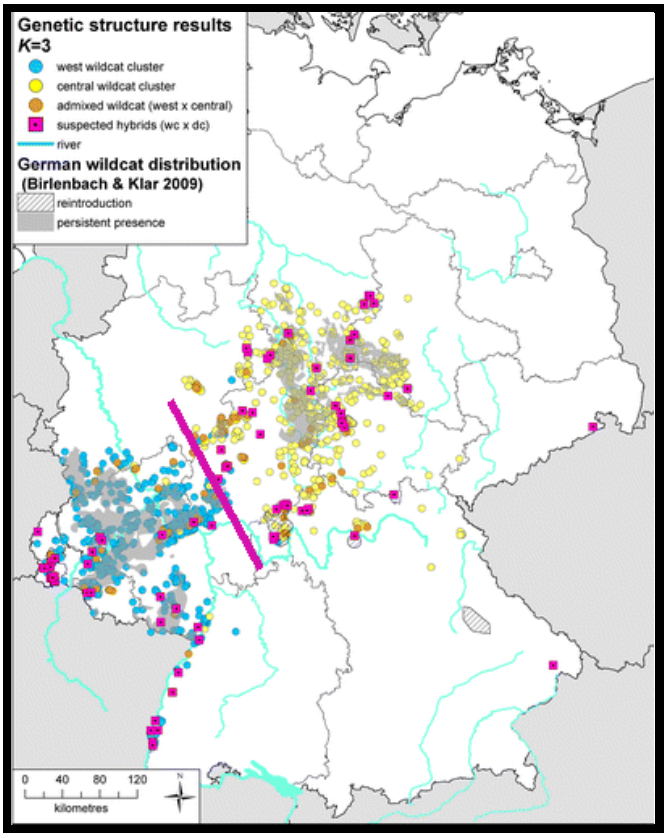


Figure 6.3. Microsatellite derived genotype clusters for *Felis silvestris silvestris* in Germany and Luxembourg. Blue circles - true wildcats belonging to the western cluster; Yellow circles – true wildcats belonging to the central cluster; Orange circles - admixed (west x central) wildcats; Purple squares – suspected hybrid wildcat x domestic cat.

Some authors have argued for the existence of multiple independent domestication events at different time points and geographic locales for certain species (Larson & Burger 2013). However, migrating domesticate populations often remain inter-fertile with their wild progenitors and genetic introgression can occur via hybridization between the resident wild population and migrating domestic population (Larson & Burger 2013). This hybridization process can result in the replacement and turnover of mtDNA over relatively small timescales creating a genetic signature that can mask prior demographic processes (Ballard & Whitlock 2003). This can ultimately be misinterpreted as a second, independent domestication event. The nuclear genome will retain information relating to domestication and demographic processes over

longer timescales than mtDNA but, thus far, these studies have tended to focus exclusively on humans (Green *et al.* 2010; Reich *et al.* 2010).

There has been very little research undertaken on the fine-scale spatio-temporal domestication process in wildcats. While Single Nucleotide Polymorphism (SNP) nuclear markers have been detected that allow quantification of the level of introgression between some *F. silvestris* subspecies (Driscoll *et al.* 2011; Oliveira *et al.* 2015) no research has, as yet, been conducted to utilise the nuclear genome to understand the full suite of demographic events that occurred during domestication of *F. s. catus*. Evidence of cat domestication exists from China at *c.* 5.5 kya and Egypt *c.* 4 kya (Hu *et al.* 2014) but whether these represent independent events or, and perhaps more likely, are interwoven with a more complex chain of domestication followed by migration and multiple occurrences of admixture with wild progenitors remains unknown. The domestic cat has subsequently spread throughout the world, primarily as a consequence of human-mediated dispersal along major terrestrial and oceanic trade routes, and is now so widespread that it is found in all continents (including islands) bar Antarctica (Driscoll *et al.* 2011; Zeder 2012). While little is known about the fine-scale timing and pattern of this process, the increase of domestic cats in Asia and Europe around the 10th century BC is largely attributed to dispersal by Romans, Etruscans and Greeks. The current global population is now estimated at *c.* 1.2 billion individuals living both freely and in household association with humans (Driscoll *et al.* 2011).

Felis silvestris catus is now geographically ubiquitous and while it possesses many phenotypic traits (such as coat colour variation) and behavioural characteristics (such as tameness) not found in pure-bred wildcats, it has retained similar skeletal and muscular morphology and many behavioural traits of its wild progenitors (such as hunting ability). As such they can freely interbreed with other subspecies in the *F. silvestris* subspecies complex. This has led to the existence of a number of mixed ancestry 'types' in the cat population with overlapping and/or sympatric niches; true wildcats, random-bred feral cats, random-bred house cats, fancy breeds and hybrid wild subspecies (Matucchi 2014). Furthermore, morphological integration as a result

of hybridisation has created a 'morphological cline' where characteristics such as size and pelt patterns follow a continuum from purebred wildcats to the domestic house cat (Davis & Gray 2010; Matucchi 2014). Identification via morphology is thus very difficult, particularly in zooarchaeological material. Admixture has also had the effect of producing continual horizontal gene flow within the global *F. silvestris* population and may represent a threat to the genetic integrity of wild living pure-bred populations such as the European wildcat *F. s. silvestris*. The extent of hybridization between the European wildcat and the domestic cat varies greatly across remnant populations in central and southern Europe. Microsatellite markers suggest the existence of around 5-10 percent cryptic hybrids in most European populations (Mattucci *et al.* 2016). However, the Hungarian and remnant Scottish populations show deep signatures of genetic introgression with *F. s. catus* (Oliveria *et al.* 2015; Mattucci *et al.* 2016).

6.1.4 Conservation and the history of the European wildcat in Britain

The earliest evidence of the wildcat in Britain dates back to the Cromerian (*c.* 866,000 to 478,000 years ago) where a lower forth premolar was identified at West Runton as belonging to *F. s. lunensis*, the putative ancestral wildcat (Kurtén 1965). Records also exist for ancestral wildcat types at Swancombe, Kent during the Hoxnian Interglacial (*c.* 424,000 – 374,000 years ago; Sutcliffe 1964) and the Ipswichian (or Eemian) Interglacial at Joint Mitnor and Tornewton, Devon (*c.* 130,000 – 155,000 years ago; Sutcliffe 1960; Sutcliffe & Zeuner 1962). Lateglacial records are extremely scarce in the UK although the wildcat is recorded at Langwith Cave (Yalden 1999) and putative Lateglacial material also exists from Gough's Cave and Sun Hole Cave in the SW of England (Collcutt *et al.* 1981) - although none of this material has been dated. Holocene records are more numerous but undated material suffers from difficulties in correct identification due to potential chronological overlap with the morphologically similar *F. s. catus*. Holocene finds are recorded from pre-and sub-Boreal Thatcham, Gales Dale, Robins Hood and Steetly Caves (Yalden 1999), Windmill Hill and Glastonbury (Jope & Grigson 1965). In Scotland, undated Holocene records are available from Creag nan Uamh Caves, Loch Borralie in Lairg and Dunagoil Cave on the Isle of Bute (Lawson 1981; Richie 1920; Kitchener & O'Connor 2010).

Interpretations made from the zoo-archaeological record for this species should be made with caution, particularly inferences made from the apparent absence of wildcat remains (a rare species anyway) at particular time periods or sites. The subfossil record for European wildcats in Britain is scarce and incomplete. Kitchener & O'Connor (2010) attribute this to its status as a top predator and the fact that this species was not exploited by humans to the extent that allowed large aggregations of skeletal material to accumulate at sites. European ^{14}C radiocarbon dated records indicate that Lateglacial recolonisation of central Europe, from where the wildcat could have entered Britain, had been completed by at least the Allerød (Feustal 1980; Hedges *et al.* 1998) and that the wildcat persisted here into the Younger Dryas (Degerbøl 1933; Altuna *et al.* 1991; Staesche 1994). However, little can be inferred about Lateglacial recolonisation of Britain from the scant fossil record available. Radiocarbon dates from the putative Lateglacial material at Sun Hole Cave and Goughs Cave could provide information on whether the wildcat had recolonised SW England during this period and by which date. The gap of 5000 years between the pre-and sub-Boreal records of *F. s. silvestris* has been interpreted by Sommer & Benecke (2005) as evidence of a slow Postglacial recolonisation of Britain by the wildcat. However, further evidence is required to fully support this assertion, since remains of this animals are generally uncommon.

The history and fate of the wildcat in Britain is intricately linked with human activity, namely changes in land-use, game hunting and the introduction of *F. s. catus*. The transition to woodland during the Early Holocene and subsequent spread of woodland ecotypes throughout Britain during this period would have provided abundant habitat for wildcats. Despite the paucity of their faunal record here, there is no reason to doubt that they were once widespread and relatively abundant throughout this region, being absent only from exposed mountains, coastal areas and open habitats (Easterbee *et al.* 1991). Archaeological evidence also suggests they may have been present in the Inner Hebridean Islands, Shetland and Orkney. However, considerable uncertainty exists over whether these were wildcats that naturally colonised these islands, wildcats that were imported by humans for exploitation,

domestic cats, hybrid forms or combinations thereof (Sommer & Benecke 2006; Kitchener & O'Connor 2010).

The transition from forested to cleared land (termed 'culture steppe' by some authors; Hammond 1974) over the last 5000 years of the Holocene greatly reduced the availability of habitat for woodland fauna (Buckland & Dinnin 1993; Whitehouse 2006). However, it is not until relatively recently that wildcat populations experienced catastrophic declines. By the early 18th century, the wildcat still had a stronghold in the Scottish Highlands but in England, its range had been reduced to Northern areas only and it was found only sparingly in Wales (Kitchener & O'Conner 2010). The rise of sporting estates and a concordant increase in persecution coupled with continued deforestation finally resulted in the complete extirpation of the wildcat from England and Wales by the end of the 19th century. It is now found only in areas of Scotland north of the Central Belt with recent population surveys estimating the population at only *c.* 400 individuals (Kitchener *et al.* 2005). *Felis s. catus* is known to have been introduced to Britain in large numbers by the Romans (Kitchener & O'Connor 2010) but there is archaeological evidence (although somewhat ambiguous) of *F. s. catus* that pre-dates this. Harcourt (1979) notes the presence of articulated cat bones at the Iron Age human settlement of Gussage All Saints in Dorset, which he attributes to the domestic cat, and Grant (1984) also interprets cat remains from an Iron Age hill-fort at Danebury as domestic cat. The extent of the early interactions between *F. s. silvestris* and *F. s. catus* in Britain, including the extent and origin of hybridisation between these two subspecies, remains unknown.

The European wildcat is a legally protected species throughout Europe due to historic and ongoing population declines and hybridization with *F. s. catus* (Yamaguchi *et al.* 2015). It is included in Annex IV of the European Habitats Directive (92/43/CEE) for strictly protected species, in Annex II of the Bern Convention and listed as 'Vulnerable' by the IUCN (Yamaguchi *et al.* 2015). The European Council identified two main extinction risks for *F. s. silvestris*, *i*) habitat loss and fragmentation and *ii*) admixture with the domestic cat (Stahl & Artois 1991). Fragmentation and destruction of habitat negatively affects both population level processes and the evolutionary

potential of species. Small and isolated populations with low levels of gene flow typically experience loss of both neutral and adaptive genetic diversity, reduced effective population size and elevated rates of genetic drift (Hedrick 2011). This may lead to increased levels of inbreeding via reduced mating opportunities, fixation of deleterious mutations and a loss of genetic diversity. The overall loss of adaptive genetic potential leaves populations vulnerable to the effects of inbreeding depression, novel pathogens and stochastic environmental events, effectively increasing population extinction risk (Hedrick 2011).

The effects of hybridization between wild and domestic forms are largely considered to be negative. These can include loss of reproductive potential, a decrease in fitness at both the individual and population level, introduction of maladaptive domestic alleles into wild populations, disease transfer and loss of legal protection if the taxonomic status of the wild population is altered (Leonard *et al.* 2014). Driscoll *et al.* (2007) have suggested that the introgression of domestic cat alleles into European wildcat populations poses a number of threats to autochthonous wildcat gene complexes. As hybrid recombinant genotypes resulting from *F. s. silvestris* – *F. s. catus* admixture have not had time for the evolutionary process of natural selection to act on deleterious genes and mutations, hybrid genotypes may be expected to be less fit than either parental type (Turelli *et al.* 2001). Furthermore, as the landscape within which wildcats exist becomes increasingly human-dominated, there may be a selective advantage in domestic cat traits within hybrids. These may decrease suitability to the natural environment and increase fitness in anthropogenically-modified environments (*i.e.,* less fearful behaviour towards humans and greater exploitation of human generated resources; Anderson *et al.* 2009; Macdonald *et al.* 2010; Driscoll *et al.* 2011). No studies have yet attempted to quantify the effects of *domestic × wildcat* hybridization on the fitness and behaviour of wild-living populations so these hypothesised threats remain largely speculative and untested.

In Scotland, the wildcat is protected under the Wildlife & Countryside Act 1981 and has its own Species Action Plan (Davis & Gray 2010). Conservation of wildcats in Scotland is deeply challenging. While persecution and habitat availability remain of

concern, arguably the primary obstacle to wildcat conservation is genetic introgression with the domestic cat (Senn & Ogden 2015). Deep levels of admixture exist in the Scottish wildcat population and genetic research indicates that there is a complete genetic continuum between pure bred wildcats and domestic cats in the wild-living wildcat population (Senn & Ogden 2015). This has led to the description of the wildcat population as a 'hybrid swarm', where the majority of individuals show some level of hybridisation (Senn & Ogden 2015). This deep level of admixture means that simply defining a wildcat is problematic. The first wildcat type specimen was collected in 1904, at least 2000 years after domestic cats are known to have been present in Britain (Senn & Ogden 2015) and thus may have admixed ancestry. No wildcat soft tissue or pelt specimens exist that pre-date the introduction of the domestic cat and so wildcat specimens that have been used as baseline reference material may themselves have experienced some introgression with domesticates. Identifications are generally carried out using pelt characteristics and a scoring system developed via analyses of museum specimens and deceased cats (Kitchener & O'Connor 2010). However, a high degree of error has been observed among experts in identifying purebred wildcats from hybrids on the basis of pelage characteristics (Devillard 2014). Pelt characters often do not show congruence with genetic data on levels of admixture. This likely due to *i*) the high extent of admixture in the Scottish wildcat population and *ii*) lack of correspondence and/or different genomic histories between the regions of the genome being examined for introgression and the few genes in control of phenotypic traits.

Scottish Natural Heritage (SNH) now has a captive breeding program, which aims eventually to supplement the wild population of *F. s. silvestris,* in addition to carrying out other ongoing conservation measures such as control of feral cats and habitat management (David & Gray 2010; Senn & Ogden 2015). However, the difficulties involved in identifying the level of admixture in wildcats has led to challenging and sometimes controversial decisions over how to best conserve the wildcat in Scotland. While preservation of pure-bred wildcats may be a viable conservation option in mainland Europe the deep level of introgression in the Scottish population means that this is not, in all likelihood, a realistic option. SNH now proposes a 'cut-off' point

for selecting which wildcats to conserve, which relies on corroborating pelage and genetic evidence (Senn & Ogden 2015). Using SNP data, a cat will be considered a 'wildcat' if there is 95 percent confidence of it being closer than a first generation backcross to a purebred wildcat, *i.e* an equivalent genetic contribution of one of four grandparents being a domestic cat and the remaining three pure wildcats.

6.1.5 Aims and research questions

Ancient DNA has clear utility in unravelling the genetic history of the wildcat in Britain. It has huge potential to uncover both the effects of abrupt climate change on wildcat populations over the Late Pleistocene - Holocene boundary and to provide genetic data from individuals that pre-date the introduction of the domestic cat thus aiding conservation research for this critically endangered felid. The sample sizes in this study are prohibitively small and do not allow many key research questions to be addressed. However, this is the first research undertaken on the European wildcat using aDNA and therefore kick-starts the field of aDNA research into *Felis silvestris silvestris*. Specific aims are to:

- Generate the first whole ancient mitogenomes and nuclear genes from British wildcats of Lateglacial to Late Holocene age
- Generate the first draft low-coverage whole genome of a putative wildcat dating to the Romano-British period
- Quantify and assess levels of genetic diversity in the mitogenome of ancient wildcats and infer from this information relating to early founder populations
- Use ancient mtDNA to validate the current mtDNA admixture test employed by Scottish Natural Heritage
- Use AMS ^{14}C radiocarbon dating to determine the age of a wildcat specimen from Gough's Cave and two Holocene specimens from Creswell Crags
- Use mtDNA and AMS ^{14}C evidence to determine whether wildcats show population continuity over the Younger Dryas cold phase in Britain

6.2 Methods

6.2.1 Sampling strategy

Obtaining samples of subfossil ancient wildcat from Britain is confounded by the rarity of this species in the British archaeological record and potential chronological overlap of this subspecies with domesticates during the Holocene (Kitchener & O'Connor 2010). The high degree of morphological similarity between skeletal elements of the European wildcat, *F. s. silvestris,* and the domestic cat *F. s. catus* makes separation of these species from excavated material extremely difficult. In the absence of a large program of radiocarbon dating (beyond the financial scope of this study), sampling was limited to material that could (as far as possible) be associated with stratigraphic layers that definitively date to periods prior to the introduction of *F. s. catus*. Genetic methods exist with which to ascertain both subspecific taxonomy within the *F. silvestris* subspecies complex and determine the presence and degree of introgression between European wildcats and the domestic cat (Driscoll *et al.* 2007; Driscoll *et al.* 2011; Oliveira *et al.* 2015). This could, in theory, provide a means to identify pure-bred ancient wildcats from archaeological material that cannot be directly dated due to *i*) financial constraints *ii*) prohibitively small amounts of available material and/or *iii*) lack of stratigraphic information. However, this method would require screening extremely large amounts of individuals and the ability to retrieve both mitochondrial DNA and nuclear SNPs. As a result, this was not considered a feasible approach in this study due to time and financial constraints. The sample size in this dataset was therefore low, consisting of only nine bone samples, eight of which were from the Holocene period and one of which was putatively assigned to the Lateglacial period (Table 6.1; Fig 6.4). While sample sizes were small, they did provide an excellent spatio-temporal spread from the Lateglacial site of Gough's Cave in the SW of England to a variety of Holocene sites throughout the Midlands, northern England and up to Loch Borralie in the far north of Scotland (Table 6.1; Fig 6.4). Attempts were made to locate all Lateglacial wildcat material recorded in the UK. However, the Sun Hole Cave wildcat remains appear lost or potentially destroyed during World War II (Linda Wilson, Pers comms.) and the Langwith Cave material could not be identified with any degree of certainty.

6.2.2 AMS ^{14}C Radiocarbon dating and age determination

Financial support and sampling permission could only be obtained to radiocarbon date four specimens from this dataset. These were MM098, (Gough's Cave, Cheddar); MM092, Robin Hood Cave; MM093, Pin Hole Cave (both Creswell Crags, Derbyshire) and MM091, Dunagoil Cave, Isle of Bute. Samples ranged from 0.2g to 0.5g in weight and all samples were composed of bone collagen. Specimen MM053 (Gully Cave, Ebbor Gorge) had a pre-existing radiocarbon date. Samples MM076 and MM077 (Victoria Cave and North End Pot, Derbyshire) remain undated but have been excavated from sites that are unlikely to have post-Iron Age deposits. Contextual information would confer on these specimens a likely date from the Late-Mid Holocene. However, without direct dating information this remains only a best estimate.

Figure 6.4. Distribution of sites from where British *Felis silvrestis silvestris* material was sampled.

Table 6.1. Sample identification numbers, collections and sites for all British wildcats sampled in this study.

ID	Collection	Collection ID	Site
MM036	Somerset County Council Museum, Taunton	41/1995/76	Bleadon Bone Cave, Somerset
MM053	Royal Holloway University of London	Unreg	Gully Cave, Somerset
MM076	Private Collection, Tom Lord	Unreg	Victoria Cave, Yorkshire
MM077	Private Collection, Tom Lord	Unreg	North End Pot, Yorkshire
MM090	National Museum of Scotland	Unreg	Loch Borralie, Sutherland
MM091	National Museum of Scotland	Unreg	Dunagoil Cave, Isle of Bute
MM092	Manchester Museum	LL.11993	Pin Hole Cave, Derbyshire
MM093	Manchester Museum	P3247	Robin Hood Cave, Derbyshire
MM098	Natural History Museum, London	M13798	Gough's Cave, Somerset

6.2.3. Ancient DNA

Screening for endogenous content

For each sample between 10-50mg of powdered bone was used as the starting material for DNA extractions and library builds as described in Appendix A, Protocols A1 (Ancient DNA Extraction) and A2 (DNA Library Build). These were shotgun sequenced on an Illumina NextSeq500 and paired-end reads were paired, trimmed, merged and aligned to the domestic cat genome using the Burrows Wheeler Aligner (BWA) with parameters as described in Chapter 2, Section 2.3.2. Sequence data was mapped to an unmasked version of the *Felis catus* [1]genome v8.0 (Li *et al.* 2016; accession numbers NC_018723.2 – 41.2). Statistics for all mapping assemblies were calculated with the SAMtools and BEDtools suites (Li *et al.* 2009) in a Linux-GNU environment. Mapping paired-end Illumina reads to the domestic cat nuclear genome enabled quantification of the percentage of endogenous (target species) reads that had been sequenced amongst the complex background of contaminant reads. Only samples that had > 0.5 percent endogenous content were taken through to downstream analyses.

[1] The authors of this mitochondrial genome, Li *et al.* (2016), ascribe full species status to the domestic cat. However, no clarity on the reasons behind this taxonomic preference is provided in the manuscript.

Authentication of ancient DNA

To check for a signature of ancient DNA degradation and thus ensure that no modern contaminant DNA had been mapped to the *F. s. silvestris* or *F. s. catus* reference genomes this study used mapDamage2.0 (Jónsson *et al.* 2013). Degradation of ancient DNA introduces specific nucleotide mis-incorporations and DNA fragmentation signatures in sequencing reads and these signatures of post-mortem damage can be used to authenticate ancient DNA. Typical patterns are an excess of cytosine to thymine (C – T) misincorporations at the 5' end, guanine to adenine (G – A) misincorporations at the 3' termini (resulting from cytosine deamination in single stranded 5' overhangs), and an excess of purines at the genomic locus just before sequencing start). The mapDamage2.0 algorithm uses read alignment of reference genomes and maximum likelihood optimization under the HKY transition matrix to deliver quantitative estimates of damage parameters under the assumption that post-mortem DNA damage and mutations within a fragment are independent (for full details of the mapDamage algorithm see Jónsson *et al.* 2013).

Nuclear insertions of mitochondrial DNA (NUMTs)

The cat genome is known to harbour a wide variety of nuclear insertions of mitochondrial DNA (NUMTs; Lopez *et al.* 1996; Antunes *et al.* 2007). The first of these was detected by Lopez *et al.* (1996) where a translocation of a 7.9kbp tandem repeat of mtDNA was discovered on chromosome D2. The subsequent sequencing of the domestic cat genome has allowed a greater depth of detection and characterisation of NUMTs and it is now known that at least 176kb of the cat genome, spanning the majority of chromosomes, are mtDNA translocations (Tamazian *et al.* 2014). These homologues have the potential to confound phylogenetic and population genetics studies if they are amplified and incorporated into the final consensus sequence. However, Li *et al.* (2012), found that in-solution capture enrichment followed by NGS sequencing does not generally incorporate large numbers of NUMT alleles and that those that are amplified generally have a frequency of <5 percent. Target enrichment usually provides high coverage sequence data and therefore low frequencies of NUMT alleles are unlikely to influence the final

consensus sequence when this method is used (Li *et al.* 2012). While there are difficulties in differentiating low frequency mtNUMTs from true low-level heteroplasmic mutations, the presence of low-frequency mtNUMTs in the data generated in this study was not considered to have a realistic potential of affecting base calls.

In-Solution target enrichment

Biotinylated RNA baits were synthesised from the *F. catus* mitogenome (accession NC_001700.1) and five functional nuclear protein-coding genes: MC1R (involved in coat colour determination; 954bp), COL9A (associated with hearing; 17,555bp), SIRPB2 (inter-photoreceptor retinoid binding protein; 11,405bp), APOBR (apolipoprotein binder; 6,559.2) and LY6K (lymphocyte antigen 6; 5,913bp). Nuclear genes were chosen as they had typically been used in previous phylogenetic studies and were considered potentially informative for gene evolution between wild and domestic cats. In-solution target enrichment was carried out as in Appendix A, Protocol A4 (In-solution Hybridisation Capture Enrichment) before sequencing on the Illumina NextSeq500 platform. Paired-end Illumina reads were quality filtered, trimmed, merged and aligned using BWA to either a consensus sequence of the two available *Felis silvestris* mitogenomes (Li *et al.* 2015; accessions: NC_028310.1, KP_202278.1) or the reference sequence of the nuclear gene used to synthesise the baits. Mapping parameters followed those described in Chapter 2, Section 2.3.2.

Quantification of genetic introgression in the mitochondrial genome

For samples that potentially contained some level of genetic introgression with domestic cats, determination of subspecific rank was carried out via the suggested methods in Driscoll *et al.* (2011). This protocol uses mitochondrial DNA to assign specimens to one of the five mtDNA clades within the *F. silvestris* subspecies phylogeny: Clade I – European wildcat *F. s. silvestris*, Clade II Southern African wildcat *F. s. cafra*, Clade III Asian wildcat *F. s. ornata*, Clade IV Near Eastern wildcat *F. s. lybica* and domestic cat *F. s. catus*. Due to its matrilineal pattern of inheritance, mitochondrial DNA can only infer uniparental ancestry and so can only provide information on

introgression that has occurred throughout the maternal line. Within matrilineal limits, however, it is considered to be an informative marker at detecting historical introgression events and has been shown to reliably recover wildcat subspecific clades (Driscoll *et al.* 2007; Driscoll *et al.* 2011).

For samples MM076 and MM077, which were not directly dated and sample MM091 which returned a ^{14}C age that placed it in the Romano-British period (*i.e* after the introduction of the domestic cat – see section 6.3.1), 2,604bp length sequences of the NADH dehydrogenase 5 and 6 (ND5 and ND6) genes were aligned with previously published data for *F. s. silvestris*, *F. s. caffra*, *F. s. ornata*, *F. s. lybica* and *F. s. catus*. The final dataset contained 61 sequences in total and consisted of ten sequences from *F. s. caffra* (accession numbers EF587016-025), seven sequences from *F. s. ornata* (accession numbers EF587026-062), twenty-three sequences of *F. s. silvestris* (accession numbers EF587154-177), eleven sequences of *F. s. catus* (accession numbers KP279609, KP279615, KP279619-621, KP279623-628), one sequence of *F. s. lybica* (accession no KP202275) and two sequences of the Chinese desert wildcat, *F. s. beiti* (accession numbers EF587178-179). Four sequences of the sand cat *F. s. margarita* (accession numbers EF587033-036) were used to root the phylogeny and only unique haplotypes were included.

Sequences were aligned with data generated in this study for MM076, MM077 and MM091 using Muscle before the alignment was manually edited. Two tree-building methods, Neighbour-Joining (NJ) and Bayesian Markov Chain Monte Carlo (MCMC) analysis were used to assess the robustness of the tree topology. The appropriate model of nucleotide substitution was determined using JModelTest (Darriba *et al.* 2012). The Akaike, Akaike Corrected and Bayesian Information Criteria (AIC, AICc and BIC) as well as the Decision Theory method all estimated the General Time Reversal (GTR) model with a gamma (G) shape parameter of 0.022 to be the most likely model for this dataset. The NJ tree was constructed in MEGA v. 7.0.18 (Kumar *et al.* 2016) using the Kimura 2-parameter model, which includes transitions and transversions with G set to 0.022. The Bayesian phylogeny was created in MrBayes (Huelsenback *et al.* 2001) on the Cipres portal (Miller *et al.* 2010) using the GTR+G

parameters suggested by JModelTest (above). Two independent chains ran for $2x10^7$ iterations each, sampling trees and model parameters from the posterior every 1000 iterations. The first 10 percent of each run was discarded as burn-in. TRACER v.1.5 (Rambaut *et al.* 2014) and Are We There Yet (AWTY: Wilgenbusch *et al.* 2004) were used to check for chain convergence and sufficient sampling before a 50 percent consensus tree was created. All trees were edited in FigTree v.1.3 (Rambaut 2009).

Phylogenetic analyses

Available comparative data for wildcats is very scarce and the sequence data that is available represents a variety of incompatible mtDNA and nuclear markers. Furthermore, most of these datasets do not span the full European range of *F. s. silvestris* due to geographically restricted sampling. In terms of mitochondrial DNA, only two whole mitochondrial genomes exist and the individuals that these were taken from have not been examined for admixture (Li *et al.* 2015). There are a number of mtDNA sequences from the ND5/ND6 genes but those representing non-admixed wildcats are located in France and Spain only and this data is not informative enough to recover the known split between these regions inferred by microsatellites in Matucci *et al.* (2016); this study, not shown. A number of partial CR sequences are available from purebred wildcats in Germany and Luxembourg (Steyer *et al* 2016). These individuals have been tested for levels of introgression via microsatellite and Bayesian STRUCTURE simulations so are known to be true wildcats. Despite the relatively small geographic range spanned by these samples (see Fig 6.3), this area shows two distinct microsatellite clusters that correspond to two of the biogeographic lineages suggested by Mattucchi *et al.* (2016): a western cluster with samples from west Germany and Luxembourg (corresponding to the Mattucchi *et al.* central Europe lineage which also incorporates France and Belgium) and a central Germany cluster (corresponding to the lineage of the same name in Mattucchi *et al.* 2016; Steyer *et al.* 2016). This offers the opportunity to reconstruct the phylogeographic relationship of British wildcats to the central European and central German clades.

Accordingly, a 433bp sequence of the CR was extracted from British wildcat mitogenomic sequences generated here and aligned with 14 CR sequences

representing the central Europe lineage (accession numbers: GQ268268-70, GQ268287, GQ268279-98, GQ268309-16) and 20 sequences representing the central German lineages (accession numbers: GQ268271-84, GQ268290-95, GQ268299-304, GQ268307-308). The Chinese desert cat (*F. s. bieti*, accession number KP_202273) was used as the outgroup taxon. The appropriate model of nucleotide substitution was determined in JModelTEST (Darriba *et al.* 2012) to be the Hasegawa, Kishino and Yano (HKY) model. A Bayesian MCMC phylogenetic reconstruction was run on MrBayes v3.2.6 (Huelsenbeck *et al.* 2001) with two independent chains running for 4 x 10^6 iterations each, sampling trees and model parameters from the posterior every 4000 generations. The first 10 percent of each run was discarded as burn-in. TRACER v.1.5 (Rambaut *et al.* 2014) and Are We There Yet (AWTY: Wilgenbusch *et al.* 2004) were used to check for chain convergence and sufficient sampling before a 50 percent consensus tree was created and edited in FigTree v.1.3 (Rambaut & Drummond 2010).

To examine relationships within the British data only, a MJ network was calculated in PopART (http://popart.otago.ac.nz) with (ε) set to zero. The dataset contained 6 taxa and was performed using alignments of the whole mitogenome and single gene mtDNA datasets where haplotype diversity was sufficient to infer some degree of distinction. Two methods were employed to investigate patterns of population differentiation, the fixation index (F_{ST}). Within population sizes were too small to conduct a more appropriate test such as an AMOVA. In addition, Tajima's D, Fu's F (Fu 1997) and a mismatch distribution were used to search for evidence of population expansions (see chapter 5 for a description of these statistical tests). Data were pooled for analysis as sample sizes in putative chronological populations were too small.

Genetic diversity and fixed nucleotide discrimination of wild and domestic cats

Mitochondrial genomes generated here were annotated via the MITOS web server (Bernt *et al.* 2013). Genetic diversity indices for the mitogenome as a whole and for major coding and non-coding genes were estimated between British individuals both using pairwise genetic distances in net percent and with the Kimura-2 metric implemented in DNAsp (Rozas 2010). As part of a genetic screening test

implemented by SNH to identify domestic cat introgression in wildcat mtDNA fixed nucleotide positions identified by Driscoll *et al.* (2007) were used to distinguish between the two subspecies (McEwing *et al.* 2012). These fixed nucleotide positions are located over 2,604bp of the ND5/ND6 genes and were identified using extant *putative* wildcats as a reference (Driscoll *et al.* 2007). As such these individuals may have experienced historical admixture with domestic or hybrid feral cats. To add another line of evidence in the power of this test to distinguish domestic and wildcats in Scotland, all ancient wildcats that yielded mitogenomes in this study were examined for the presence of these fixed base mutations. These were at locus positions (defined on the *F. catus* mitogenome) as: 12663, 13265, 13326, 13554, 13611, 13656, 13679, 13770, 14187, 14555, 15082 and 15164.

Whole genome sequencing of sample MM091

Endogenous content in the first screening run for sample MM091 from Dunagoil Cave, Isle of Bute was surprisingly high (>50 percent). This presented an opportunity to assemble the first, draft, light-coverage genome from an ancient European wildcat via shotgun sequencing. A total of six DNA libraries on two separate DNA extracts were built using the protocol outlined in Appendix A, Protocol A2 (DNA Library Build). Of these, one library was sequenced on the Illumina NextSeq500 platform alongside all other wildcat samples on the screening run, one was sequenced in isolation to optimize the number of reads returned and four were sequenced together without any other samples. Illumina paired-end reads from all runs were then concatenated with those returned from the *in-situ* hybridisation procedure. Concatenated reads were then paired, trimmed, merged and assembled to the domestic cat *F. catus* genome v8.0 using the assembly parameters described in Chapter 2, Section 2.3.2.

The effective 'mappable' region of a genome can be difficult to define as it is influenced not only by the characteristics of the reference sequence but also by sequence read properties, the mapping algorithm used and the parameters used in the read assembly (*i.e* number of mismatches allowed). A distinction is made between *uniquely mapping reads* which can only be aligned to a single genomic location and

multiple mapping reads where read properties (DNA sequence, quality scores, error rates etc.) mean it can be aligned to two or more genomic locations (Whiteford *et al.* 2005). For a reference genome, the proportion of reads that can be thought of as uniquely mapping depends on the length of the sequence reads produced and the number of mismatches allowed (Derrien *et al.* 2012). It is therefore possible to determine the mappability of a genome if the technical details of the sequencing process are known. Genome mappability may be determined as 'the inverse of the number of times that a read originating from any position in the reference genome maps to the genome itself' (Derrien *et al.* 2005, pp.1). Regions of the genome that have high mappability will produce unique, unambiguous reads while regions with low mappability will produce ambiguous, multiple mapping reads.

The *F. s. catus* nuclear genome consists of 18 autosomes (A1-3. B1-4, C1-2, D1-4, E1-3, F1-2) in addition to the sex chromosomes and the latest version Fca-8.0 is 2.56Gb in length. Repetitive elements are estimated to make up around 55.7 percent of the genome and are found on every chromosome (Tamazian *et al.* 2014). The majority of these are long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), satellite DNA, and long terminal repeats (LTRs). These regions will tend to have low mappability and, as such a large proportion of the *F. catus* genome is composed of them, it should be kept in mind that these will significantly influence final coverage estimates. Coverage values reported here should therefore be viewed as a minimum estimate.

6.3 Results

6.3.1 ^{14}C radiocarbon dating

Four samples which were submitted to ORAU for radiocarbon dating returned successful dates (Table 6.2, Fig 6.5; full chemical results are in Appendix A, Fig AF1, Table AT1). Samples MM053 (Gully Cave) and MM092 (Pin Hole Cave) both dated to the Early Holocene (10, 227 – 9, 921 cal BP and 8,715 – 8,484 cal BP respectively) and sample MM093 (Robin Hood Cave) dated to the Mid Holocene (7,239 – 6,999 cal BP; climatic phases of the Holocene defined as in Walker *et al.* 2012). Sample

MM091 from Dunagoil Cave, Isle of Bute returned a young (Late Holocene) date of 1,814 – 1,618 cal BP, which places it in the Romano-British period, and thus post-dates the known introduction of the domestic cat into Britain. The sole sample which was assumed of Lateglacial age, sample MM098 from Gough's Cave, returned an Early Holocene age of 10,655 – 10, 299 cal BP. This represents the oldest directly dated *F. s. silvestris* specimen from Britain but meant that this study was unable to address the question of whether wildcats persisted in Britain over the Younger Dryas using ^{14}C dating methods.

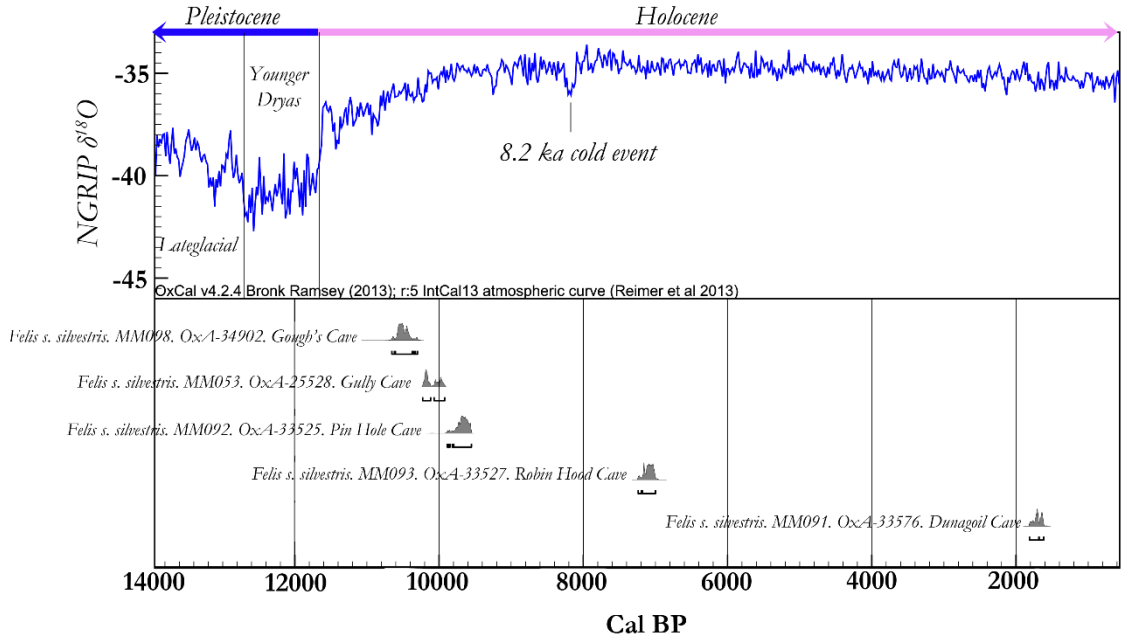


Figure 6.5. Accelerator Mass Spectrometry (AMS) ^{14}C radiocarbon ages for four European wildcat *Felis silvestris silvestris* individuals. Also shown is the age estimate for MM053, which was dated as part of a previous study.

Table 6.2. Accelerator Mass Spectrometry (AMS) ^{14}C radiocarbon ages for four European wildcat *Felis silvestris silvestris* individuals. Also shown is the age estimate for MM053, which was dated as part of a previous study.

Sample ID	Locality	Age cal BP (95.4% probability)	Period
MM098	Gough's Cave, Somerset	10,655 - 10,299	Early Holocene
MM053	Gully Cave, Somerset	10,227 - 9,921	Early Holocene
MM092	Pin Hole Cave, Derbyshire	8,715-8,484	Early Holocene
MM093	Robin Hood Cave, Derbyshire	7,239-6,999	Mid Holocene
MM091	Dunagoil Cave, Isle of Bute	1,814-1,618	Late Holocene

6.3.2 Screening and DNA authentication

Initial screening of wildcats (prior to target enrichment) via read alignment to the domestic cat *F. s. catus* genome v8.0 gave all samples bar MM090 and MM091 an end target DNA content of between 1 percent - 8 percent (not shown). Sample MM090 had a very poor endogenous DNA content and a low number of mapped reads and thus was not taken forward to the target enrichment procedure. MM091 had a very high percent of endogenous reads (> 50 percent). Mapping statistics for pooled shotgun and target enrichment reads for all wildcats are shown in Table 6.3. The DNA degradation profiles output by mapDamage2.0 (Jónsson *et al.* 2011) indicated that all specimens showed diagnostic signatures of DNA damage associated with ancient DNA (Appendix E, Fig EF1, a-f.). These included an excess of cytosine to thymine changes at the 5' end and an excess of guanine to adenine changes at the 3' end. Levels of DNA damage are influenced by a multitude of factors and are not necessarily indicative of the age of a specimen. However, there appears to be rough correlation in these specimens with the oldest wildcat, MM053 (10,227 – 9,921 cal BP) showing a highly damaged profile while the youngest specimens, MM091 (1,814 – 1,618 cal BP) showed lower levels of DNA damage compared to the other specimens.

Table 6.3. Mapping results for British *F. s. silvestris* specimens aligned to the domestic cat *Felis catus* genome v8.0. These results are for pooled reads generated from shotgun sequencing and target enrichment (bar MM090 which was not subjected to target enrichment due to low endogenous content). The inclusion of reads derived from target enrichment leads to the elevated levels of duplicate removal. The final end percent is affected by the high number of duplicates; a more informative parameter in this case is the total number of reads mapping to *F. catus* v8.0 after quality filtering and duplicate removal.

SampleID	TotalReads	Readsmapped	Readsmapped after quality (q30) filter	Readsmapped after quality filter (q30) and duplicate reads removed	duplication (%)	Final End (%)
MM036	16704809	132578	46105	16055	65.18	0.1
MM053	22622511	161259	96792	46610	51.85	0.2
MM076	20474136	206966	119358	31902	73.27	0.2
MM077	13477198	123955	86662	68792	20.62	0.5
MM090	6604405	2505	1024	803	21.00	0.0
MM091	230428351	152519568	122858149	67994993	44.66	29.5
MM092	11272358	1998690	1315696	501077	61.92	4.4
MM093	12432278	1059051	731372	458238	37.35	3.7
MM098	4788163	356961	292290	255306	12.65	5.3

6.3.3 Nuclear DNA

For all the specimens apart from MM091 nuclear genes were not able to be sequenced with sufficient fraction of the reference gene covered and of sufficient read depth at covered regions (Table 6.4). The Dunagoil cave specimen (MM091) had reasonable coverage with at least 80 percent of the total length of all genes covered at least 1 read depth but for all other specimens, no more than a third of the gene had mapped reads. This is likely due to the low copy number of nuclear DNA present in ancient bone material in conjunction with an enrichment protocol that used RNA baits diluted 1:10. Therefore nuclear DNA was not able to be incorporated into any analyses in this study.

Table 6.4. Fraction (%) of *Felis silvestris silvestris* nuclear genes covered at 1 read depth or more.

ID	COL9A (17,555bp)	SIRPB2 (11,405bp)	MC1R (954bp)	APOBR (6,559bp)	LY6K (5,913bp)
MM036	0	1.16	4.92	0	0
MM053	5.57	5.39	9.11	2.63	5.46
MM076	4.83	0	0	1.29	1.87
MM077	13.92	18.42	10.79	21.07	14.54
MM091	88.52	92.16	97.27	80.95	86.6
MM092	9.86	20.13	0	8.91	16.99
MM093	29.42	28.34	47.39	28.99	32.94

6.3.4 Mitochondrial DNA

Whole mitogenomes were successfully returned for six out of the eight specimens subjected to the target enrichment procedure and x-fold coverage values ranged from 9.6x in MM076 to 180x in MM091. These represent the first ancient whole mitochondrial genomes generated from the European wildcat. Very low coverage data was generated for samples MM036 (Bleadon Bone Cave, Somerset) and MM098 (Gough's Cave, Somerset) so these datasets could not be included in the analyses. Unfortunately, as no material used in this study had a late Pleistocene age, the remainder of the analyses had to focus exclusively on material from the British

Holocene. Whole mitochondrial genomes were sequenced and assembled for MM053, Gully Cave, Somerset, 26.7x; MM076, Victoria Cave, Yorkshire, 9.6x; MM077, North End Pot, Yorkshire, 67.2x; MM091, Dunagoil Cave, Isle of Bute, 180x; MM092, Pin Hole Cave Derbyshire, 60.3x; MM093, Robin Hood Cave, Derbyshire, 98.3x; Table 6.5; Fig 6.6 a-f). Two insertions were detected in the CR, which all specimens possessed and which are not present in the modern reference wildcat mitogenome, a thymine (T) between base positions 63 and 64 in the reference sequence and an adenine (A) between bases 16,198 and 16,199. This gave a total length for the ancient mitogenomes as 16,642bp. The domestic cat *F. s. catus* genome is 17,009bp in length, a difference of 367 bases. Whole wildcat mitogenome alignments to the domestic cat *F. s. catus* mitogenome showed this to be due to a shorter CR length in the European wildcat. The domestic cat has a long region of repeat sequences largely composed of minisatellite-like 'CTACACs' inserted in positions 264 – 474 and a 160 base insertion (non-repetitive) between bases 16,608 and 16,767 (reference positions from the *F. catus* mitogenome accession number NC_001700). Both of these insertions are located in the hypervariable regions of the CR. Examination of the data showed no evidence of heteroplasmy.

Table 6.5. Read mapping results for *Felis silvestris silvestris* against a modern European wildcat cat mitochondrial genome (length 16,640bp). A high proportion of duplicates is the product of the target enrichment process.

Sample ID	Total reads	Reads mapped after quality filter	Quality filtered reads mapped after duplicate removal	% duplicate reads removed	% reference sequence covered at min 1 read depth	X Coverage
MM053	22,622,511	103,874	6,261	94	99.83	26.7
MM076	20,474,136	158,541	2,764	98	98.62	9.6
MM077	24,425,557	138,146	15,769	88	99.98	67.2
MM091	230,428,351	1,360,363	32,538	97	100	180
MM092	16,043,260	1,841,484	14,538	99	99.84	60.3
MM093	26,200,314	1,331,630	21,882	98	99.99	98.3

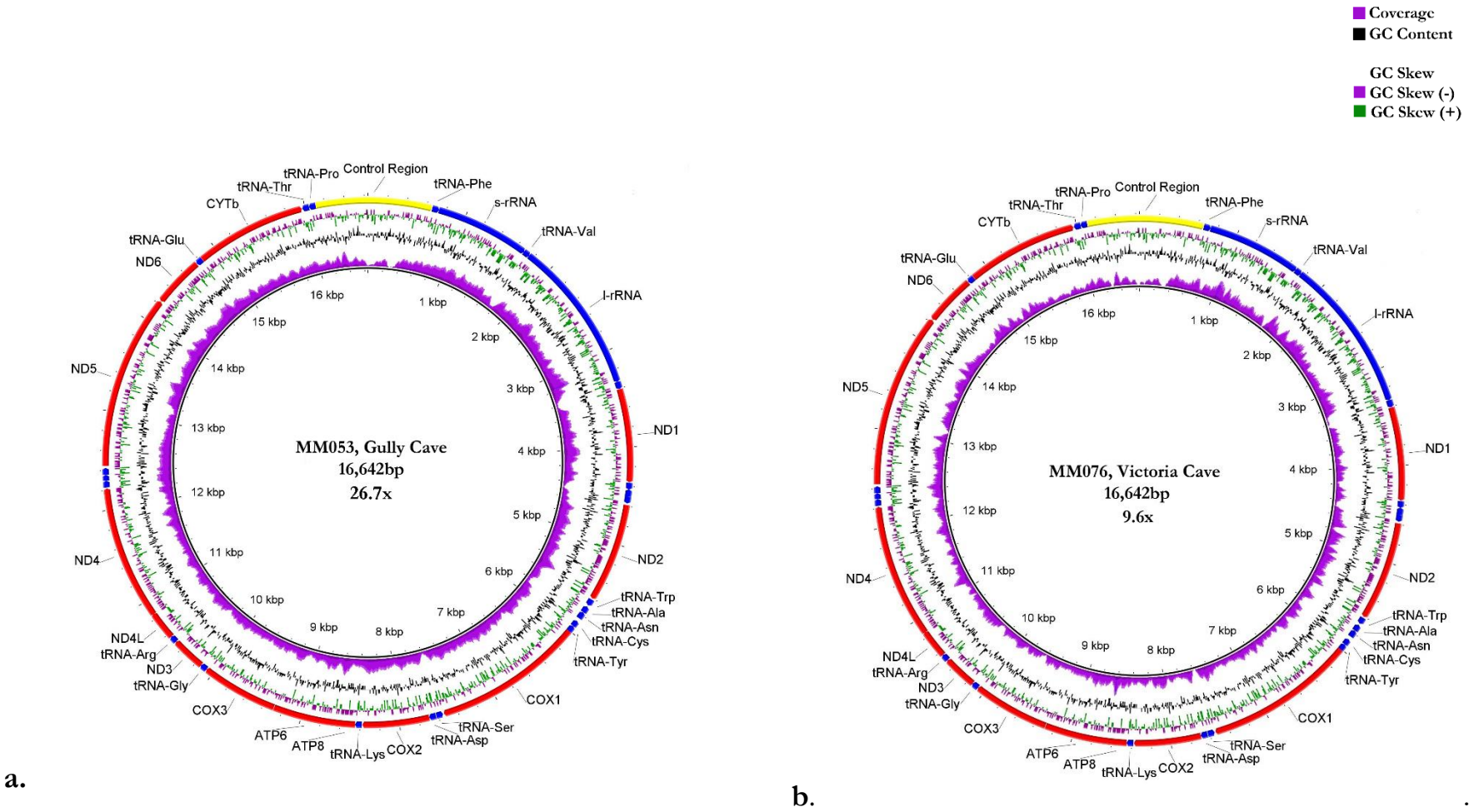
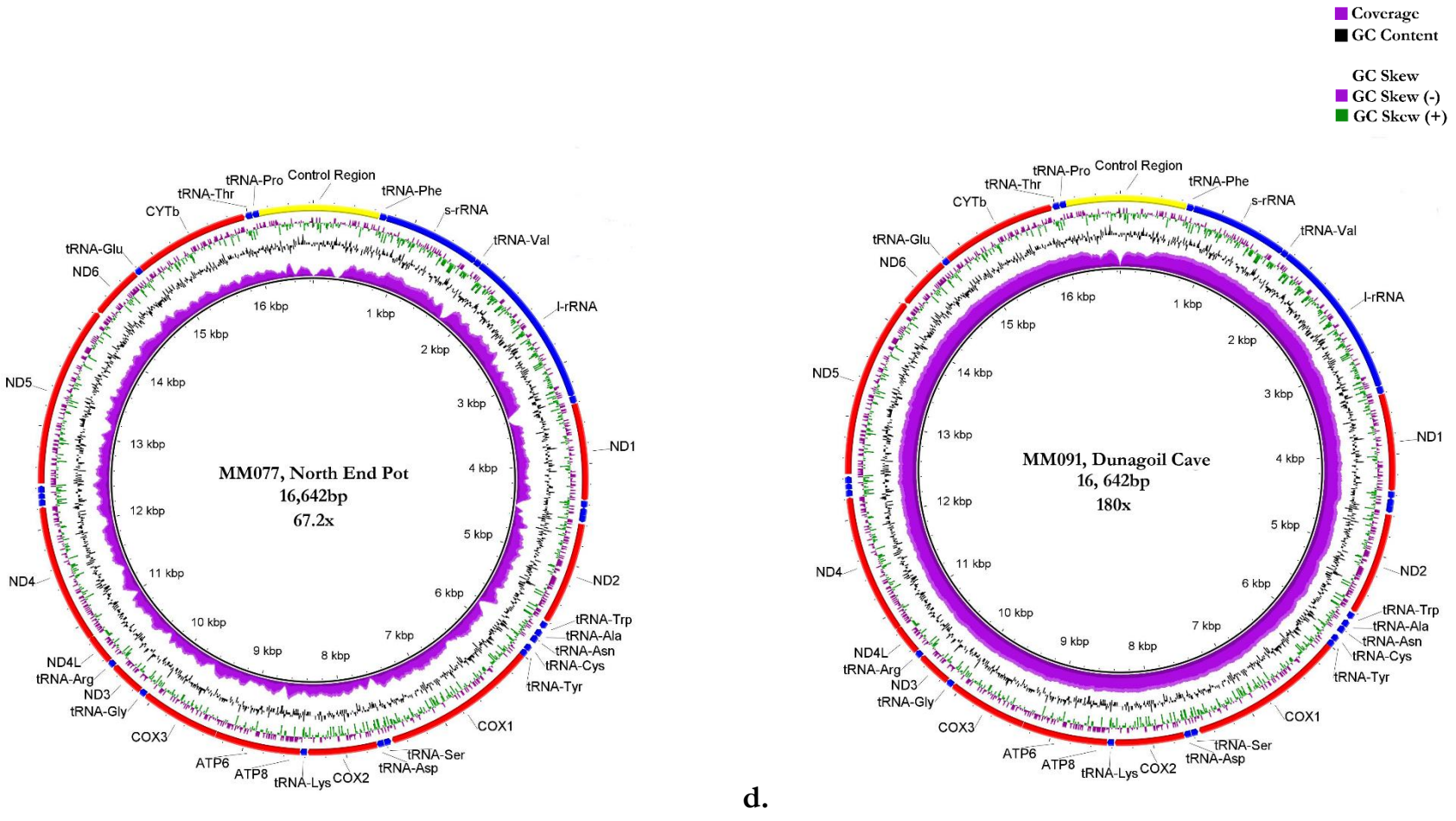
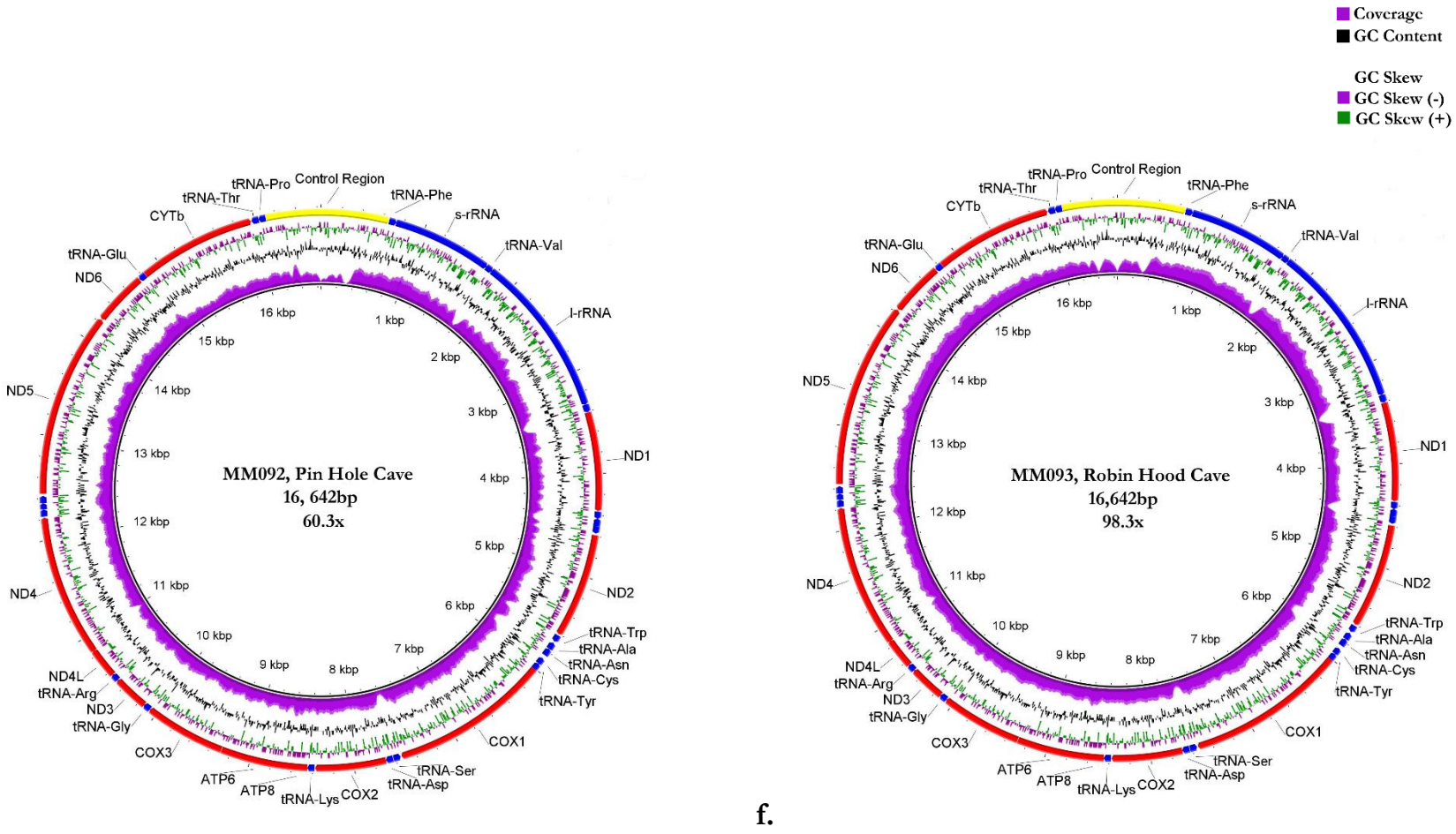


Figure 6.6 a – f. Annotated mitogenomes for *Felis silvestris silvestris* specimens used in this study. Images generated in Blast Ring Image Generator (BRIG: Alikham *et al.* 2011).



c. d.

Figure 6.6 a – f. Annotated mitogenomes for *Felis silvestris silvestris* specimens used in this study. Images generated in Blast Ring Image Generator (BRIG: Alikham *et al.* 2011).



e. **f.**

Figure 6.6 a – f. Annotated mitogenomes for *Felis silvestris silvestris* specimens used in this study. Images generated in Blast Ring Image Generator (BRIG: Alikham *et al.* 2011).

6.3.5 Phylogenetic placement

The phylogenetic placement of three individuals, MM076, MM077 and MM091 for which ages could either not be inferred or for which AMS ^{14}C dating placed them within the timeframe of the introduction of the domestic cat, were investigated via phylogenetic reconstruction of the *Felis silvestris* species group. Of the 61 sequences included in the alignment of the 2,604bp length sequence of the ND5/ND6 genes, 20 unique haplotypes were found and haplotype diversity (*Hd*)=0.865. An accurate estimation of per-site nucleotide diversity (π) and number of polymorphic sites was difficult to obtain here due to the large number of sites in the external (Genbank) sequences with missing data or ambiguous base calls. If these sites are included in the calculation of polymorphism metrics, then these statistics will be overinflated and if they are excluded, then they will be underestimated. Therefore, it was decided not to calculate these. Within the European wildcat clade, a total of 27 individuals shared 7 haplotypes. None of the individuals sequenced here had haplotypes unique to the ancient dataset with individuals MM076, MM077 and MM091 all sharing a haplotype that was also shared with 15 other extant *F. s. silvestris* samples.

Both the NJ and Bayesian MCMC phylogenetic trees recovered identical topologies at higher levels, which were also largely concordant, but not identical to, the phylogeny reported by Driscoll *et al.* 2007 for the *F. silvestris* group (Fig 6.7). The earliest clade to diverge in the phylogeny is the Chinese sand cat *F. s. bieti* - estimated to have occurred *c.* 230 000 kya (divergence date estimation follow those in Driscoll *et al.* 2007), followed by the European wildcat clade *F. s. silvestris* for which no divergence date estimate is available. The Central Asian wildcat, *F. s ornata*, Southern African wildcat *F. s. cafra* and Near East wildcat *F. s. lybica* (inc domestic cat *F. s. catus*) clade diverges later in the phylogeny at an estimated 173 000 kya with a divergence date for the *F. s. lybica*/*F. s catus* clade of 131 000 kya (node ages from Driscoll *et al.* 2007). Bootstrap and posterior probabilities for nodes and branches leading to the main clades were generally high bar the branch and node of the *F. s. silvestris* clade for which nodal support was 0.68 posterior probability and bootstrap support just 32. All three ancient wildcats group unambiguously within the monophyletic *F. s. silvestris* clade, indicating that these individuals do not have domestic mtDNA types at the

ND5/ND6 loci. While there is high bootstrap support (1) for a split within the group, with MM076 and (MM077 + MM091) placing in alternate groups, these do not represent discrete geographic areas and thus this is unlikely to represent biogeographic disparity.

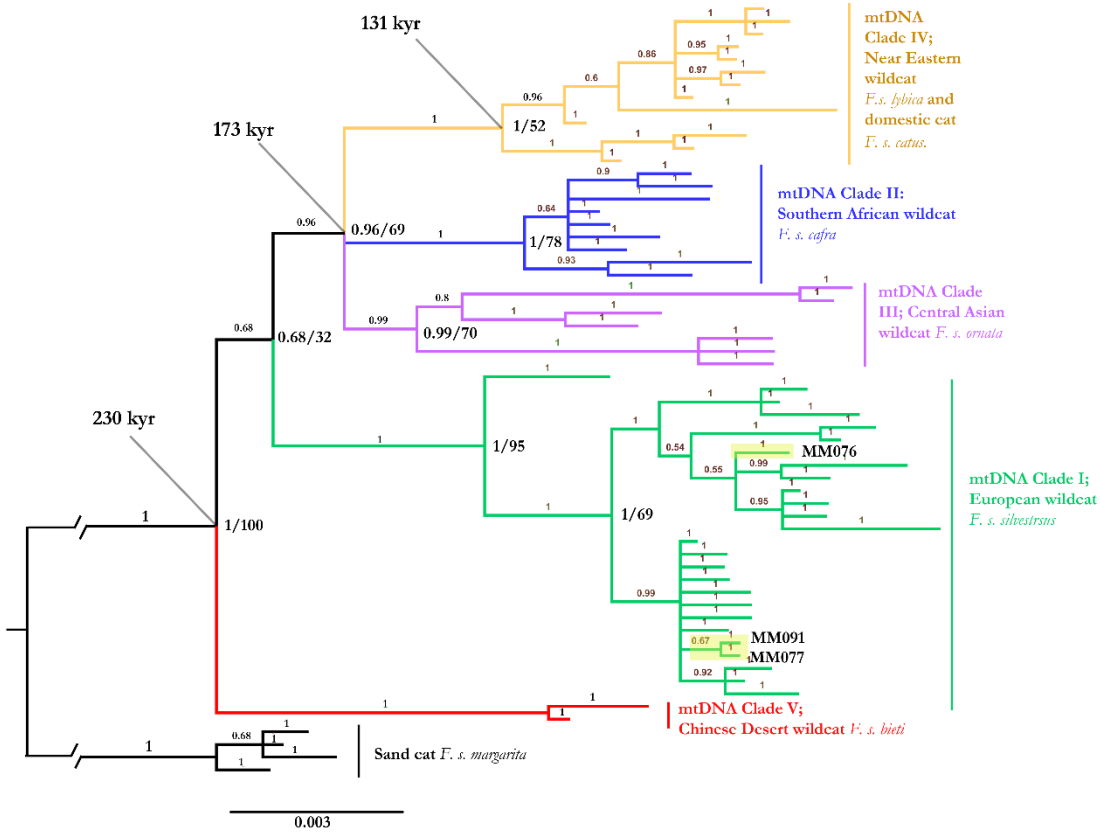


Figure 6.7. Phylogenetic reconstruction (Neighbour Joining/Bayesian MCMC) of wildcat *Felis silvestris* subspecies based on ND6 and ND6 mtDNA genes and including specimens MM076, MM077, MM091. Support values for major nodes in the phylogeny shown as posterior probability/bootstrap support and divergence dates between major clades are from Driscoll *et al.* 2007. All ancient British wildcats place in the *F. s. silvestris* clade.

6.3.6 Genetic diversity in the mitochondrial genome

As part of quantifying mitochondrial genetic diversity in this dataset, all specimens were examined for the presence of fixed nucleotide substitutions employed by Scottish Natural Heritage (SNH) to discriminate between wild and feral domestic cats. All of the ancient British wildcats showed fixed base substitutions at 12 loci diagnostic

for the discrimination of feral domestic cats and wildcats in the Scottish population (Table 6.6). This confirms that these mtDNA substitutions were present in the ancestral British wildcat population as far back as the Early Holocene and strongly suggests that these base locations have powerful utility to discriminate between extant pure-bred wildcats and feral domestic/hybrids individuals.

Table 6.6 Twelve mtDNA loci located in the ND5/ND6 genes which show fixed differences between wild and domestic cats. Row one shows bases found in domestic cats, row two shows fixed differences based on analysis of modern wildcats and subsequent rows show these base calls in the ancient British specimens.

	Nucleotide Position											
	12663	13265	13326	13554	13611	13656	13679	13770	14187	14555	15082	15164
Domestic Cat	T	T	G	A	T	A	G	A	C	A	T	T
Modern Wildcat	C	C	A	C	C	G	A	G	T	G	C	C
MM053	C	C	A	C	C	G	A	G	T	G	C	C
MM076	C	C	A	C	C	G	A	G	T	G	C	C
MM077	C	C	A	C	C	G	A	G	T	G	C	C
MM091	C	C	A	C	C	G	A	G	T	G	C	C
MM092	C	C	A	C	C	G	A	G	T	G	C	C
MM093	C	C	A	C	C	G	A	G	T	G	C	C

Pairwise measures of genetic distances were calculated for the mitochondrial genome as a whole and twelve protein-coding genes plus the non-coding control region for the British ancient wildcat sequences (Table 6.7, a-f). When sites with missing data and ambiguities were removed pairwise comparisons included 16,367 bases out of a total mitogenome length of 16,642 bases. Haplotype diversity (*Hd*) = 1, with all mitogenomes being unique. However, overall genetic diversity was remarkably low, considering the spatio-temporal spread of these individuals with only 38 polymorphic (segregating) sites across the entire mitogenome, of which 18 were parsimony informative. There were no sites with more than two variants. Nucleotide diversity (π) = 0.009, overall mean number of base differences (*V*) between individuals was 17.5 and mean Kimura-2 (K-2) distance was 0.001 (Table 6.x, a-f).

The number of pairwise base differences ranged from 2 (Kimura-2 = 0; 0.02 percent genetic distance) between MM053 and MM076 to 28 between MM092 and MM093 (K-2 =0.002; 0.168 percent genetic distance). Both of these are surprising results. A

difference of just two bases between MM053 from Gully Cave, Somerset and MM076 from Victoria Cave, Yorkshire is a remarkably low amount of genetic difference between these samples when measured across the entire mitogenome. No age is available for sample MM076 but contextual evidence would suggest a likely age of *c.* Mid Holocene while MM053 has an Early Holocene AMS ^{14}C date of 10, 227 – 9, 921 cal BP. These specimens are therefore not chronologically contemporaneous, potentially being separated by several thousand years, and come from geographic areas separated by over 300km. Conversely, specimens MM092 and MM093 have the highest pairwise genetic distances (K-2 = 0.168) and number of base differences (28) between them, despite both being from the same locality (Pin Hole Cave and Robin Hood Cave, Creswell Crags, Derbyshire) and separated by just *c.* 1,500 years.

Of the coding sequences, lowest diversity was seen in the ATP8, COX3 and ND6 genes, which all have identical haplotypes in this dataset and appear to be highly conserved. Low diversity and genetic distances were also observed in COX1 where only two haplotypes were observed, of which five individuals shared a haplotype and one sample (MM077) differed by a single base. Similarly, low genetic distances were also seen between wildcats for COX2 and ND1, which both had only two haplotypes of which only one individual differed by a single base substitution. The highest levels of variability ($Hd \geq 0.8$) were observed in the non-coding Control Region (Hd = 0.933, number of haplotypes = 5, V = 2.6, π =0.002, mean K-2 = 0.003), ATP6 (Hd = 0.8, number of haplotypes = 4, V = 1.86, π =0.002, mean K-2 = 0.003), ND2 (Hd = 0.866, number of haplotypes = 4, V = 2.2, π =0.002, mean K-2 = 0.002) and ND4 (Hd = 0.8, number of haplotypes = 4, V = 1.27, π =0.001, mean K-2 = 0.001). The cytochrome*b* gene, often employed on population level studies, showed moderate levels of variability with three haplotypes (Hd = 0.73), 3 polymorphic sites across the 1,134b length of the gene, nucleotide diversity of 0.001, mean number of base differences = 1.5 and a mean K-2 distance of 0.001.

A MJ network utilising the whole mitogenome distinguished between two groups separated by a minimum of 1 median vector and 21 base substitutions when the shortest route between haplotypes is taken (Fig 6.8). Samples MM072 (Yorkshire,

undated), MM091 (Isle of Bute, 1,814 – 1,618 cal BP) and MM093 (Derbyshire, 7,239 – 6,999 cal BP) form a group at one extremity of the bipartition while MM053 (Somerset, 10,227 cal BP), MM073 (Yorkshire, undated) and MM092 (Derbyshire, 8,715 – 8,484) form a group at the other. It is difficult to ascertain whether this represents a geographical split as there are samples from Yorkshire and Derbyshire in each group, with one group containing an individual from Somerset and the other one individual from Scotland. There is a potential chronological split with mid-late Holocene individuals forming one group while early Holocene individuals form another. Interpretation of this result should be approached cautiously due to the low sample size and the inclusion of two individuals for which no AMS ^{14}C age is available (MM076 and MM077) but the network suggests the existence of two groups that, interestingly, occur pre-and post-the 8.2 ka cold event. The fixation index gave an F_{ST} = 0.729, indicating a very high level of population differentiation (although this may be affected by low sample size, n=3 in each putative population). Fu's F and Tajima's D were non-significant (F = -0.0025, $p > 0.1$; Tajima's D = 0.315, $p > 0.1$) and mismatch analysis showed that the data deviated significantly from a pattern consistent with a demographic expansion (SSD = 0.032, p = 0.041). While these results may be influenced by low sample size and undetected population substructure, they suggest that the genetic patterns observed here are not the result of a population expansion.

Median Joining networks constructed with single coding genes and the non-coding Control Region failed to provide any resolution in the networks they were used to construct (Fig 6.9). The mitochondrial genome evolves at different rates across coding and non-coding regions and thus different loci may produce different scenarios of phylogenetic relationships given the individual history of different regions and sub-regions. While the CR, ATP6, ND2 and ND4 regions all showed the highest levels of haplotype, nucleotide and pairwise K-2 distances, variability was still too low to produce a resolved network. This demonstrates the utility of using whole mitogenomes in phylogenetic reconstruction when genetic diversity in specific sub-regions is low.

Mitogenome. Length 16,642bp. No analysed bases: 16,367.						
No haplotypes: 6. Mean no of base differences: 17.467 Overall mean distance: 0.001						
	MM053	**MM076**	**MM077**	**MM091**	**MM092**	**MM093**
MM053		0.000	0.001	0.001	0.000	0.001
MM076	2 (0.012)		0.002	0.001	0.000	0.002
MM077	23 (0.138)	25 (0.15)		0.000	0.002	0.001
MM091	21 (0.126)	23 (0.138)	8 (0.048)		0.001	0.001
MM092	4 (0.024)	6 (0.036)	27 (0.162)	25 (0.15)		0.002
MM093	24 (0.114)	26 (0.156)	11 (0.066)	9 (0.05)	28 (0.168)	

Hd	1
Haplotype 1	MM053
Haplotype 2	MM076
Haplotype 3	MM077
Haplotype 4	MM091
Haplotype 5	MM092
Haplotype 6	MM093
No polymorphic sites	38
(π)	0.009

Cytochrome*b*. Length 1,134bp. No analysed bases: 1,134.						
No haplotypes: 3. Mean no of base difference: 1.533. Overall K2 mean distance: 0.001						
	MM053	**MM076**	**MM077**	**MM091**	**MM092**	**MM093**
MM053		0.001	0.002	0.002	0.000	0.002
MM076	1 (0.09)		0.003	0.003	0.001	0.003
MM077	2 (0.176)	3 (0.26)		0.000	0.002	0.000
MM091	2 (0.176)	3 (0.26)	0 (0.00)		0.002	0.000
MM092	0 (0.00)	1 (0.09)	2 (0.176)	2 (0.176)		0.002
MM093	2 (0.176)	3 (0.26)	0 (0.00)	0 (0.00)	2 (0.176)	

Hd	0.73
Haplotype 1	MM053, MM092
Haplotype 2	MM076
Haplotype 3	MM077, MM091, MM093
No polymorphic sites	3
(π)	0.001

Control Region. Length 1,188bp. No analysed bases: 1,021						
No haplotypes: 5 . Mean no of base differences: 2.6. Overall K2 mean distance: 0.003						
	MM053	**MM076**	**MM077**	**MM091**	**MM092**	**MM093**
MM053		0.000	0.002	0.003	0.003	0.001
MM076	0 (0.000)		0.002	0.003	0.003	0.001
MM077	2 (0.19)	2 (0.19)		0.003	0.003	0.003
MM091	3 (0.29)	3 (0.29)	3 (0.29)		0.004	0.004
MM092	3 (0.29)	3 (0.29)	3 (0.29)	4 (0.39)		4.000
MM093	1 (0.09)	1 (0.09)	3 (0.29)	4 (0.39)	4 (0.39)	

Hd	0.933
Haplotype 1	MM076, MM.53
Haplotype 2	MM091
Haplotype 3	MM093
Haplotype 4	MM077
Haplotype 5	MM092
No polymorphic sites	7
(π)	0.002

Tables 6.7 (a-c). Pairwise genetic distance and measures of genetic diversity for all *Felis silvestris silvestris* individuals that yielded full-length mitogenomes. Lower left areas show number of nucleotide difference between aligned pairs of sequences and net percentage difference while the upper right areas show Kimura-2 distances. Also given re the number of bases included in the analyses (*i.e* minus missing data and ambiguities), number of haplotypes, haplotype diversity (Hd), and nucleotide diversity (π) and overall mean number of base difference (V).

ATP6. Length 675bp. No analysed bases: 675 No haplotypes: 4. Mean no of base differences: 1.86. Overall K2 mean distance: 0.003						
	MM053	**MM076**	**MM077**	**MM091**	**MM092**	**MM093**
MM053		0.000	0.000	0.004	0.004	0.003
MM076	0 (0.000)		0.000	0.004	0.004	0.003
MM077	0 (0.000)	0 (0.000)		0.004	0.004	0.003
MM091	3 (0.293)	3 (0.293)	3 (0.293)		0.003	0.001
MM092	3 (0.293)	3 (0.293)	3 (0.293)	2 (0.195)		0.001
MM093	2 (0.195)	2 (0.195)	2 (0.195)	1 (0.097)	1 (0.097)	

Hd	0.8
Haplotype 1	MM053, MM076, MM092
Haplotype 2	MM077
Haplotype 3	MM091
Haplotype 4	MM093
No polymorphic sites	4
(π)	0.002

ATP8. Length 198bp. No analysed bases: 198 No haplotypes: 1. Mean no of base differences: 0.00. Overall K2 mean distance: 0.000						
	MM053	**MM076**	**MM077**	**MM091**	**MM092**	**MM093**
MM053		0.000	0.000	0.000	0.000	0.000
MM076	0 (0.000)		0.000	0.000	0.000	0.000
MM077	0 (0.000)	0 (0.000)		0.000	0.000	0.000
MM091	0 (0.000)	0 (0.000)	0 (0.000)		0.000	0.000
MM092	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)		0.000
MM093	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	

Hd	0
Haplotype 1	MM053, MM076, MM077, MM091, MM092, MM093
No polymorphic sites	0
(π)	0.000

Cox1. Length 1,533bp. No analysed bases: 1,530. No haplotypes: 2. Mean no of base difference: 0.333. Overall K2 mean distance: 0.000						
	MM053	**MM076**	**MM077**	**MM091**	**MM092**	**MM093**
MM053		0.000	0.001	0.000	0.000	0.000
MM076	0 (0.00)		0.001	0.000	0.000	0.000
MM077	1 (0.065)	1 (0.065)		0.001	0.001	0.001
MM091	0 (0.00)	0 (0.00)	1 (0.065)		0.000	0.000
MM092	0 (0.00)	0 (0.00)	1 (0.065)	0 (0.00)		0.000
MM093	0 (0.00)	0 (0.00)	1 (0.065)	0 (0.00)	0 (0.00)	

Hd	0.333
Haplotype 1	MM053, MM076, MM091, MM092, MM093
Haplotype 2	MM077
No polymorphic sites	1
(π)	0.000

Tables 6.7 (d-f). Pairwise genetic distance and measures of genetic diversity for all *Felis silvestris silvestris* individuals that yielded full-length mitogenomes. Lower left areas show number of nucleotide difference between aligned pairs of sequences and net percentage difference while the upper right areas show Kimura-2 distances. Also given re the number of bases included in the analyses (*i.e* minus missing data and ambiguities), number of haplotypes, haplotype diversity (Hd), and nucleotide diversity (π) and overall mean number of base difference (V).

Cox2. Length 681bp. No analysed bases: 681. No haplotypes: 2. Mean no of base differences: 0.333. Overall K2 mean distance: 0.000						
	MM053	**MM076**	**MM077**	**MM091**	**MM092**	**MM093**
MM053		0.000	0.000	0.000	0.000	0.001
MM076	0 (0.000)		0.000	0.000	0.000	0.001
MM077	0 (0.000)	0 (0.000)		0.000	0.000	0.001
MM091	0 (0.000)	0 (0.000)	0 (0.000)		0.000	0.001
MM092	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)		0.001
MM093	1 (0.147)	1 (0.147)	1 (0.147)	1 (0.147)	1 (0.147)	

Hd	0.333
Haplotype 1	MM053, MM076, MM077, MM091, MM092
Haplotype 2	MM093
No polymorphic sites	1
(π)	0.000

Cox3. Length 783bp. No analysed bases: 776. No haplotypes: 1. Mean no of base differences: 0.000 Overall K2 mean distance: 0.000						
	MM053	**MM076**	**MM077**	**MM091**	**MM092**	**MM093**
MM053		0.000	0.000	0.000	0.000	0.000
MM076	0 (0.000)		0.000	0.000	0.000	0.000
MM077	0 (0.000)	0 (0.000)		0.000	0.000	0.000
MM091	0 (0.000)	0 (0.000)	0 (0.000)		0.000	0.000
MM092	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)		0.000
MM093	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	

Hd	0
Haplotype 1	MM053, MM076, MM077, MM091, MM092, MM093
No polymorphic sites	0
(π)	0.000

ND1. Length 951bp. No analysed bases: 926. No haplotypes: 2. Mean no of base differences: 0.333. Overall K2 mean distance: 0.000						
	MM053	**MM076**	**MM077**	**MM091**	**MM092**	**MM093**
MM053		0.000	0.000	0.000	0.001	0.000
MM076	0 (0.000)		0.000	0.000	0.001	0.000
MM077	0 (0.000)	0 (0.000)		0.000	0.001	0.000
MM091	0 (0.000)	0 (0.000)	0 (0.000)		0.001	0.000
MM092	1 (0.105)	1 (0.105)	1 (0.105)	1 (0.105)		0.001
MM093	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.105)	

Hd	0.333
Haplotype 1	MM076, MM092, MM053, MM077, MM091
Haplotype 2	MM093
No polymorphic sites	1
(π)	0.000

Tables 6.7 (g-i cont.). Pairwise genetic distance and measures of genetic diversity for all *Felis silvestris silvestris* individuals that yielded full-length mitogenomes. Lower left areas show number of nucleotide difference between aligned pairs of sequences and net percentage difference while the upper right areas show Kimura-2 distances. Also given re the number of bases included in the analyses (*i.e* minus missing data and ambiguities), number of haplotypes, haplotype diversity (Hd), and nucleotide diversity (π) and overall mean number of base difference (V).

ND2. Length 1,029bp. No analysed bases: 1,029. No haplotypes: 4 . Mean no of base differences: 2.2. Overall K2 mean distance: 0.002						
	MM053	**MM076**	**MM077**	**MM091**	**MM092**	**MM093**
MM053		0.000	0.002	0.003	0.002	0.002
MM076	0 (0.000)		0.002	0.003	0.002	0.002
MM077	2 (0.194)	2 (0.194)		0.001	0.004	0.000
MM091	3 (0.291)	3 (0.291)	1 (0.097)		0.005	0.001
MM092	2 (0.194)	2 (0.194)	4 (0.388)	5 (0.486)		0.004
MM093	2 (0.194)	2 (0.194)	0 (0.000)	1 (0.097)	4 (0.388)	

Hd	0.866
Haplotype 1	MM053, MM076
Haplotype 2	MM077, MM093
Haplotype 3	MM091
Haplotype 4	MM092
No polymorphic sites	5
(π)	0.002

ND3. Length 345bp. No analysed bases: 334. No haplotypes: 2. Mean no of base differences: 0.6. Overall K2 mean distance: 0.002						
	MM053	**MM076**	**MM077**	**MM091**	**MM092**	**MM093**
MM053		0.000	0.003	0.003	0.000	0.003
MM076	0 (0.000)		0.003	0.003	0.000	0.003
MM077	1 (0.299)	1 (0.299)		0.000	0.003	0.000
MM091	1 (0.299)	1 (0.299)	0 (0.000)		0.003	0.000
MM092	0 (0.000)	0 (0.000)	1 (0.299)	1 (0.299)		0.003
MM093	1 (0.299)	1 (0.299)	0 (0.000)	0 (0.000)	1 (0.299)	

Hd	0.6
Haplotype 1	MM053, MM076, MM072
Haplotype 2	MM077, MM091, MM093
No polymorphic sites	1
(π)	0.001

ND4. Length 1,368bp. No analysed bases: 1,367 No haplotypes: 4. Mean no of base differences: 1.267 Overall K2 mean distance: 0.001						
	MM053	**MM076**	**MM077**	**MM091**	**MM092**	**MM093**
MM053		0.000	0.000	0.001	0.001	0.001
MM076	0 (0.000)		0.000	0.001	0.001	0.001
MM077	0 (0.000)	0 (0.000)		0.001	0.001	0.001
MM091	2 (0.146)	2 (0.146)	2 (0.146)		0.001	0.001
MM092	2 (0.146)	2 (0.146)	2 (0.146)	2 (0.146)		0.001
MM093	1 (0.073)	1 (0.073)	1 (0.073)	1 (0.073)	1 (0.073)	

Hd	0.8
Haplotype 1	MM076, MM092, MM053
Haplotype 2	MM077
Haplotype 3	MM093
Haplotype 4	MM091
No polymorphic sites	3
(π)	0.001

Tables 6.7 (j-l cont.). Pairwise genetic distance and measures of genetic diversity for all *Felis silvestris silvestris* individuals that yielded full-length mitogenomes. Lower left areas show number of nucleotide difference between aligned pairs of sequences and net percentage difference while the upper right areas show Kimura-2 distances. Also given re the number of bases included in the analyses (*i.e* minus missing data and ambiguities), number of haplotypes, haplotype diversity (Hd), and nucleotide diversity (π) and overall mean number of base difference (V).

ND5. Length 1,782bp. No analysed bases: 1,766 No haplotypes: 3. Mean no of base differences: 2.73. Overall K2 mean distance: 0.002						
	MM053	**MM076**	**MM077**	**MM091**	**MM092**	**MM093**
MM053		0.000	0.002	0.002	0.001	0.002
MM076	0 (0.000)		0.002	0.002	0.001	0.002
MM077	4 (0.226)	4 (0.226)		0.000	0.003	0.000
MM091	4 (0.226)	4 (0.226)	0 (0.000)		0.003	0.000
MM092	1 (0.056)	1 (0.056)	5 (0.283)	5 (0.283)		0.003
MM093	4 (0.226)	4 (0.226)	0 (0.000)	0 (0.000)	5 (0.283)	

Hd	0.73
Haplotype 1	MM053, MM076
Haplotype 2	MM077, MM091, MM093
Haplotype 3	MM092
No polymorphic sites	3
(π)	0.001

ND6. Length 519bp. No analysed bases: 519 No haplotypes: 1. Mean no of base differences: 0.000. Overall K2 mean distance: 0.000						
	MM053	**MM076**	**MM077**	**MM091**	**MM092**	**MM093**
MM053		0.000	0.000	0.000	0.000	0.000
MM076	0 (0.000)		0.000	0.000	0.000	0.000
MM077	0 (0.000)	0 (0.000)		0.000	0.000	0.000
MM091	0 (0.000)	0 (0.000)	0 (0.000)		0.000	0.000
MM092	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)		0.000
MM093	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	

Hd	0
Haplotype 1	MM053, MM076, MM077, MM091, MM092, MM093
No polymorphic sites	0
(π)	0.000

Tables 6.7 (m-n cont.). Pairwise genetic distance and measures of genetic diversity for all *Felis silvestris silvestris* individuals that yielded full-length mitogenomes. Lower left areas show number of nucleotide difference between aligned pairs of sequences and net percentage difference while the upper right areas show Kimura-2 distances. Also given re the number of bases included in the analyses (*i.e* minus missing data and ambiguities), number of haplotypes, haplotype diversity (Hd), and nucleotide diversity (π) and overall mean number of base difference (*V*).

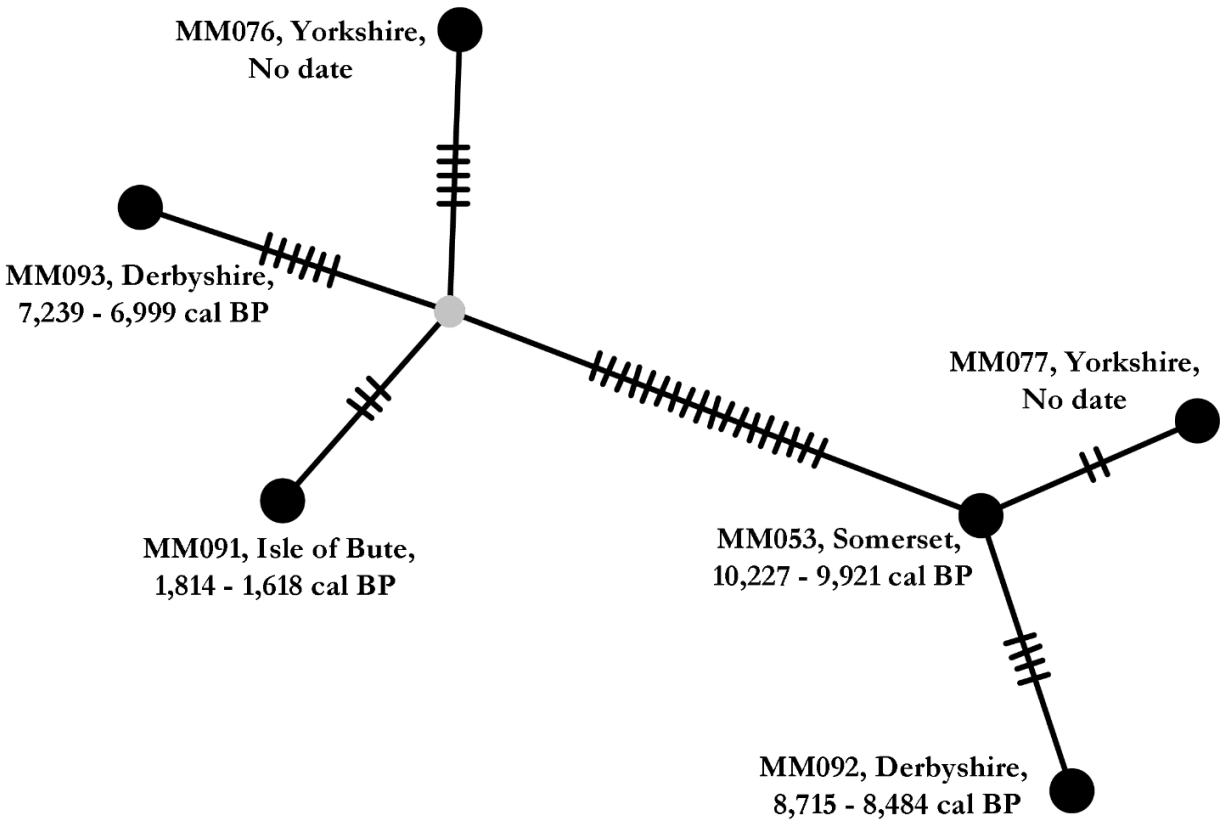
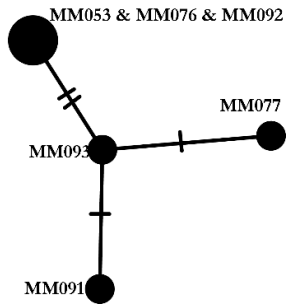
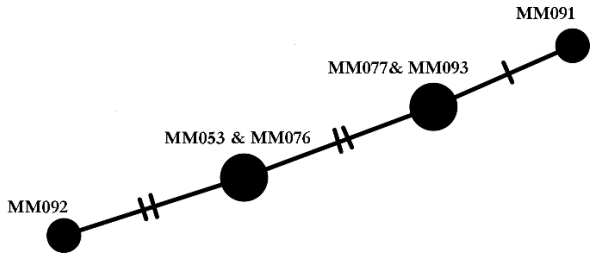


Figure 6.8. Median Joining (MJ) network for European wildcat *Felis silvestris silvestris* whole mitochondrial genomes generated in this study. Mitogenome haplotypes fall into two groups separated by a minimum of 21 substitutions from MM053 to MM091. Grey median vector represents an absent, hypothesised, missing haplotype.

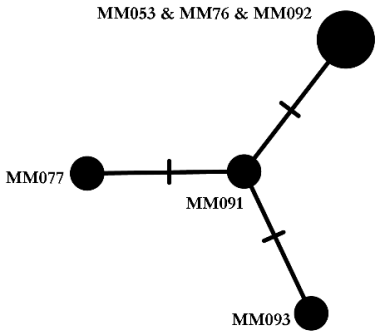
a. ATP6



b. ND2



c. ND4



d. Control Region

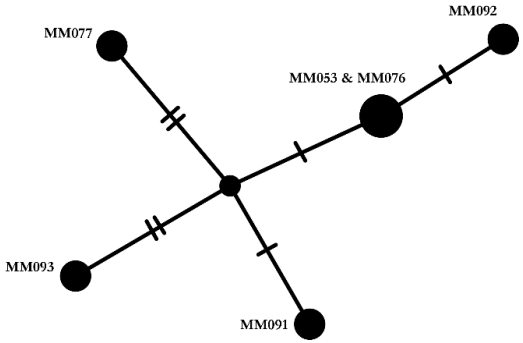


Figure 6.9. Single gene (ATP6, ND2, ND4 and Control Region) Median-Joining networks for British wildcat *Felis silvestris silvestris*.

6.3.7 Phylogeographic placement

Of the 40 *F. s. silvestris* individuals in this dataset sequenced at a 433bp region of the CR, 10 unique haplotypes were identified with *Hd* = 0.747 and π =0.156. Samples MM091, MM093, MM077 each had unique haplotypes not shared by any other ancient or modern wildcat whereas MM053, MM076 and MM092 had haplotypes both identical to each other and to five modern wildcats from west Germany and Luxembourg. The number of polymorphic (variable) sites was 19, of which 14 were parsimony informative and 5 were singleton variants. There were no sites which showed more than two variants.

Two strongly-supported monophyletic groups with nine fixed nucleotide differences were distinguished in the Bayesian MCMC phylogenetic reconstruction (Fig 6.10). These were a central Germany group and a west Germany/central Europe group. Within Lineage 1 (west Germany/central Europe) the number of polymorphic sites were 7 and there were 7 unique haplotypes. Variability was lower within Lineage 2 (central Germany) with the number of polymorphic sites 3 and the number of unique haplotypes was just 3. British samples grouped together and unambiguously within the matrilineage representing west Germany/central Europe (Fig 6.10). The modern CR samples in this clade were all sampled from localities in either Luxembourg or west Germany but this area is likely to belong to (and be representative of) a much larger clade that includes France and Belgium (Figs 6.2 – 6.3). Within these two clades, there was little sub-structure which can be indicative either of rapid population expansion or, and perhaps more likely in this case given the above sequence statistics, a lack of informative characters from which to infer finer-scale topologies.

Mitochondrial sequence data spanning the full European range of *F. s. silvestris* is not currently available and so the full mitochondrial phylogeography of this species in Europe is unknown. Without range-wide data population level topologies cannot be obtained. However, some inferences can be made with these results regarding the phylogenetic history of ancient Holocene wildcats from Britain. The phylogenetic reconstruction obtained here strongly suggests that all British Holocene wildcats

belong to a large clade that spanned central to west-central Europe, including the north-west European coastline. Consequently, it is reasonable to conclude that this was the most likely geographical area for post-glacial recolonisation of Britain. Further sampling of mitochondrial markers or whole mitogenomes from extant wildcat populations across Europe, and in particular the NW coast, would be needed to fully confirm this assertion.

6.3.8 Light-Coverage whole genome of MM091

Repeated DNA library builds and shotgun sequencing on an Illumina NextSeq500 platform enabled retrieval of DNA sequence reads from specimen MM091 (Dunagoil Cave, Isle of Bute) that, when assembled to the domestic cat genome v8.0, represented a 1.9x light coverage whole *F. s. silvestris* genome. This is the first ancient European wildcat genome to be sequenced and assembled. Table 6.8 shows summary statistics for the assembly and Table 6.9, Figs 6.11 - 6.13 show coverage statistics. Coverage metrics have not been corrected for the 'mappability' of the domestic cat genome and should therefore be viewed as minimum estimates only. A total of 67,994,993 uniquely mapping reads were aligned to the domestic cat *F. catus* genome v8.0 using the BWA parameters indicated in Chapter 2. Read length was typical for ancient DNA with a mean length of 69.8 bases and a range of 25 to 141 bases. Coverage (x) was generally similar over the 18 autosomes, ranging from 1.88x on ChrE1 to 2.08x on ChrF2. A decrease in coverage is associated with GC rich areas (Fig 6.11) and overall GC content was 40.50 percent. A noticeably lower read coverage of 1.04x was on ChrX. This is likely the result of the unique arrangement of certain regions of the X chromosome, which contain repetitive elements as well as inverted repeats, both features that can reduce the ability to map short-fragment NGS reads.

Read depth varied considerably across genomic locations as would be expected from the methods employed here which pooled reads derived from both shotgun sequencing and target enrichment (Table 6.10). The overall fraction of the genome

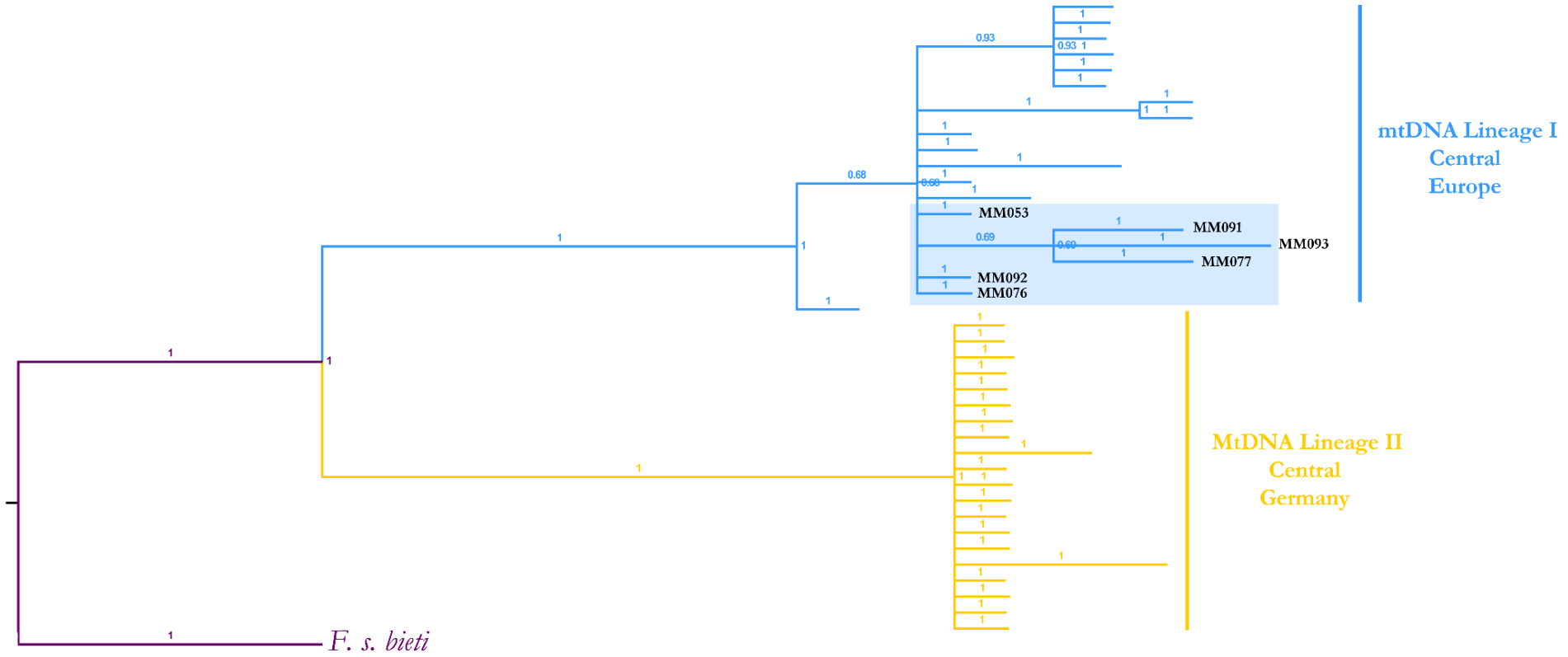


Figure 6.10. Bayesian MCMC phylogeny of European wildcat *Felis silvestris silvestris* using a 433bp sequence of the Control, Region (CR). The phylogenic reconstruction recovers two well-supported monophyletic clades with samples from Central Germany forming a distinct lineage and samples from Luxembourg and western Germany forming a phylogroup with the British samples. The latter likely represents a large western European lineage which also includes France and Belgium.

covered at ≥ 1 read depth is 79.5 percent, leaving 20.5 percent of the *F. catus* genome with zero base coverage. All chromosomes had at least 75 percent or more of the reference sequences covered at a minimum of one read depth with the majority of regions covered at between 1-10 reads. The exception to this was the X chromosome which had only 59.4 percent of the reference sequence covered. The large fraction of repetitive elements present in the domestic cat genome (*c.* 57 percent Tamazian *et al.* 2014) undoubtedly had a large influence on both coverage and read depth estimates, likely making up a large proportion of the unmapped regions in this assembly (although the mappability if different genomic regions was not quantified during this study).

Table 6.8 Summary properties for the alignment of NGS reads from specimens MM091 (Dunagoil Cave, Isle of Bute) to the domestic cat *Felis catus* genome v 8.0.

Reference Size (# bases)	2,419,212,910
Number of mapped reads	67,994,993
Read Min/Max/Mean Length	25/141/69.81
Percentage A's	29.74%
Percentage C's	20.25%
Percentage T's	29.76%
Percentage G's	20.26%
GC Percent	40.50%

Table 6.9 Whole genome and chromosome coverage (x) for alignment of NGS reads from specimens MM091 (Dunagoil Cave, Isle of Bute) to the domestic cat *Felis catus* genome v 8.0.

	Coverage (x) *Felis catus* v. 8.0		
Region	**Length**	**Mean**	**Standard Deviation**
ChrA1	240,380,223	1.96	1.64
ChrA2	168,638,799	1.92	1.72
ChrA3	140,925,898	2.06	1.93
ChrB1	206,538,554	2.05	1.72
ChrB2	152,998,503	2.04	1.64
ChrB3	148,068,395	2.01	1.62
ChrB4	142,431,058	2.02	1.66
ChrC1	222,198,629	2.04	1.64
ChrC2	159,252,932	2.05	1.61
ChrD1	115,468,741	2.03	1.69
ChrD2	88,096,124	2.02	1.67
ChrD3	94,101,111	2.03	1.63
ChrD4	94,492,513	2.02	1.64
ChrE1	61,081,816	1.88	1.87
ChrE2	61,960,243	1.94	1.9
ChrE3	41,224,383	1.96	1.75
ChrF1	70,119,229	1.99	1.69
ChrF2	83,953,389	2.08	1.65
ChrX	127,282,370	1.04	1.29
Whole Genome	**2,419,212,910**	**1.96**	**1.69**

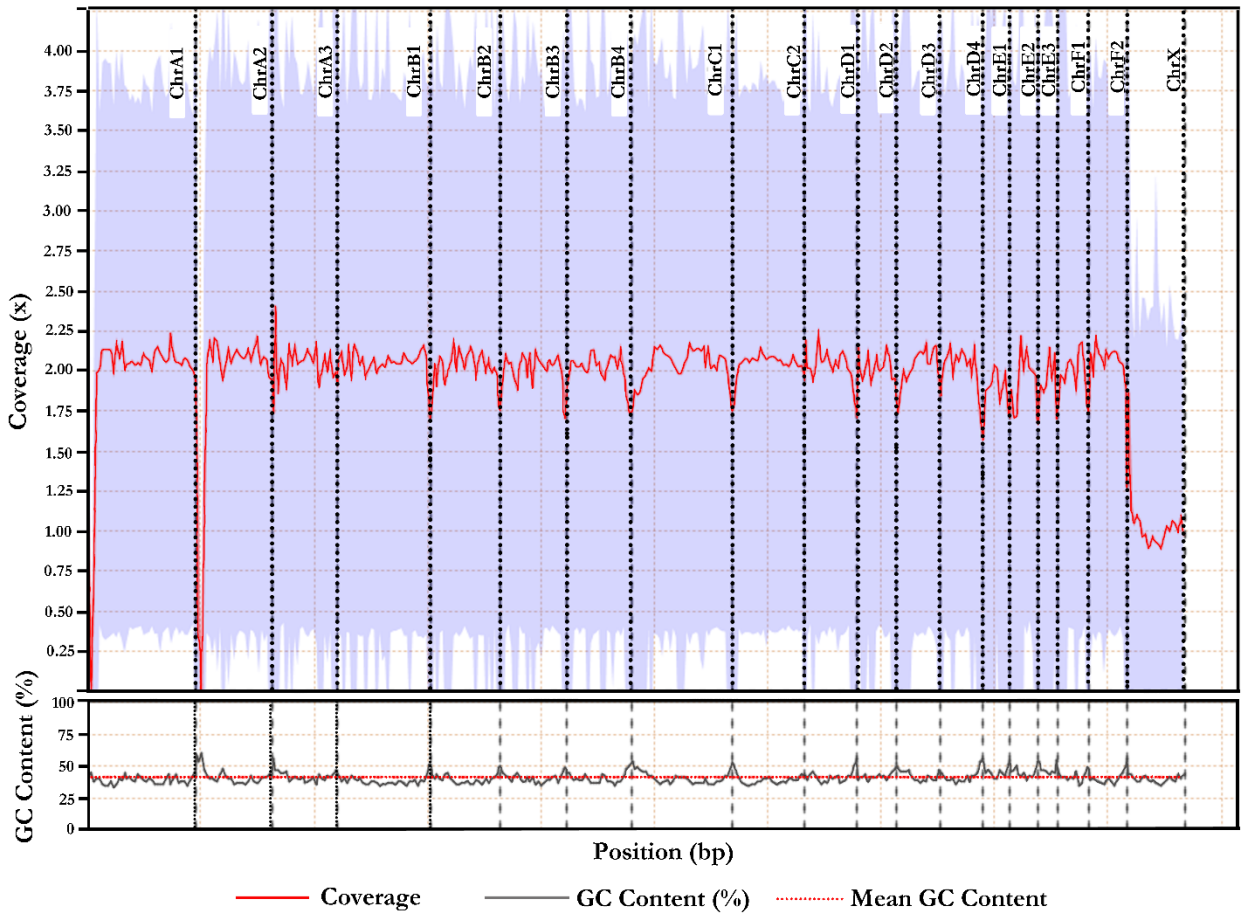


Figure 6.11. Coverage (x) across all 18 autosomes and the X chromosome for mapping of Illumina reads from specimen MM091 to the domestic cat *Felis catus* genome v8.0. Blue region shows standard deviation. Coverage is lower across the X chromosome and varies in association with GC content.

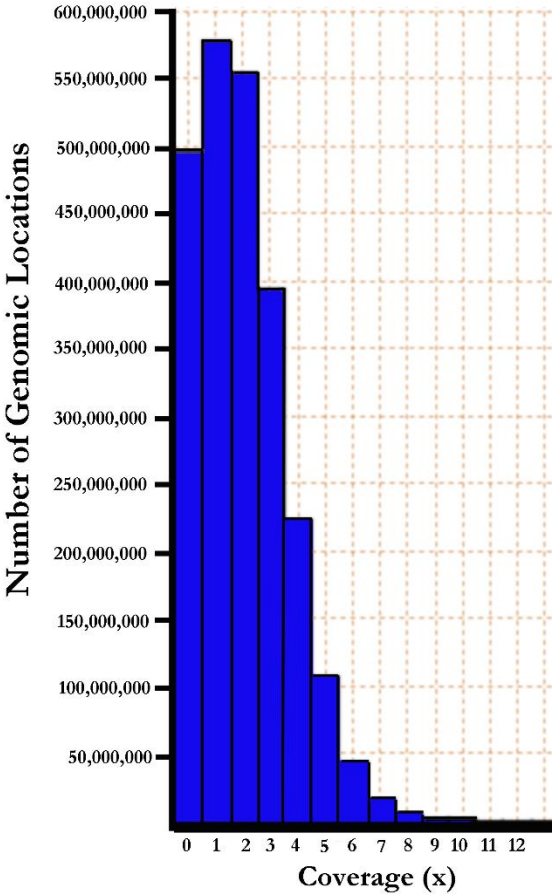


Figure 6.12. Number of genomic locations at x coverage for specimen MM091

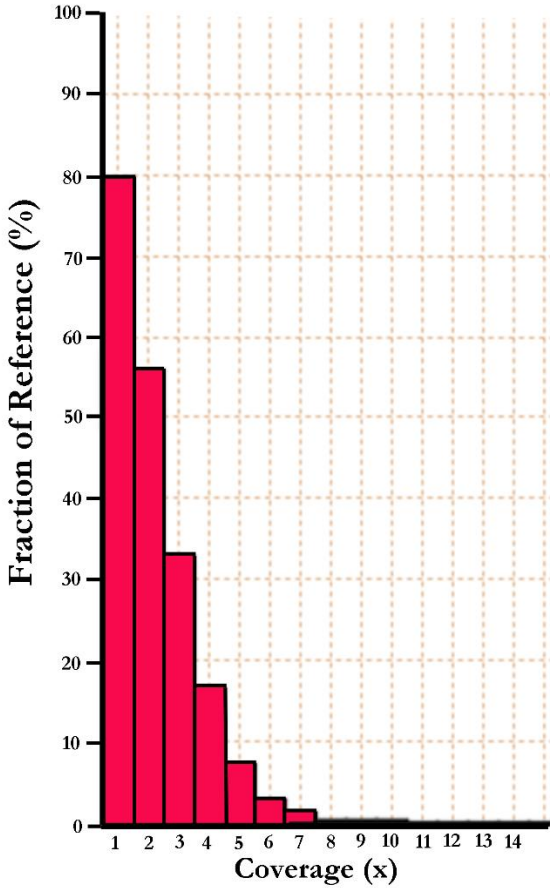


Figure 6.13. Fraction of reference sequence (*F. catus* v8.0) at x coverage for specimen MM091

Table 6.10. Minimum and maximum read depth per chromosome for MM091 Illumina reads aligned to the *Felis catus* v8.0 genome.

Region	Min Read Depth	Max Read Depth	Percent bases covered at ≥ 1
ChrA1	0	173	78.00
ChrA2	0	154	76.12
ChrA3	0	213	81.72
ChrB1	0	173	81.57
ChrB2	0	178	81.54
ChrB3	0	163	80.98
ChrB4	0	200	81.18
ChrC1	0	159	81.99
ChrC2	0	123	82.00
ChrD1	0	150	81.35
ChrD2	0	160	81.42
ChrD3	0	144	81.58
ChrD4	0	191	81.40
ChrE1	0	182	79.19
ChrE2	0	131	79.63
ChrE3	0	200	80.06
ChrF1	0	163	80.30
ChrF2	0	153	82.37
ChrX	0	177	59.41
Whole Genome	**0**	**213**	**79.52**

6.4 Discussion and main findings

This chapter aimed to generate the first ancient DNA sequence data from the European wildcat *Felis silvestris silvestris* with specimens sourced from the British Lateglacial - Holocene zooarchaeological record. Ancient DNA was used in conjunction with Accelerator Mass Spectrometry (AMS) radiocarbon ^{14}C dating to investigate phylogeographic origins, population history and patterns of genetic diversity in this subspecies. The failure to generate ancient DNA and to receive a ^{14}C age estimate from the sole putative Lateglacial specimen sampled here, MM098 (Gough's Cave), restricted this study to analyses of Holocene wildcats only. The main findings were:

- New AMS ^{14}C age estimates for four specimens were obtained:
    - MM091, Dunagoil Cave, Isle of Bute: 1.814 – 1,618 cal BP
    - MM092, Pin Hole Cave, Derbyshire: 8,715 – 8,414 cal BP

  - MM093, Robin Hoods Cave, Derbyshire: 7,239 – 6,999 cal BP
  - MM098, Gough's Cave, Cheddar: 10,655 – 10,299 cal BP

- Six mid-to high coverage mitochondrial genomes for specimens MM053, MM076, MM077, MM091, MM092 and MM093 were sequenced and assembled using *in-situ* hybridisation target enrichment, representing the first ancient mitogenomes to be sequenced from *F. s. silvestris*.

- Neighbour Joining and Bayesian MCMC phylogenetic reconstruction incorporating ND5/ND6 sequence data from extant European wildcats did not detect any phylogenetic signal in samples from this study. All three specimens (MM076, MM077 and MM091) grouped within the European wildcat clade.

- All ancient mitogenomes generated here showed fixed nucleotide substitutions at 12 key loci used by Scottish Natural Heritage (SNH) for rapid genetic discrimination between true wildcats and feral hybrids/domestics.

- Analysis of nucleotide variability across mitogenomes indicated an overall pattern of low variability but with significant variation between genes.

- Analyses of pairwise genetic distance between samples indicated greatest divergence between specimens MM092 and MM093 and lowest divergence between MM053 and MM076.

- Bayesian MCMC phylogenetic reconstruction using a 433bp length of the Control Region and incorporating data from extant wildcats strongly supported the placement of all British wildcats in the 'central Europe' clade.

- The first light-coverage (1.9x) whole genome for *F. s. silvestris* was obtained via shotgun sequencing and assembly to the domestic cat *F. catus* genome v8.0.

- A Median Joining (MJ) network and fixation index (F_{st}) indicated a *possible* chronological distinction within the British specimens in to two haplogroups representing the Early Holocene and Mid-Late Holocene

6.4.1 Genetic diversity

In analyses of mtDNA variability within the British wildcat mitochondrial dataset only 38 polymorphic sites were detected out of a total of 16,367 bases (the remaining 275 bases being excluded due to missing data or ambiguous base calls). Concordant low levels of nucleotide diversity and Kimura-2 pairwise genetic distances were also observed. Assuming that this sample size (n = 6) is representative of the British Holocene wildcat population, this is a very low overall level of mitogenomic variability, particularly given that this is a heterochronous dataset spanning *c.* 8,609 years and encompassing the entire geographic area of Britain. The observed low mtDNA genetic diversity observed in Holocene British wildcats may be due to three alternative effects *i*) low study sample number *ii*) founder effects or *ii*) inheritably low mtDNA diversity in this subspecies.

In all of these scenarios, further sampling of ancient and extant wildcats from Britain and Europe is ultimately necessary to confirm or disprove these hypotheses. However, with regard to Holocene recolonisation of Britain, low mtDNA diversity could reasonably, but tentatively, be interpreted as a signature of 'founder effects'. This occurs when the number of individuals in a founding population is small, leading to subsequent generations showing low genetic diversity or by chance showing high or low frequencies of particular alleles due to the effects of drift (founder mutations; Hedrick 2011). While the small sample size obtained here must be kept in mind, the low genetic diversity observed in the Holocene British wildcat population could be the result of a small Early Holocene founder population derived from a single source on the European mainland. *Felis silvestris silvestris* is thought to have low dispersal capabilities (Hartmann *et al.* 2013) and genetic studies have shown population differentiation across geographical barriers such as riverine systems (such as the Rhine in Germany; Hartmann *et al.* 2013). The major riverine systems in NW Europe during the Early Holocene may therefore have represented major biogeographical barriers for wildcats, effectively limiting dispersal into Britain and causing the signature of low mitochondrial variability observed here.

Alternatively, low mitochondrial diversity could be an inherent feature of this species, such as has been observed in the European Lynx (*Lynx lynx*; Hellborg *et al.* 2002) and Iberian lynx (*Lynx pardinus* Rodriguez *et al.* 2011) – although both of these studies used only single gene mtDNA datasets. Without comparable data it is difficult to comment on whether low genetic diversity across the mitochondrial genome is typical for European wildcats. This is further complicated by the difficulty in extrapolating this information from extant wildcat population. Habitat fragmentation and isolation coupled with population reductions across Europe (Stahl & Artois 1991; Nowell & Jackson 1996; Mitchell-Jones *et al.* 1999) can lead to genetic bottlenecks and result in anthropogenically-mediated reduction in genetic diversity. This is turn can lead to genetic partitioning among fragmented populations, which may obscure past phylogeographic events. In extreme cases, such as in the European beaver *Castor fiber*, such scenarios have led authors to interpret genetic partitioning as genuine taxonomic divisions (Durka *et al.* 2005 and references therein). This is turn can impact on conservation measures due to the erroneous allocation of subspecific ranks, confusion over correct designation of Evolutionary Significant Units (ESUs), Management Units (MUs) and associated degree of legal protection. While the wildcat is not yet critically endangered in mainland Europe, increasing anthropogenic threats including habitat fragmentation, isolation and genetic admixture with the domestic cat *F. s. catus* pose serious threats to its conservation (Yamaguchi *et al.* 2015). The European wildcat would therefore benefit from range-wide phylogenetic analyses, ideally using whole mitogenomes and incorporating ancient individuals. This would facilitate a full evaluation of ancient and historic mitochondrial diversity levels, phylogeny and population processes since the LGM.

Comparably low levels of variability across the mitogenome have been reported for other mammals. For example, in a data set of extant brown bears, *Ursus arctos*, from NW Eurasia, whole mitogenomic analyses found 117 polymorphic sites (Keis *et al.* 2013). However, when this dataset is subdivided into haplogroups, which roughly correspond to country boundaries, the number of polymorphic sites falls to 9 (haplogroup D), 12 (haplogroup B & E), 26 (haplogroup A) and 59 (haplogroup C; Keis *et al.* 2013). Low levels of variability have also been observed across the

mitogenome of a carnivore of similar size and ecological niche to the *F. s. silvestris* – the North American fisher *Martes pennanti* (Knaus *et al.* 2011). This species shows only 66 segregating sites across the 16,290 bp length of its mitogenome. Interestingly, both these studies echo what was also observed here; that in species where genetic variability is low, single gene datasets often fail to provide sufficient phylogenetic information to produce a fully-resolved network or phylogeny (Knaus *et al.* 2011; Keis *et al.* 2013).

Utilising the whole mitochondrial genome for intraspecific phylogenetic inference is still relatively rare (although see otter *Lutra lutra* Ki *et al.* 2010; killer whale *Orcinus orca* Morin *et al.* 2010; North American fisher *Martes pennanti* Knaus *et al.* 2011; brown bear *Ursus arctos* Keis *et al.* 2013). Many studies of modern and ancient European mammal species employ single or partial gene sequence data to reconstruct phylogeography (for example, Red fox *Vulpes vulpes* Teacher *et al.* 2011; European lynx *Lynx lynx* Hellborg *et al.* 2002; Iberian lynx *Lynx pardinus* Rodriguez *et al.* 2011; collared lemming *Dicrostonyx torquatus* Brace *et al.* 2012; European beaver *Castor fiber* Horn *et al.* 2014; water vole *Arvicola amphibius* Brace *et al.* 2016). In species where mitochondrial genetic diversity is sufficiently high, single gene datasets may indeed provide sufficient signal to reconstruct intraspecific relationships and phylogeographic history. This study highlights that in species with inherently low levels of mtDNA variability, these loci may be insufficient to reveal true phylogeographic patterns. In these instances, mitogenome-wide analyses may be required, a conclusion also shared with Keis *et al.* (2013). Analyses of whole mitogenome diversity for the Iberian and European lynx (Hellborg *et al.* 2002; Rodriguez *et al.* 2011) and comparison with existing data on single mtDNA gene datasets for these species, alongside further sampling of the European wildcat, would help establish whether low mtDNA diversity is an inherent characteristic of European felids.

Different regions of the mitochondrial genome evolve at different rates and single gene phylogenies essentially recapitulate gene rather than species history. Consequently, different genes and non-coding sub-regions of the mitogenome are often utilised for alternate phylogenetic inferences. For example, cyt*b* and the Control

Region (CR) are often employed in population level analyses due to their relatively high mutation rate, which can resolve topologies in lower levels of a tree but which suffer from site saturation at deeper levels (Avise 2000). Here, the CR was found to harbour the highest levels of variability but still could not produce a resolved network comparable to that achieved with a whole mitogenomic approach (Figs 6.8 & 6.9). Moreover, alternative protein-coding markers ATP6, ND2 and ND4 also showed high haplotype diversity and have the added advantage of not suffering from indels, tandem repeats and homoplasy that reduce the information content of the CR. Knaus *et al.* (2011) found that when CR homoplasy was controlled for, protein coding genes ND4L and ND5 in the brown bear *Ursus arctos* had greater nucleotide and haplotype diversity than the CR. While the short length of ATP6 (675bp) may mean that it has a smaller phylogenetic contribution, ND2 and ND4 may prove to be useful markers in future population level studies of *F. s. silvestris.*

6.4.2 Phylogeny

Based on a relatively short length of the CR (433bp) the British wildcat samples all placed within a well-supported monophyletic clade alongside individuals from west Germany and Luxembourg (Steyer *et al.* 2016). Analysis of microsatellite markers suggest that this group is part of a larger central European clade that includes France and Belgium (Mattucchi *et al.* 2016). With the assumption that mitochondrial lineages show concordance with those elucidated by the STR phylogeny in Mattucchi *et al.* (2016), this suggests that British wildcats are part of a large west-central European matrilineage. The geographic origin of Holocene British wildcats can therefore be confidently identified as the NW coast of France or Belgium and the likeliest glacial refugia for this species are the Iberian or Italian peninsulae. The results here show a likely broad west - eastern European split in *F. s. silvestris,* which is also present in the majority of European mammals, most likely the result of isolation in glacial refugia during the LGM (Hewitt 1999, 2000, 2004). Due to the lack of comparative range-wide mtDNA and the almost complete absence of information on wildcats in the Eastern part of their range it is not possible to identify finer-scale genetic structuring.

A Median Joining network of the mitochondrial genome of the British wildcats revealed the possible existence of two haplogroups (Fig 6.8) with 18 fixed nucleotide differences. These *may* show a north-south geographical split and appear to show divergence between Early Holocene specimens and Mid to Late Holocene specimens. These assertions would need to be confirmed with radiocarbon age estimates for samples MM076 and MM077 and ideally supplemented with further data from more geographically widespread material. Intriguingly this split occurs either side of the 8.2 ka Holocene cold event. The Holocene is often cited as a period of climatic stability. However, sedimentary and ice-core data clearly show this period was punctuated by a series of short, cold, isolations occurring at *c.* 11.4, 11.2, 10.7, 10.4, 10.4, 9.3 and 8.2 cal BP (Bond *et al.* 1997; Klieven *et al.* 2008; Rasmussen *et al.* 2007). Of these the 9.3 and 8.2 cal BP events are strongly expressed in the oxygen isotope and chironomid records from British sites (Lang *et al.* 2010). These abrupt cooling events have been associated with landscape and vegetation change. Prolonged snow-cover and frequent and intense frosts would likely have caused destruction of ground vegetation and long periods of frozen ground (Lord *et al.* 2015), an effect that is likely to have been most severe in northern regions of Britain (Thomas *et al.* 2007).

The 8.2 ka event is the most pronounced of the Holocene cold oscillations (Alley *et al.* 2007). This short cold phase lasted for around 200 years with a central phase characterised by markedly colder temperatures lasting for *c.* 60-70 years; a temperature drop of 6 ± 2°C is recorded at Summit, Greenland (Alley *et al.* 1997; Kobashi *et al.* 2007; Thomas *et al.* 2007). In western Scotland temperature simulations have proposed a mean annual cooling of 1-2°C (Renssen *et al.* 2001). Vegetation responses have been recorded from both Ireland and Scotland where there is a general fall in the number of thermophilous plants and an increase in cool-tolerant taxa, including woodland transition to trees favouring cold conditions such as *Pinus* and *Betula* (Ghilardi & O'Connell 2013; Wicks & Mithen 2014). Loss of woodland in Scotland appears to have preceded the minimum temperatures for this event, being recorded from 8,560 cal BP onwards (Edwards *et al.* 2007. The impacts of these short and abrupt cold phases on mammal populations during the Holocene is little studied and thus evidence for patterns of population change, range contractions and/or local

extirpations is scarce. However, Lord *et al.* (2015) associate the preceding 9.3 ka event with die-off of wild boar in NW England as the consequence of prolonged periods of annual snow cover. Wicks & Mithen (2014) also uncovered archaeological evidence of a drastic reduction in the human Mesolithic population of western Scotland, synchronous with the 8.2 ka event. It is therefore possible that this brief cold phase impacted on British wild cat populations, particularly in northern areas of Britain. This could have led to populations tracking their habitat south or becoming locally extinct prior to recolonisation when the climate stabilized – potentially creating the genetic partitioning observed here.

6.4.3 Conservation

The data generated here has huge utility in wildcat conservation, both in the UK and throughout the European range of *F. s. silvestris* by allowing analyses of phylogeography, demographic history and introgression prior to the introduction of the domestic cat. No evidence of admixture was uncovered in the mitochondrial genome of any of the undated or Late Holocene wildcats in this study. Fixed nucleotide substitutions at 12 loci located in the ND5/ND6 region of ancient British wildcats were in concordance with those used by SNH to differentiate between wild and hybrid feral cats. Ideally nuclear markers should be employed to quantify introgression levels as mitochondrial DNA only provides evidence information transmitted through the maternal line *i.e* one of the 2^n ancestors an individual has *n* generations back (Senn & Ogden 2015). Mitochondrial DNA can also be rapidly replaced during admixture events between closely related wild and domestic species (Whiteford *et al.* 2006; Larson & Burger 2013). Senn & Ogden (2015) report introgression within mtDNA of Scottish individuals, which do not show nuclear signatures of admixture, and suggest either ancestral polymorphism (where the mitochondrial haplotype is not fixed between these species) or a long history of domestic x wildcat hybridisation as an explanation.

It is clear that mtDNA is not an ideal marker for quantification of admixture. However, the wild-living domestic cat population is often monitored using capture –

neuter – release programs (McEwing *et al.* 2012). Licensing conditions for these stipulate swift determination of genetic admixture status so that the welfare considerations of trapped cats are met (McEwing *et al.* 2012). To these ends, real-time PCR is used to amplify swiftly fixed nucleotide differences located in ND5/ND6 mtDNA as it is not possible to genotype and analyse nuclear markers within the time-frame allowed. Reference data for these markers was taken from wildcats that themselves may have had undetected, genomic introgression. The results obtained here showing that all ancient, and thus undoubtedly pure, British wildcats also show these fixed base substitutions. The results strongly support this test as a means of differentiating between wild, hybrid and domestic cats with respect to this particular mtDNA locus.

As closely-related subspecies, large portions of the genomes of the wild and domestic cat will be virtually indistinguishable from each other (Chambers & MacAvoy 1999; Senn & Ogden 2015). Furthermore, as introgression progresses via the process of backcrossing, the proportion of the genome that has introgressed in an individual reduces by around 50 percent per generation (although there is a large amount of variation; Chambers & MacAvoy 1999; Senn & Ogden 2015). The combined effects of both factors mean that distant genetic introgression between subspecies is very difficult to detect (Baack & Rueseberg 2008). In such scenarios, large numbers of genome-wide nuclear markers may be required to allow discrimination between taxa (Baack & Rueseberg 2008). This study assembled the first, draft, low-coverage (1.9x), whole ancient genome from a Romano-British age wildcat excavated from Dunagoil Cave, Isle of Bute. While, in isolation, this genome has little utility in allowing the detection of novel polymorphic SNP variants, which allow admixture detection between pure-bred wildcat and hybrids, it represents a significant first step in sequencing whole ancient genomes from *F. s. silvestris*. Future sequencing of modern and ancient wildcat genomes and development of genome-wide wildcat SNP arrays, however, will allow detection of such variants. At such a time, ancient wildcat genomes could play a key role in unravelling the timing, progression and consequences of hybridisation between *F. s. silvestris* and *F. s. catus*. Moreover, further

work with this data will potentially allow investigation of a number of population genetics metrics such as levels of inbreeding and population size.

Chapter 7. Discussion

7.1 Introduction

The following chapter will give a summary of each data chapter, after which the aims set out in Chapter 1, Section 1.2 are addressed. This is followed by a brief summary of the thesis and suggestions for future research are made, both on the thesis topic and the relevance of the data obtained here in other areas of research.

7.2 Summary of the data chapters

7.2.1. Chapter 3.

This chapter investigated the use of 2D geometric morphometric techniques in identifying shape and size variables involved in discriminating between common voles, *Microtus agrestis,* and field voles, *Microtus arvalis,* from the lower first molar (M_1). The M_1 is a commonly recovered dental element from these species but morphological similarity, coupled with the partially sympatric distributions of *M. arvalis* and *M. agrestis* during the Late and Postglacial periods, make species discrimination from this tooth extremely difficult. Accurate species discrimination for these voles is essential if the individual species histories are to be investigated. The utility of the shape space variables generated here (constructed by principal components analysis, PCA, and canonical variates analysis, CVA) was examined for its utility in identifying ancient *Microtus* spp. of unknown identity. Shape changes in the M_1, as represented by 2D landmark and semi-landmark configurations, were modelled across the PCA axes. Ancient DNA was used as a second line of evidence. This utilised the cytochrome *b* locus in conjunction with BLAST searches and phylogenetic reconstruction methods to investigate *i*) the taxonomic signal in this locus and *ii*) the accuracy of species identification from morphology.

Results showed that *M. agrestis* and *M. arvalis* can be accurately separated on the basis of M_1 dental morphology alone, with CVA providing the highest degree of between-group separation. Strobe models identified areas of shape change associated with species discrimination with the 2D landmarks shown to be as accurate in between-

group separation as 2D semi-landmarks (which were used in conjunction with extended eigenshape techniques). However, the latter provided more information on gross morphological shape change in the enamel boundary of the tooth. Ancient specimens were accurately and consistently assigned to a species group within the CVA space with > 95% accuracy and this result was supported via ancient DNA. These methods also allowed the identification of the common vole, *M. arvalis*, in the British Holocene faunal record for the first time.

7.2.2. Chapter 4.

This chapter used lower first molars (M_1's) from the common vole, *M. arvalis*, excavated through a stratigraphic sequence at a single site (Gully Cave, Somerset) to examine whether this species shows population continuity or discontinuity throughout the climatic intervals comprising the Lateglacial Interstadial, Younger Dryas and early Holocene. Methods employed were 3D geometric morphometrics and ancient DNA focusing on the cytochrome *b* locus. Three dimensional models of the tooth were created via microCT before 3D landmarks were collected on the CT models to create a 3D configuration. These were used as shape and size proxies for the tooth. Shape co-ordinates were analysed via discriminant function analyses with a novel data augmentation technique applied to investigate the effect of sample size on results. Cytochrome *b* was sequenced from DNA extracted from the tooth and these sequences subjected to phylogenetic and population genetic analyses to explore the genetic relationship of individuals throughout the climatic groups.

Morphometric results showed a moderately positive correlation between M_1 shape change and the mean NGRIP temperature proxy, $\delta^{18}O$ (Svensson *et al.* 2006, 2008). Results from PCA decomposition and between-group CVA analysis demonstrated that shape change between these climatic groupings were subtle but generally well-supported. Augmentation of the sample size increased the distinctiveness of between-group support inferring that small sample size was a factor in lowering statistical support values. While the morphometric results suggested climate-correlated M_1 shape change, they could not discriminate between scenarios of, *i*) a single population persisting throughout the various climatic phases and showing morphological M_1

shape change in relation to climate, *ii*) three distinct populations representing the Lateglacial Interstadial, Younger Dryas and Holocene, which show turnover events in relation to climate and, *iii*) a two-group scenario where there is turnover between the early phase of the Lateglacial Interstadial (Bølling) and the Younger Dryas but continuity between the Younger Dryas and Holocene. Ancient DNA analyses indicated that the early Lateglacial Interstadial samples were divergent from the later Younger Dryas and Holocene samples and showed greater population level subdivision. While this result was somewhat ambiguous, it suggests a turnover event in the early phase of the Lateglacial (Bølling, c, 14,000 – 15,000 cal BP), possibly relating to the onset of the Older Dryas stadial which commenced *c.* 15,000 cal BP. Common voles have a high level of phylogeographic substructure in Europe with three lineages in close geographical proximity to the south coast of Britain; Western-South, Western-North and Central (Stojak *et al.* 2015). These are located from the north-west coast of Spain to the north coast of Germany and all represented potential source populations for the Lateglacial and Holocene population/s of Britain. This study identified, for the first time, that British common voles belong to the Western-North lineage and thus have a geographical origin on the NW coast of Europe (*i.e.,* France and Belgium).

7.2.3. Chapter 5.

This chapter examined phylogeographic origins, population turnover and levels of genetic diversity in the Eurasian beaver, *Castor fiber,* from the British Late Pleistocene into the Holocene. Ancient DNA and accelerator mass spectrometry (AMS) ^{14}C radiocarbon dating were the principal means of investigation with bone and tooth samples sourced from a range of cave and surface sites throughout the UK. No successful aDNA extractions from the Lateglacial Interstadial were obtained, leaving a dataset that represented the Younger Dryas and Holocene climatic phases. Ancient DNA analyses focused on the tRNAPro and Hypervariable-I (HV-1) regions of the mitochondrial genome and utilised existing ancient and modern *C. fiber* sequences in phylogenetic reconstruction and population genetic analyses.

Radiocarbon dating identified at least one, and most likely two, individuals that persisted in the SW of England during the period of glacial re-advance associated with the Younger Dryas. This result suggests that the SW of England may have acted as a cryptic northern refuge for temperate adapted species during the Younger Dryas and also that *C. fiber* has some degree of resilience to climatic changes and sub-optimal cold environments. British beavers from the Late Pleistocene (Younger Dryas) and Early Holocene grouped alongside other western European individuals to form a large, panmictic, pan-Western European clade with no internal substructure. Divergence estimates for this 'Western' clade, alongside the Tajima's D result, suggest a refugial area in the Mediterranean peninsulas during the LGM followed by a rapid population expansion associated with climatic amelioration. The lack of substructure within the phylogeny suggests that the ancient British beavers are not particularly more closely related to one extant European population over another. This is a timely result as wider re-introductions of beavers into Britain has been extensively discussed and debated both by conservation bodies and in the literature (Halley 2011; Frosch *et al.* 2014; Senn *et al.* 2014). The question of which European population from which to source individuals for this purpose has been a central question in this debate. Ideally, a source population should be both genetically diverse so as to maximise the genetic adaptive potential of translocated individuals to the new environment and show a close phylogenetic affinity to the extirpated population (IUCN 2013). The results obtained here show that, given such close phylogenetic relatedness between all modern and ancient western European beavers to the ancient British beaver population (at the tRNAPro/HV-I CR region), phylogenetic relatedness need not be a chief concern in conservation planning.

7.2.4. Chapter 6.

This chapter investigated levels of genetic diversity in British Holocene populations of the European wildcat, *Felis silvestris silvestris,* using ancient DNA analysis in conjunction with in-solution target enrichment and Next Generation Sequencing (NGS). Network analysis and population genetics statistics were employed to investigate genetic diversity both throughout the mitogenome and within individual

genes, to quantify their relative contribution to determining phylogenetic and population substructure and their utility in assessing levels of genetic introgression with the domestic cat, *F. s. catus*. A small, 433bp, region of the CR was used alongside existing European data for this subspecies to identify a broad European lineage for the British wildcats, in addition to inferring the geographical source region for this population. Repeat library builds and NGS sequencing were used to assemble the first draft *F. s. silvestris* genome.

Six whole mitogenomes, which ranged from 9.6 to 108 x-fold coverage and spanned the Early to Late Holocene, were generated in this study. Levels of genetic diversity across the mitogenome were very low and network constructions showed that individual genes did not have sufficient resolution to identify population-level structure in the dataset. Whole mitogenome networks revealed two haplogroups, with 18 fixed substitutions between them, which may correspond to a temporal division between Early and Mid + Late Holocene wildcats. Intriguingly, this appears to be associated with the 8.2 ka cold event. Evidence exists for vegetation decline at northern latitudes in both Scotland and Ireland which are associated with the 8.2 ka event (Edwards *et al.* 2007; Ghilardi & O'Connell 2013; Wicks & Mithen 2014). While the effects of this cold event on fauna are, as yet, unknown they may have included a reduction in habitat availability and/or, reduced availability of prey species. These factors are hypothesised here to have initiated a short phase of population decline, southern habitat tracking or population replacement in this species. There is evidence which associates the 8.2 ka event with dramatic population reductions in Mesolithic humans (Wicks & Mithern 2014) and which links the preceding 9.3 ka event with die-offs of wild boar (Lord *et al.* 2015). However, this is the first time in the British Holocene subfossil mammal assemblage that the 8.2 ka event has been demonstrated to have had such a strong impact on a non-human mammal population. Further supporting evidence in the form of additional ^{14}C dates and increased sample size is needed to confirm these findings.

British wildcats grouped within a large west-central European clade, suggesting the NW coast of Europe as the most likely source area for Holocene colonisation of

Britain. The paucity of comparative mitochondrial sequence data for this species prevented any further examination of European phylogenetic substructure (which almost certainly exists) in wildcats. All ancient British wildcats possessed fixed nucleotide substitutions in a region of the mitogenome used by conservation agencies for wildcat/hybrid discrimination, strongly suggesting that these are robust discriminative markers. Lastly, this study succeeded in assembling the first draft genome of the European wildcat, at 1.9 x-fold coverage. This represents a significant first step towards initiating whole-genome research in ancient wildcats, with the potential to uncover the demographic history of the subspecies as well as the initiation, pattern and process of wildcat x domestic hybridisation.

7.3 Thesis Aims

7.3.1 **Aim 1.** Can the geographical source area for Late Pleistocene and Holocene British populations of *M. arvalis*, *C. fiber* and *F. silvestris* be identified and can this be used to infer the most likely glacial refugium?

The study incorporated species of differing size, dispersal capabilities and ecological requirements. This was in order to investigate whether common, climate-related, patterns of population turnover and phylogenetic history could be inferred in relation to the abrupt climatic changes observed over the Late Pleistocene – Holocene boundary. While individualistic responses to climate change have been suggested for many mammals at both the phylogeographic and population level (Stewart 2008, Stewart *et al.* 2010), the genetic structure of European mammals show a number of common divisions. Most temperate species show a pattern of genetic subdivision into broad lineages associated with isolation in southern refugia during the LGM (Hewitt 1999, 2004). The geographical source location for all species studied here can confidently be identified as the NW coast of Europe. This was demonstrated by the high support in the phylogenetic analyses for the placement of all study species within lineages known to represent these regions. Furthermore, this region was the source population for both Lateglacial and Holocene British populations and suggests refugial areas consistent with isolation in either the Italian or Iberian peninsulae.

Known phylogeographic structure for the study species differed. The common vole has five main European lineages (Fink *et al.* 2004; Jaarola *et al.* 2004; Tougard *et al.* 2008; Bužan *et al.* 2010; Stojak *et al.* 2015) with two that could reasonably be identified as potential source populations, *i*) the Western-South lineage which has a post-LGM origin in the Iberian Peninsula and *ii*) the Western-North lineage representing the NW coast of France and Belgium and which likely originates from a glacial refuge in the Italian Peninsula. The absence of evidence of individuals from the Western South lineage in Britain suggests that the Pyrenees acted as a significant geographical barrier, allowing colonisation of Britain by the more proximal Western North lineage. This species is known to have exploited cryptic northern refugia (Tougard *et al.* 2008) but there is no evidence here to suggest a phylogenetic affinity between the British common voles and known lineages derived from these. However, it must be kept in mind that this dataset was based on a single cave site situated in a region which likely represents a cryptic northern refuge. While some authors have successfully made inferences about Late Pleistocene and Holocene population events from a low number of study sites (for instance Brace *et al.* 2016, where a composite of just two sites are utilised), more complex recolonisation patterns than are observed here (such as multiple recolonisation events) may have occurred.

The Eurasian beaver, *Castor fiber*, shows remarkably little phylogenetic substructure within the Western clade. This is likely to be related to its apparent ability to rapidly recolonise suitable habitat (Hartman 1996; Nolet & Baveco 1996) and its robustness to the genetic effects of genetic bottlenecks (see evidence presented in Chapter 5). The British beavers all grouped within this large Western clade within which no genetic subdivision was observed between British individuals and beavers from other regions. Based on this data, they are likely to have retreated to either the Iberian or Italian peninsulae during the LGM before a rapid population expansion occurred as the climate warmed. Dispersal and recolonisation of Europe and Britain by *C. fiber* during the Holocene is likely to have been facilitated by the abundance of large riverine systems, which effectively acted as dispersal corridors for this species.

This scenario contrasts sharply with the genetic evidence obtained for the European wildcat, a species with a similar body size to *C. fiber* and which is also associated with temperate woodland habitats. While virtually nothing is known regarding the mitochondrial lineages of this species in Europe, evidence presented in this study shows that this species also belongs to a western European matrilineage. However, while *C. fiber* shows a large degree of genetic diversity in the ancient British Holocene population, *F. s. silvestris* is characterised by a distinct lack of genetic diversity – even when measured across the entire mitogenome. This may be related to the differing demographic and dispersal potentials of these species. The Eurasian beaver is a semi-aquatic herbivore that can exist in large population densities (when conditions allow; Campbell *et al.* 2005), has a higher generational turnover then *F. silvestris* (Müller-Schwarze & Sun 2003*),* and can utilise riverine systems for dispersal (Hartman 1996; Nolet & Baveco 1996). In contrast, the European wildcat is a solitary carnivore, which is known to be a poor disperser and shows population subdivision over geographical barriers such as large river systems (Hartmann *et al.* 2013). Therefore, while the evidence for *C. fiber* suggests that it rapidly colonised Europe and Britain during the Late Pleistocene and Early Holocene, results for the European wildcat suggest that this species may have had a much slower rate of recolonisation and a smaller founder population.

7.3.2 **Aim 2.** Can patterns of population continuity or population turnover for *M. arvalis*, *C. fiber* and *F. silvestris* over the Lateglacial Interstadial – Younger Dryas - Holocene climatic periods be detected?

The sole dataset with specimens representing the full span of the Lateglacial Interstadial – Younger Dryas – Holocene intervals was for the common vole. This dataset suggested that there may have been a turnover event in the Early Lateglacial (Bølling phase). The Lateglacial was not a period of uniformly warm temperatures with two abrupt phases of cooling observed before and after the Bølling - the Oldest and Older Dryas, corresponding to GI-1d and GI-1d respectively (Svensson *et al.* 2006, 2008; see Chapter 1, Fig 1.4). As yet, there have been no studies examining the effects of Lateglacial climatic oscillations on mammal populations (bar this one).

However, there are strong reasons to suspect that this may have been a tumultuous period for mammal species recolonising Britain from mainland Europe. Temperate mammal species had emerged from LGM glacial refugia only relatively recently and populations may not have been well established. In addition, this would have been a period where new faunal communities were coming together, with species coming into contact with one another that may have been formerly been separated into glacial refugia. Further research into the effects of the Lateglacial climatic oscillations on mammal populations would be most informative in this respect.

Population continuity over the Younger Dryas and into the Holocene is suggested for both common voles and Eurasian beavers on the basis of both genetic evidence and their presence during these intervals in cave sites situated in the SW of England (Gully Cave and Gough's Cave, Somerset). This leads to the conclusion that the SW of England may have acted as a cryptic micro-refugium for temperate mammals during the Younger Dryas cold phase, from which they could expand their range and populations during the Holocene warming. Pollen records show elevated levels of conifer tree species (*Picea, Pinus and Abies*) and a low but persistent presence of broadleaf tree species in the SW of England during the LGM (Kelly *et al.* 2010). Consequently, the authors of this study have suggested this region as a cryptic northern refugium for these floral species. No evidence is currently available to support the presence of conifer and broadleaved tree communities at either of these sites during the Younger Dryas. However, if these tree species were able to survive in this region at the height of the last glaciation, it is entirely possible that sheltered regions in SW England could have supported a habitat that was partially forested during the shorter period of the Younger Dryas. The dataset for the European wildcat, *F. s. silvestris,* contained individuals from the Holocene Interglacial only. Despite the small sample size and limited temporal spread, this species showed a divergence between Early Holocene specimens and those dating to the Mid-Late Holocene. This divide coincides with the 8.2 ka cold event and suggests that populations of this species, which is known to be a poor disperser, may be more vulnerable to the effects of abrupt climate change than either the common vole or European beaver.

Ancient DNA studies have consistently shown that in many mammal species, ancient populations harboured far greater levels of genetic diversity than are observed in extant populations (*e.g* Hofrieter & Barnes 2010). In the tRNAPro/CR data generated in this study for the Eurasian beaver, this species showed a pattern of high genetic diversity in the ancient Late Pleistocene and Early Holocene populations, followed by diversity depletion in modern populations. The loss of a large proportion of the genepool in this species is undoubtedly the result of anthropogenic-related activity (Lavrov 1983; Nolet & Rosell 1998; Horn *et al.* 2014). This, unfortunately, makes it extremely difficult to detect whether loss in diversity at this locus can be attributed to natural processes, such as population responses to climate change. The Eurasian wildcat showed very low genetic diversity across the mitogenome in the British Holocene individuals. However, the absence of Late Pleistocene and modern comparative data means that is impossible to come to any clear conclusions regarding trends in genetic diversity levels over time. Further sampling of Late Pleistocene wildcats would be required to allow a full evaluation of this.

7.3.3 **Aim 3**. Does morphological evidence exist for eco-phenotypic adaptation to abrupt changes in climate and/or population turnover events?

Eco-morphological analyses were only performed for the common vole as this was the only species for which a dataset possessing both excellent morphological preservation and a sufficiently large sample size over the study intervals could be generated. The results obtained here were somewhat ambiguous but clearly indicated that climate-correlated shape change in the *M. arvalis* M_1 was present and was not an artefact of allometric effects (as no significant size variation was detected). The analyses supported both a two group scenario, where a turnover event occurred in the early part of the Lateglacial, and a three-group scenario of three distinct populations over the Lateglacial Interstadial – Younger Dryas – Holocene. However, these results do not rule out a scenario of a single, continuous, population showing morphological adaptation in M_1 shape in response to climate change. This component of the study also allowed an evaluation of the application of both 2D and 3D geometric morphometric techniques to zooarchaeological material dating as far

back as the Lateglacial. Low sample size and poor condition of material is a pervasive problem in this field. However, when these factors can be overcome, this investigation demonstrated that these techniques are powerful tools both to discriminate between closely related species, quantify changes in form and relate such change to eco-morphological correlates.

7.4 Additional Findings

During the course of these investigations, a number of additional research findings of interest were uncovered. The first record of the common vole in the British Holocene was identified by morphological analyses and further supported by ancient DNA analyses. Up until this point it was assumed that this species had not successfully recolonised Britain after the Younger Dryas (Corbet 1961, 1986; Yalden 1999). In addition, conservation implications for the Eurasian beaver, *Castor fiber,* on the basis of the phylogenetic results could also be suggested – namely that British reintroduction programs need not focus on identifying the closest phylogenetically related European population. The identification of two temperate species, common vole *M. arvalis*, and Eurasian beaver *C. fiber*, in Younger Dryas faunal assemblages at two cave sites in SW England strongly suggest that this region acted as a cryptic northern refugium for temperate adapted mammals. Ancient fixed nucleotide differences in the mitochondrial genome of the European wildcat, *F. s. silvestris,* provided a means of validating the introgression test currently in use by conservation bodies in Scotland. The superiority of using whole mitogenomes, rather than single mtDNA markers, for species with inherent low genetic diversity was demonstrated.

7.5 Summary

This investigation demonstrated the huge potential of both ancient DNA and geometric morphometrics in the investigation of ecomorphological, phylogenetic and population level studies of British mammals dating from the Lateglacial Interstadial to the Holocene. The difficulties associated with small sample size, successful extraction and sequencing of ancient DNA and obtaining material in good morphological condition did not allow for the spatio-temporal sampling distribution that was desired at the outset of the project. Nevertheless, this investigation succeeded

in reconstructing ancient phylogenies for all study species and was able to generate new insights into the histories of these species in Britain.

A common geographical origin was identified for all British study species as the NW cost of Europe and likely the NW coast of France and Belgium – the closest geographical regions to Britain. Differential patterns of genetic diversity and population-level responses to abrupt climate change could be broadly related to the dispersal capabilities and ecological preferences of individual species. European beavers and common voles show continuity over the Younger Dryas cold phase, suggesting a certain level of tolerance to sub-optimal environmental conditions. In contrast, European wildcats appear to be more sensitive to the effects of climate change. This species showed fine scale population divergence observed over the relatively short 8.2 ka Holocene cold interval. Taken together, these results demonstrate that while the study species shared a common geographical origin, species-specific dispersal capabilities and ecological preferences influence their phylogenetic structure, levels of mitochondrial genetic diversity and their ability to tolerate changes in abrupt climatic fluctuations.

7.6 Future work

Further sampling over a wider geographical and chronological range would allow the results obtained here to be expanded upon. For the common vole, *M. arvalis*, it would be particularly interesting to see if this species shows the 'Celtic Fringe' pattern of haplotype distribution (Piertney *et al.* 2005; Searle *et al.* 2009) observed for other small mammals (*e.g.,* bank vole *Myodes glareolus*, field vole *Microtus agrestis*, pygmy shrew *Sorex minutus* and water vole *Arvicola amphibius*; Searle *et al.* 2009; Brace *et al.* 2016). In order to accomplish this, samples from more northern locales and, in particular, Scottish sites would be necessary. This study recorded the common vole in the British Holocene record for the first time. The early Holocene population of this species at Gully Cave may have been a small, remnant population that did not expand its range beyond the south-west of England. Alternatively, it may have expanded its range and size in synchrony with Holocene warming. What *is* known is that the species

subsequently became extirpated at some, unknown, point in the Holocene (Corbet 1961, 1986; Yalden 1999), although the factors behind this could only be speculated on here. The 2D geometric morphometric data produced in this study can provide an accurate, non-destructive, and relatively swift means of identifying this species from Holocene deposits based on the M_1 alone. Further research could utilise this method to identify the presence *M. arvalis* at British Holocene sites, which would allow the selection of individuals for ancient DNA analyses. Unravelling the fate of the common vole in Britain, its relationship to the introduced Orkney population and the factors behind its eventual extirpation would be an intriguing prospect.

For the Eurasian beaver, *C. fiber*, additional sampling should focus on acquiring Lateglacial material and expanding the sampling distribution of individuals from all climatic periods to include northern England and Scotland. This again would allow an evaluation of the 'Celtic Fringe' haplotype distribution, which has not yet been examined in medium-sized mammals. The relationship of the Holocene and Younger Dryas beaver populations to the Lateglacial Interstadial population could not be established in this study. It would be fascinating to uncover whether the Lateglacial Interstadial population was also part of the large Western European clade and also whether *C. fiber* exhibits any population responses to climate change over the Lateglacial climatic fluctuations, as seen with the common vole.

For the European wildcat, *F. s. silvestris*, it would be of particular interest to obtain DNA of Lateglacial Interstadial and Younger Dryas age. The results for the Holocene data suggest this species shows a marked response, even to (relatively) short cold events. It may therefore be hypothesised that wild cat populations display clear responses to more abrupt periods of cooling. Pre-Holocene material for this species in Britain is difficult to find and not every specimen will yield sufficient quality and quantity of ancient DNA for downstream analyses. However, the extraction and in-solution target enrichment methods used in this study demonstrated high efficacy at optimising ancient DNA yields and enriching DNA libraries for whole mitochondrial genomes. Therefore, it is proposed that these methods be employed for future whole

mitogenome research on ancient British *F. s. silvestris* so as to maximise the likelihood of successful ancient DNA retrieval.

The theory, practical applications and analytical techniques in geometric morphometrics are already very well developed for application to ancient material. Principal challenges instead relate largely to accumulating ancient material in sufficient numbers and of good enough morphological quality. This study utilised teeth as an eco-morphological proxy for population response to abrupt climate change. The drawback with such a proxy is that when form variation is subtle, it is difficult to make interpretations about the functional relevance of any observed shape changes. Variation in tooth shape may be related to a host of developmental, genetic and environmental influences (Teaford 2007). Targeting morphological studies at skeletal elements that can clearly be shown to have a functional response to climate change in size or shape (for example as was observed in wild horse *E. ferus* metapodia, Bignon *et al.* 2005 and arctic fox metapodia, Dalén *et al.* 2007) would be an interesting addition to a study such as this.

Mitochondrial DNA has been the marker of choice in ancient DNA phylogenetic and population level studies for some time, largely due to its much higher cellular copy number (Willerslev & Cooper 2005). Ancient DNA studies that focus on using whole genomes, single nuclear genes and markers such as Single Nucleotide Polymorphisms (SNPs) in non-human species are becoming more commonplace but are still considered relatively rare (although for example see for canids, Skoglund *et al.* 2015; Frantz *et al.* 2016, equids, Pruvost *et al.* 2011; Orlando *et al.* 2013; Shubert *et al.* 2014; moa, Huynen *et al.* 2003; mammoths, Römpler *et al.* 2006; Palkopoulou *et al.* 2015; Caribbean mammals, Woods *in press*). However, the field of ancient DNA is rapidly developing. Methodological, technological and bioinformatics techniques have progressed enormously, even over the duration of this project. It is likely that large, genome-wide NGS datasets will become increasingly cost effective to assemble, thus allowing these studies to increase in number and to extend their focus to a larger diversity of species.

Inclusion of genome-wide data, nuclear genes and SNPs would enhance a study such as this in a number of ways. First, single nuclear genes have the potential to highlight the genetic basis for adaptation to climatic change (Franks & Hoffman 2012) and also provide a means of calibrating mitochondrial phylogenies which suffer from low resolution. In addition, they can be employed to gather information on phenotypic traits such as coat colour in ancient populations, which has been inferred from aDNA in dogs (Olivier *et al.* 2013) and horses (Pruvost *et al.* 2011). While the mtDNA markers used in this study were sufficient for the purposes of the genealogical and phylogenetic analyses employed, there are undoubtedly undiscovered population-level processes that have not been detected. Incorporation of SNPs and nuclear genes would allow a finer degree of population structure and dynamics to emerge and potentially answer some of the remaining questions regarding mammalian Late and Postglacial recolonisation of Britain.

References

Abramson, N. I., Lebedev, V. S., Tesakov, A., Bannikova, A. A. (2009). Supraspecies relationships in the subfamily Arvicolinae (Rodentia, Cricetidae): an unexpected result of nuclear gene analysis. *Molekulyarnaya Biologiya* **43**:879-909

Adams, D. C., Rohlf, F. J., Slice, D. E. (2004). Geometric morphometrics: ten years of progress following the "revolution". *Italian Journal of Zoology* **71**: 5–16

Adams, D. C., Rohlf, F. J., Slice, D. E. (2013). A field comes of age: geometric morphometrics in the 21st century. *Hystrix, the Italian Journal of Mammalogy* **24**: 1–8.

Adler, C. J., Haak, W., Donion, D., Cooper, A. (2011). Survival and recovery of DNA from ancient teeth and bones. *Journal of Archaeological Science* **38**: 956-965

Alley, R. B., Mayewski, P. A., Sowers, T., Stuiver, M., Taylor, K. C., Clark, P. U. (1997). Holocene climatic instability: a prominent event 8200 years ago. *Geology* **25**: 483-486

Alroy, J. (2001). A multispecies overkill simulation of the end-Pleistocene megafaunal mass extinction. *Science* **292:** 1893-1896

Altschul. S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410

Altuna, J. (1967). Fauna de mamíferos del yacimiento pre-histórico de Marizulo (Urnieta, Guipúzcoa). *Munibe* **19**: 271-298

Amato, G., DeSalle, R., Ryder, O. A., Risenbaum, H. C. *(Eds.)* (2009). *Conservation Genetics in the Age of Genomics*. Columbia University Press: New York, Chichester

Anderson, T. M., vonHoldt, B. M., Candille, S. I., Musiani, M., Grecco, C., Stahler, D. R., Smith, D. W., Padhukasahasram, B., Randi, E., Leonard, J. A., *et al.* (2009). Molecular and evolutionary history of melanism in North American gray wolves. *Science* **323**: 1339-1343

Antunes, A., Pontius, J., Ramos, M. J., O'Brian, S. J., Johnson, W. E. (2007). Mitochondrial introgression into the nuclear genome of the domestic cat. *Journal of Heredity* **98**: 141-420

Atkinson, T. C., Briffa, K. R., Coope, G. R. (1987). Seasonal temperatures in Britain during the past 22, 000 years, reconstructed using beetle remains. *Nature* **325**: 587-92

Avise, J. C. (2000). Phylogeography. *The History and Formation of Species*. Harvard University Press: Cambridge, Massachusetts

Baack, E. J and Rieseberg, L. H. (2007). A genomic view of introgression and hybrid speciation. *Current Opinions in Genetics and Development* **17**: 513-518

Balharry, D and Daniels, M. J. (1998). Wild living cats in Scotland. Scottish *Natural Heritage Research, Survey and Monitoring Report No. 23*. Perth and Edinburgh: Scottish Natural Heritage

Ballantyne, C. K. (2010). Extent and de-glacial chronology of the last British-Irish Ice Sheet: implications of exposure dating using cosmogenic isotopes. *Journal of Quaternary Science* **17**: 1149-1184

Ballard, J. W. O and Whitlock, M. C. (2003). The incomplete natural history of mitochondrial DNA. *Molecular Ecology* **13**: 729-744

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

Barbosa, S., Pauperio, J., Searle, J. B., Alves, P. C. (2012). Genetic identification of Iberian rodent species using both mitochondrial and nuclear loci: application to noninvasive sampling. *Molecular Ecology Resources* **13**: 43-56

Barčiová, L. (2009). Advances in insectivore and rodent systematics due to geometric morphometrics. *Mammal Review* **39**: 80-91

Barnes, I & Young, J. P. W. (2000). DNA-based identification of Goose species from two archaeological sites in Lincolnshire. *Journal of Archaeological Science* **27**: 91-100

Barnes, I., Matheus, P., Shapiri, B., Jensen, D., Cooper A. (2002). Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science* **195**: 2267-2270

Barnes, I., Shapiro, B., Lister, A., Kuznetsova, T., Sher, A., Guthrie, D., Thomas, M. G. (2007). Genetic structure and extinction of the Woolly Mammoth, *Mammuthus primigenius*. *Current Biology* **17**: 1072-1075

Barnett, R., Shapiro, B., Barnes, I., *et al.* (2011). Phylogeography of lions (*Panthera leo ssp*). Reveals three distinct taxa and a late Pleistocene reduction in genetic diversity. *Molecular Ecology* **18**: 1668-1677

Barnosky, A. D. & Bell, C. J. (2003). Evolution, climatic change and species boundaries: perspectives from tracing *Lemmiscus curtatus* populations through time and space. *Proceedings of the Royal Society B* **270**: DOI: 10.1098

Barnosky, A. D., Koch, P. L., Feranec, R. S., Wing, S. L., Shabel, A. B. (2004) Assessing the causes of Late Pleistocene Extinctions on the Continents. *Science* **306**: 70-75

Batson, W. G., Gordon, I. J., Fletcher, D. B., Manning, A. D. (2015). Translocation tactics: a framework to support the IUCN Guideline for wildlife translocations and improve the quality of applied methods. *Journal of Applied Ecology* **52**: 1598-1607

Beaumont, M., Barratt, E. M., Gottelli, D., Kitchener, A. C., Daniels, M. J., Pritchard, J. K., Bruford, M. W. (2001). Genetic diversity and introgression in the Scottish wildcat. *Molecular Ecology* **10**: 319-336

Bennett, K. D. (1990*)*. Milankovitch cycles and their effects on species in ecological and evolutionary time. *Palaeobiology* **16***:* 11–21

Bernt. M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., Stadler, P. F. (2013). MITOS: Improved de novo Metazoan Mitochondrial Genome Annotation. *Molecular Phylogenetics and Evolution* **69**:313-319

Biedrzycka, A., Konior, M., Babik, W., Świslocka, M., Ratkiewicz, M. (2014). Admixture of two phylogeographic lineages of the Eurasian beaver in Poland. *Mammalian Biology* **79**: 287-296

Bignon, O., Baylac, M., Vigne, J-D, Eisenmann, V. (2005). Geometric morphometrics and the population diversity of Late Glacial horses in Western Europe (*Equus caballus arcelini*): phylogeographic and anthropological implications. *Journal of Archaeological Science* **32**: 375-391

Björk, S. (1995). A review of the history of the Baltic Sea, 13.0-8.0 ka BP. *Quaternary International* **27**: 19-40

Björk, S., Walker, Micheal, J. C., Cwynar, L., C., Johnsen, S., Knudsen, K-L., Lowe, J. J., Wohlfarth, B. (1998). An event stratigraphy for the Last Termination in the North Atlantic region based on the Greenland ice-core record: A proposal by the INTIMATE group. *Journal of Quaternary Science* **13:** 283–292

Blois, J. L., McGuire, J. L., Hadly, E. A. (2010). Small mammal diversity loss in response to late-Pleistocene climatic change. *Nature* **465**: 771-774

Blow, N. (2008). Metagenomics: benchmarks and standards. *Nature* **453**: 687-690#

Bond, G., Showers, W., Chesby, M. *et al*. (1997). A pervasive millennial-scale cycle in North Atlantic Holocene and glacial climates. *Science* **278**: 1257-1266

Bookstein, F. L. (1991). *Morphometric tools for landmark data: geometry and biology*. Cambridge University Press: New York.

Bookstein, F. L. (1996). Biometrics, biomathematics and the morphometric synthesis. *Bulletin of Mathematical Biology* **58**: 313–365

Bookstein, F. L. (1997). Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Medical Image Analysis* **1**: 97-118

Boscato, P and Palma di Cesnola, A. (2000). Nuovi ritrovamenti di epigravettiano antico 'iniziale' a Grotta Pagliacci (Rignano Garganico)

Brace, S., Palkopoulou, E., Dalén, L., Lister, A. M., Miller, R., Otte, Marcel., Germonpré, M., Blockely, S. P. E., Stewart, J. R., Barnes, I. (2012). Serial population extinctions in a small mammal indicate Late Pleistocene ecosystem instability. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 20532-20536

Brace, S., Ruddy, M., Miller, R., Schreve, D. C., Stewart, J. R., Barnes, I. (2016). The colonization of the water vole (*Arvicola amphibious* (Linnaeus, 175)): origins and developments of the Celtic fringe. *Proceedings of the Royal Society* **B** 283: DOI: 10.1098

Brentjes, B. (1968). Die Biberbilder vom Onegasee. *Zeitschrift für Jagdwissenschaft* **14**: 180-181

Bromham, L. (2009). Why do species vary in their rate of molecular evolution? *Biology Letters* **5**: DOI: 10.1098

Bromham, L., Rambaut A., Harvey, P. H. (1996). Determinants of rate variation in mammalian DNA sequence evolution. *Journal of Molecular Evolution* **43**: 610–621

Bronk-Ramsey., Scott, M., van der Plicht, Hans. (2013). Calibration for archaeological and environmental terrestrial samples in the time range 26-50 ka cal BP. *Radiocarbon* **55**: 2021-2027

Brown, T and Brown, T. (2011) *Biomolecular archaeology*. John Wiley and Sons: Chichester

Bruford, M. W., Bradley, D. G., Luikart, G. (2003). DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics* **4**: 900-910

Brumfield, R. T., Beerli, P., Nickerson, D. A., Edwards, S. V. (2003). The utility of nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution* **18**: 249-256

Bryant, D. & Moulton, V. (2004). Neighbour-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* **21**: 255-265

Buckland, P. C. and Dinnin, M. H. (1993). Holocene woodlands, the fossil insect evidence. In: Kirby, K. J., Drake, C. M. (Eds)., Dead Wood Matters: The Ecology and Conservation of Saproxylic Invertebrates in Britain. English Nature, Peterborough, pp 6-20

Bulatova, N. S., Golenishchev, F. N., Koval'skaya, Y. M., Emelyanova, L. G., Bystrakova, N. V., Pavlova, S. V., Nadzhafova, R. S., Lavrechenko, L. A. (2010). Cytogenetic study of the parapatric contact zone between two 46-chromosomal forms of the common vole in European Russia. *Russian Journal of Genetics* **46**: 443-448

Burger, J., Rosendahl, W., Loreille, O., Hemmer, H., Eriksson, T., Götherström, A., Hiller, J., Collins, M. J., Wess, T., Alt, K. W. (2004). Molecular phylogeny of the extinct cave lion *Panthera leo spelaea*. *Molecular Phylogenetics and Evolution* **30**: 841-849

Burney, D. A & Flannery, T. F. (2005). Fifty millennia of catastrophic extinctions after human contact. *Trends in Ecology and Evolution* **20**: 395-401

Bužan, E. V., Főrster, D. W., Searle, J. B., Kryštufek, B. (2010). A new cytochrome *b* phylogroup of the common vole (*Microtus aravlis*) endemic to the Balkans and its

implications for the evolutionary history of the species. *Biological Journal of the Linnaean Society* **100**: 788-796

Campbell, R. D., Harrington, A., Ross, A., Harrington, L. (2012). Distribution, population assessment and activities of beavers in Tayside. *Scottish Natural Heritage Commissioned Report No. 540*

Campbell, R. D., Rosell, F., Nolet, B. A., Dijkstra, V. A. A. (2005). Territory and group sizes in Eurasian beavers (Castor fiber): echoes of settlement and reproduction? *Behavioural Ecology and Sociobiology* **58**: 597-607

Cano, R. J & Borucki, M. K. (1992a). Isolation and partial characterisation of DNA from the bee *Proplebeia dominicana* (Apidae: Hymenoptera) in 25-40 million year old amber. *Medical Science Research* **20**: 249-251

Cano, R. J., Poinar, H, N,. Roublik, D. W & poinar Jr, G. O. (1992b). Enzymatic action and nucleotide sequencing of portions of the 18S rRNA gene of the bee *Proplebeia dominicana* (Apidae: Hymenoptera) isolated from 25 – 40 million year old Dominican amber. *Medical Science Research* **20**: 619-622

Cardini, A & Elton, S. (2007). Sample size and sampling error in geometric morphometric studies of size and shape. *Zoomorphology* **126**: 121-134

Carleton, M. D & Musser, G. G. (1984). Muroid Rodents. *In*: Anderson, S & Jones, J. K (eds). *Orders and families of recent mammals of the world*: pp. 289-379. Wiley and Sons: New York

Carter, I., Foster, J., Lock, L. The role of animal translocations in conserving British wildlife: An overview of recent work and prospects for the future. *EcoHealth **10:** 1-9*

Castańos, P. (1983). Estudio de los Macromamiferos de la Cueva de Santimamińe (Vizcaya). *Kobie (Serie Paleoantropologia y C. Naturales)* **14**: 235-318

Cerling, T. E., Harris, J. M., MacFadden, B. J., Leakey, M. G., Quade, J., Eisenmann, V., Ehleringer, J. R. (1997). Global vegetation change through the Miocene/Pliocene boundary. *Nature* **389**: 153-158

Chaline, J & Graf, J. D. (1988). Phylogeny of the Arvicolidae (Rodentia): biochemical and paleontological evidence. *Journal of Mammology* **69**: 22-33

Chaline, J. (1971). *Les rongeurs du Pliestocene moyan et superior de France*. Cahiers de paleontologie du CNRS: Paris

Chaline, J., Brunet-Lecomte, P., Montuire, S., Viroit, L., Courant, F. (1999). Anatomy of the arvicoline radiation (Rodentia): palaeogeographical, palaeoecological history and evolutionary data. *Annals Zoologici Fennici* **36**:239-267

Chambers, G. K and MacAvoy, E. S. (1999). *Molecular genetic analysis of hybridisation*. Department of Conservation: Wellington

Clark, C. D., Hugues, A. L. C., Greenwood, S. L., Jordan, C., Sejrup, H. P. (2012). Pattern and timing of retreat of the last British-Irish ice sheet. *Quaternary Science Review* **44**: 112-146

Clark, P. U., Dyke, A. S., Shakum, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., Mitrovica, J. X., Hostetler, S. W., McCabe, A. M. (2009). The Last Glacial Maximum. *Science* **325:**710-713

Clutton-Brock, J. (1999). *A Natural History of Domesticated Mammals.* Cambridge University press: New York

Coles, B. J. (1998). 'Doggerland': a speculative survey. *Proceedings of the Prehistoric Society* **64**: 45-81

Coles, B. J. (2006). *Beavers in Britain's Past.* Oxford: Oxbow Books

Collcutt, S. N., Currant, A. P., Hawkes, C. J. (1981). A further report on the excavations at Sun Hole, Cheddar. *Proceedings of the University of Bristol Speleological Society* **161**: 21-38

Collett, R. (1987). Bæveren I Norge, dens Utbredelsen og Levemaade (1986). *Bergens Museums Aarbog* **1**: 1-139

Comeron, J. M., Williford, A., Kliman, R. M. (2008). The Hill-Robertson effect: evolutionary consequences of weak selection and linkage. *Heredity* **100**: 19-31

Conroy, C. J. & Cook, J. A. (1999). MtDNA evidence for repeated pulses of speciation within arvlicoline and murid rodents. *Journal of Mammal Evolution* **6**: 221-245

Conroy, C. J. & Cook, J. A. (2000). Molecular systematics of a Holarctic rodent (*Microtus*: Muridae). *Journal of Mammology* **81**: 344-359

Cooke, S. B and Terhune, C. E. (2014). Form, function and geometric morphometrics. *The Anatomical Records* **298**: 5-28

Coope, G. R. (1997). Fossil coleopteran assemblages as sensitive indicators of climatic changes during the Devensian (Last) cold stage. *Philosophical Transactions of the Royal Society of London B* **280**: 313-40

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

Cooper, A., Mourer-Chauviré, C., Chambers, G. K., von Haeseler, A., Wilson, A. C., Pääbo, S. (1992). Independent origins of New Zealand moas and kiwis. *Proceedings of National Academy of Sciences* U.S.A **89:** 8741-8744

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

Corbet, G. B. (1961). Origin of British insular races of small mammals and of Lusitanian fauna. *Nature* **191**: 1037-1040

Corbet, G. B. (1986). Temporal and spatial variation of dental pattern in the voles, *Microtus aravlis*, of the Orkney Islands. *Journal of Zoology* **208**: 395-402

Corbet, G.B. and Harris, S (eds). (1991). *The handbook of British mammals*. 3rd ed. Blackwell Scientific Publications, Oxford.

Crees, J. J., Carbone, C., Sommer, R. S., Benecke, N., Turvey, S. T. (2016. Millennial-scale faunal records reveals differential resilience of European large mammals to human impacts across the Holocene. *Proceedings of the Royal Society B* **283**: 10.1098

Cucchi, T., Barnett, R., Martínková, N., Renaud, S., Renvoisé, E *et al*. (2014). The changing pace of insular life: 5000 years of evolution in the Orkney vole (*Microtus arvalis orcadensis*). *Evolution :* DOI: 10.111/evo.12476

Currant, A. P, & Jacobi, R. (2001). A formal mammalian biostratigraphy for the Late Pleistocene of Britain. *Quaternary Science Reviews* **20**: 1707-1716

Currant, A. P & Jacobi, R. (2011). The mammal faunas of the British late Pleistocene. In*: The ancient human occupation of Britain* (Eds. Ashton, N. M., Lewis, S. G., Stringer, C. B). *Developments in Quaternary Science* **14***:* 165 - 180

Dabney, J., Knapp, M., Glocke, I., Gansauge, M. T., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J-L., Meyer, M. (2013). Complete mitochondrial genome sequence of a Pleistocene cave bear reconstructed from ultra-short DNA fragments. *Proceedings of the National Academy of Sciences* **110**: 15758-15763

Dalén, l., Nyström, V., Valdiosera, C., Germonpré, M., Sablin, M., Turner, E., Angerbjörn, A., Arsuaga, J. L., Götherström, A. (2007). Ancient DNA reveals lack of postglacial habitat tracking in the arctic fox (*Alopex lagopus*). *Proceedings of the National Academy of Science U.S.A* **104**: 6726-6729

Damuth, J & Fortelius, M. (2001). Reconstructing mean annual precipitation, based on mammalian dental morphology and local species richness. In *EEDEN Plenary Workshop on Late Miocene to Early Pliocene Environments and Ecosystems* (J. Agustí and O. Oms, eds.), pp. 23–24. Sabadell, Spain: EEDEN Programme, European Science Foundation.

Dansgaard, W., Johnsen, S. J., Clausen, H. B., Dahl-Jensen, D., Gundestrup, N. S., Hammer, C. U., Hvidberg, C. S., Steffensen, J. P., Sveinbjörnsdottir., Jouzel, J., Bond, G. (1993). Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* **364**: 218-220.

Darriba, D., Taboada, G. L., Doallo, R., Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**: 772

Davis, A. R. and Gray, D. (2010). *The distribution of Scottish wildcats (Felis silvestris) in Scotland (2006-2008).* Scottish Natural Heritage Commissioned Report No. 360

Dawson, A. G. (1992). *Ice Age Earth; Late Quaternary Geology and Climate*. Routledge: London

Degerbøl, M. (1933). Danmarks pattedyr I fortiden I sammenligning med recente former. Videnskabelige. *Meddelelser fra Dansk Naturhistorisk Forening I København* **96**: 357-641

DeGusta, D & Vrba, E. (2005). Methods for inferring palaeohabitats from the functional morphology of bovid phalanges. *Journal of Archaeological Science* **32**:1099-1113

Derrien, T., Estellé, J., Sola, S. M, Knowles, D. G., Rauneri, E., Guigó, R., Ribeca, P. (2012). Fast computation and applications of genome mappability. *PLoS One***:** http://dx.doi.org/10.1371/journal.pone.0030377

Devillard, S., Jombart, T., Léger, F., Pontier, D., Say, L., Ruette, S. (2014). How reliable are morphological and anatomical characters to distinguish European wildcats, domestic cats and their hybrids in France? *Journal of Zoological Systematics and Evolutionary Research* **52**: 154-162

DeWoody, J. A., Chesser, R. K., Baker, R. J. (1999). A translocated mitochondrial cytochrome *b* pseudogene in voles (Rodentia: *Microtus*). *Journal of Molecular Evolution* **48**: 380-382

Dienske, H. (1969). Notes on differences of some external and skull characters of *Microtus arvalis* (Pallas, 1779) and of *Microtus agrestis* (Linnaeus, 1761) from the Netherlands. *Zoologishe Mededelingen* **44**:83-108

Drake, A. G. & Klingenberg, C. P. (2008). The pace of morphological change: historical transformation of skull shape in St. Bernard dogs. *Proceedings of the Royal Society B* **275**: 71-76

Driscoll, C., Juliet, C., Andrw, C., O'Brian, S. (2009). The taming of cats. *Scientific American* **300**: 68-75

Driscoll, C., Menotti-Raymond, M., Roca, A. L., *et al.* (2007). The Near Eastern origin of cat domestication. *Sciencexpress* DOI: 10.1126/science.1139518

Driscoll, C., Yamaguchi, N., O'Brian S. J., MacDonald, D. W. (2011) A suite of genetic markers useful in assessing wildcat (*Felis silvestris ssp.*) – domestic cat (*Felis silvestris catus*) admixture. *Journal of Heredity* **102**: S87-S90

Drummond, A. J, Suchard, M. A, Xie, D., Rambaut, A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7 *Molecular Biology and Evolution* **29**: 1969-1973

Drummond, A. J., Ho, S. Y. W., Rawlence, N., Rambaut, A. (2007). A rough guide to BEAST 1.4. Available at:

http://workshop.molecularevolution.org/molevolfiles/beast/BEAST14_MANUAL-7-6-07.pdf

Dryden, I. L. and K. V. Mardia. (1998). Statistical Shape Analysis. John Wiley and Sons: Chichester, England

Ducroz, J-F., Stubbe, M., Saveljev, A., P., Heidecke, D., Samjaa, R., Ulevičius., Stubbe, A., Durka, W. (2005). Genetic variation and population structure of the Eurasian beaver *Castor fiber* in Eastern Europe and Asia. *Journal of Mammalogy* **86**: 1059-1067

Durka, W., Babik, W., Ducroz, J-F., Heidecke, D., Rosell, F., Samjaa, R., Saveljev, A.P., Stubbe. A., Ulevičius. A., Stubbe. M. (2005). Mitochondrial phylogeography of the Eurasian beaver *Castor fiber* L. *Molecular Ecology* **14**: 3843-3856

Easterbee, N., Hepburn, L. V., Jefferies, D. J. (1991*). Survey of the Status and Distribution of the Wild Cat in Scotland 1983-1987*. Nature Conservancy Council for Scotland, Edinburgh

Edwards, K. J., Langdon, P. G., Sugden, H. (2007). Separating climatic and possible human impacts in the early Holocene: biotic response around the time of the 8200 cal. yr BP event. *Journal of Quaternary Science* **22**: 77-84

Edwards, C. J., Magee, D. A., Park, S. D. E. *et al.* (2010). A complete mitochondrial genome sequence from a mesolithic wild aurochs (*Bos primigenius*). Plos ONE **5**: e9255

Elias, S. A and Matthews, I. P. (2014). A comparison of reconstructions based on aquatic and terrestrial beetle assemblages: Late glacial – Early Holocene temperature reconstructions for the British Isles. *Quaternary International* **34**1: 69-79

Ellegren, H., Hartman, G., Johansson, M., Andersson, L. (1993). Major histocompatibility complex monomorphism and low levels of DNA fingerprinting variability in a reintroduced and rapidly expanding population of beavers. *Proceedings of the National Academy of Sciences of the United States of America* **90**: 8150-8153

Escudé, E., Renvoisé, É., Lhomme, V., Montuire, S. (2013). Why all vole molars (Arvicolinae, Rodentia) are informative to be considered as proxy for Quaternary paleoenvironmental reconstructions. *Journal of Archaeological Science* **40**: 11-23

Excoffier, L. (2004). Patterns of DNA sequence diversity and genetic structure after a range expansion: ;essons from the infinite-island model. *Molecular Ecology* **13**: 853-864

Excoffier, L., Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**: 564-567

Fedyk, A & Ruprecht, A. L. (1971). Taxonomic value of the M_1 measurements in *Microtus agrestis* (Linnaeus, 1761) and *Microtus arvalis* (Pallas, 1779). *Acta Theriologica* **16**:408-412

Feustel, R. (1980). Problem Zeitstellung. *Weimarer Monographien zur Ur-and Frühgeschichte* **3**: 114-120

Fink, S., Fischer, M. C., Excoffier, L., Heckel, G. (2010). Genomic scans support repetitive continental colonization events during the rapid radiation of voles (Rodentia:Microtus): the utility of AFLP's versus mitochondrial and nuclear sequence markers. *Systematic Biology* **59**:548-572

Flynn, L. J. & Jacobs, L. L. (2007). Castoroidea. In: *Evolution of Tertiary Mammals of North America*. Cambridge: Cambridge University Press

Forster P. and Toth A. (2003). Toward a phylogenetic chronology of ancient Gaulish, Celtic, and Indo-European. *Proceedings of the National Academy of Science* s USA **100:** 9079–9084.

Fortelius, M., Eronen, J., Jernvall, J., Liu, L., Pushkina, D., Rinne, J., Tesakov, A., Vislobokova, I., Zhang, Z., Zhou, L. (2002). Fossil mammals resolve regional patterns of Eurasian climate change over 20 million years. *Evolutionary Ecology Research* **4**: 1005-1016

Foster, P. (2009). Maximum likelihood and Bayesian analysis in molecular phylogenetics. Retrieved 30-09/2016 from http://research.ncl.ac.uk/microbial_eukaryotes/documents/peter_rio09_toPrint2.pdf

Frahnert, S. (2000). Ontogenetic changes of cranial proportions in beaver, *Castor fiber* L., 1758 (Rodentia, Castoridae): taxonomical implications and functional aspects. *Bonner Zoologische Beiträge* **49**: 131-153

Frankham, R. (1996). Relationship of genetic variation to population size in wildlife. *Conservation Biology* **10**:1500-1508

Frankham, R., Ballou, J.D., Briscoe, D.A. (2010). *Introduction to Conservation Genetics*, 2nd edition. Cambridge University Press, Cambridge, U.K.

Franks, S. J. and Hoffmann, A. A. (2012). Genetics of climatic change adaptation. *Annual Review of Genetics* **46**: 185-208

Frantz, L. A. F., Mullin, V, E., Pionnier-Capitan, M. (2016). Genomic and archaeological evidence suggests a dual origin of domestic dogs. Science **352**: 1228-1231

Freye, H. A. (1960). Zur Systematik der Castoridae (Rodentia, Mammalia). *Mitt Zoological Mus Berlin* **36**: 105-122

Frosch, C., Kraus, R. H. S., Angst, C., Allgöwer, R., Michaux, J., Teubner, J., Nowak, C. (2014). The genetic legacy of multiple beaver reintroductions in Central Europe. *PLoS ONE* **9**: e97619

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915-925

Fyfe, R. M., Twiddle, C., Sugita, S., Waller, M. P. (2013). The Holocene vegetation cover of Britain and Ireland: overcoming problems of scale and discerning patterns of openness. *Quaternary Science Reviews* **73**: 132-148

Geer, L. Y, Marchler-Bauer, A., Geer, R. C., Han, L., Han, J., He, J., He, S., Liu C., Shi, W., Bryant, S. H. (2010). The NCBI Biosystems database. *Nucleic Acids Resources* **38**: D492-496

Gelewski, T., Tilak, M, Sanchez, S., Chevret, P, Paradis, E., Douzery, E. J. P. (2006). BMC *Evolutionary Biology* **6**: 80-97

Ghilardi, B & O'Connell, M. (2013). Early Holocene vegetation and climate dynamics with particular reference to the 8.2 ka event: pollen and macrofossil evidence from a small lake in western Ireland. *Vegetation History and Archaeobotany* **22**: 99-114

Gibbard, P. L., Smith, A. G., Zalasiewicz, J. A., Barry, T. L., Cantrill, D., Coe, A. L., Cope, J. C. W., Gale, A. S., Gregory, F. J., Powell, J. H., Rawson, P. F., Stone, P. Waters, C. N. (2005). What status for the Quaternary? *Boreas* **34**: 1-6

Gilbert, M. T. P., Jenkins, D. L., Götherstrom, A., Naveran, N., Sanchez, J, J., Hofreiter, M., Thompsen, P. F., Binladen, J., Higham, T. F. G., Yohe II, R. M., Parr, R., Cummings, L. S., Willerslev, E. (2008). DNA from pre-clovis human coprolites in Oregon, North America. *ScienceExpress* **320**: 786-789

Gilbert, M. T. P., Willerslev, E., Hansen, A. J., Barnes, I., Rudbeck, L., Lynnerup, N., Cooper, A. (2003). Characterisation of genetic miscoding lesions caused by post-mortem damage. *American Journal of Human Genetics* **72**:32-47

Gould, S. J. (1966). Allometry and size in ontogeny and phylogeny. *Biological Reviews* **41**: 587-638

Graham, R. W., Belmecheri, S., Choy, K. *et al.* (2016). Timing and cuases of mid-Holocene mammoth extinction on St. Paul Island, Alaska. *Proceedings of the National Academy of Sciences* **133**: 9310-9314

Grant, A. (1984). 'Animal Husbandry', 496-547 in B. Cuncliffe (Ed.). *Danebury: An Iron Age Hillfort in Hampshire. Volume 2: The Excavations, 1969-1978: The Finds.* Council for British Archaeology Research Report 52: London

Green, R. E., Krause, J., Briggs, A. W., *et al.* (2010). A draft sequence of the Neanderthal genome. *Science* **328**: 710-722

Green, W. D. K. (1996). The thin-plate spline and images with curving features, p. 79-87. In K. V. Mardia, C. A. Gill and I. L. Dryden (eds.), *Image Fusion and Shape Variability*. University of Leeds Press: Leeds.

Grill, A., Amori, G., Aloise, G., Lisi, I., Tosi, G., Wauters, A, A., Randi, W. (2009). Molecular phylogeography of European *Sciurus vulgaris*: refuge within refugia? *Molecular Ecology* **18**: 2687-2699

Grupe, G. (2007). Taphonomic and diagenetic processes. In: *Handbook of Palaeoanthropology*. (Eds: Henke, I and Tattersall, T). Springer: New York

Gupta, S. K and Polach, H. A. (1985). *Radiocarbon dating practices at ANU: Radiocarbon Dating Research*. Radiocarbon dating Laboratory, Research School of Pacific Studies ANU: Canberra

Gurnell, J., Gurnell, A. M., Demeritt, D., Lurz, P. W. W., Shirley, M. D. F., Rushton, S. P., Faulkes, C. G., Nobert, S. & Hare, E. J. (2008). *The Feasibility and Acceptability of Reintroducing the European Beaver to England.* p. 106. Sheffield, UK: Natural England/People's Trust for Endangered Species, Sheffield, UK.

Guthrie, R. D. (1965). Variability in characters undergoing rapid evolution, and analysis of *Microtus* molars. *Evolution* **19**: 214-233

Guthrie, R. D. (1982). Mammals of the mammoth steppe as palaeoenvironmental indicators. In: Hopkins, D. M., Mathews, J. V., Schewger, C. E & Young, S. B (eds): *Paleoecology of Beringia* pp 307-326. Academic Press

Hadly., E. A., Ramakrishnan, U., Chan, Y. L., van Tuinen, M., O'Keefe, K., Spaeth, P. A., Conroy, C. J. (2004). Genetic responses to climate change: insights from ancient DNA and phylochronology. *Plos Biology* **2**:1600-1609

Hagelberg, E., Thomas, M. G. Cook Jr, C. E., Sher, A. V., Baryshnikov, G., Lister, A. M. (1994). DNA from ancient mammoth bones. *Nature* **370**: 333

Hajibabaei. M., Singer, G. A. C., Hebert, P. D. N., Hickey, D. A. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics* **23**: 167–172.

Halanych, K. M. & Robinson, T. J. (1999). Multiple substitutions affect the phylogenetic utility of cytochrome *b* and 12S rDNA data: Examining a rapid radiation in the leporid (Lagomorpha) evolution. *Journal of Molecular Ecology* **48**: 369-379

Halley, D. (2011). Sourcing Eurasian beaver *Castor fiber* stock for re-introduction in Great Britain and Western Europe. *Mammal Review* **41**:40-53

Halley, D. J. & Rosell, F. (2002). The beaver's reconquest of Eurasia: status, population development and management of a conservation success. *Mammal Review* **32**: 153-178

Halley, D. J. & Rosell, F. (2003). Population and distribution of European beavers (*Castor fiber*). *Lutra* 46: 91-101

Halley, D. J., Rosell, F., Saveljev, A. (2012). Population and distribution of Eurasian beaver (*Castor fiber*). *Baltic Forestry* **18**:168-175

Handt, O., Höss, M., Krings, M., Pääbo, S. (1994). Ancient DNA: methodological challenges. *Experientia* **50**:524-529

Hänni, C., Laudet, V., Stehelin, D., Taberlet, P. (1994). Tracking the origins of the cave bear (*Ursus spelaeus*) by mitochondrial DNA sequencing. *Proceedings of the National Academy of Sciences U.S.A* **91**: 12336-12340

Harcourt, R. (1979). 'The Animal Bones' 150-160 in G. J. Wainwright (Eds.). *Gussage All Saints: An Iron Age Settlement in Dorset*, (DoE Archaeological Reports 10). HMSO, London

Haring, E., Voyta, L. L., Däubl, B., Tiunov, M. P. (2015). Comparison of genetic and morphological characters in fossil teeth of grey voles from the Russian Far East (Rodentia: Cricetidae: *Alexandromys*). *Mammalian Biology – Zeitschrift für Säugetierkunda* **80**: 496-504

Harris, S., Morris, P., Wray, S., Yalden, D. (1995). *A review of British mammals: populations estimates and conservation status of British mammals other than cetaceans.* Peterborough: Joint Nature Conservancy Council, pp183

Hartman, G. (1996). Habitat selection by European beaver (*Castor fiber*) colonizing a boreal landscape. *Journal of Zoology* **240**: 317-325

Hartmann, S. A., Steyer, K., Kraus, R. H. S., Segelbacher, G., Nowak, C. (2013). Potential barriers to gene flow in the endangered wildcat (*Felis silvestris*). *Conservation Genetics* **14**: 413 doi:10.1007/s10592-013-0468-9

Haynes, S., Jaarola, M., Searle, J. B. (2003). Phylogeography of the common vole (*Microtus arvalis*) with particular emphasis on the colonization of the Orkney archipelago. *Molecular Ecology* **12**: 951-956

Hays, J. D., Imbrie, J., Shackleton, N. T. (1976). Variations in the Earth's Orbit: Pacemaker of the Ice Ages. *Science* **194**: 1121-1132

Heckel, G., Burri, R., Fink, S., Desmet, J-F, Excoffier, L. (2005). Genetic structure and colonization processes in European populations of the common vole, *Microtus arvalis*. *Evolution* **59**: 2231-2242

Hedges, R. E. M., Housley, R. A., Law, I. A., Perry, C., Gowlett, J. A. J. (1987). Radiocarbon dates from the oxford ams system: archaeometry datelist 6. *Archaeometry* **29**: 289-306

Hedges, R. E. M., Pettit, P. B., Bronk Ramsey, C., van Klinken, G. J. (1998). ORAU Datelist 25. *Archaeometry* **401**: 227-239

Hedrick, P. A. (2011). *Genetics of Populations 4th Ed.* Jones and Bartlett: Massachusetts

Heidecke, D. (1986). Taxonomical aspects of species conservation exemplified with the Eurasian beaver. *Hercynia N. F.* **2**: 146-161

Hellborg, L., Walker, C. W., Rueness, E. K., Stacy, J. E., Kojola, I., Valdmann, H., Vilá, C., Zimmerman, B., Jakobsen, K. S., Ellegren, H. (2002). Differentiation and levels of genetic variation in northern European Lynx (*Lynx lynx*) revealed by microsatellite and mitochondrial DNA analysis. *Conservation Genetics* **3**: 97-111

Herbert, P. D. N., Cywinska, A., Ball, S. L., deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B* 270: DOI: 10.1098/rspb.2002.2218

Herman, J. S and Searle, J. B. (2011). Post-glacial partitioning of mitochondrial genetic variation in the field vole. *Proceedings of the Royal Society of London B* 278: 3601-3607

Herman, J. S., McDevitt, A. D., Kawalko, A., Jaarola, M., Wójcik, J. M., Searle, J. B. (2014). Land-bridge calibration of molecular clocks and the post-glacial colonization of Scandinavia by the Eurasian field vole *Microtus agrestis*. *PLoS One* **9**: e103949

Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**: 87-112

Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907-913

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences* **359**: 183-195

Higgins, D. and Austin, J. J. (2013). Teeth as a source of DNA for forensic identification of human remains: a review. *Science & Justice* **53**: 433-441

Higham, T. F, G., Jacobi, R. N & Ramsey, C. B. (2006). AMS radiocarbon dating of ancient bone using ultrafiltration. *Radiocarbon* **48**: 179-195

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

Ho, S. Y. W., Lanfear, R., Bromham, L., Phillips M. J., Soubrier, J., Rodrigo, A. G., Cooper, A. (2001). Time dependent rates of molecular evolution. *Molecular Ecology* **20**: 3087-3101

Hofreiter, M & Stewart, J. (2009). Ecological change, range fluctuations and population dynamics during the Pleistocene. *Current Biology* **19**: 584-594

Hofreiter, M., Rabeder, G., Jaenicke-Després, V., Withalm, G., Nagel, D., Paunovic, M., Jambrěsić, G., Pääbo, S. (2004). Evidence for reproductive isolation between cave bear populations. *Current Biology* **14**: 40-43

Hofreiter, M., Paijams, J. L. A., Goodchild, H., Speller, C. F., Barlow, A., Fortes, G. G., Thomas, J. A., Ludwig, A. (2015). The future of ancient DNA: technical advances and conceptual shifts. *BioEssays* **37**: 384-293

Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M., Pääbo, S. (2001). Ancient DNA. *Nature Reviews: Genetics* **2:** 353-359

Hoftrieter, M., Capelli, C., Krigs, M. *et al.* (2002). Ancient DNA analysis reveals high mitochondrial DNA sequence diversity and parallel morphological evolution of late Pleistocene cave bears. *Molecular Biology and Evolution* **19**: 1244-1250

Holder, M & Lewis, P. O. (2003). Phylogeny estimation: traditional and Bayesian approaches. *Nature Reviews Genetics* **4**: 275-284

Horn, S., Durka, W., Wolf, R., Ermala, A., Stubbe, A., Stubbs, M., Hofreiter M. (2011). Mitochondrial genomes reveal slow rates of molecular evolution and the timing of speciation in beavers (*Castor*), one of the largest rodent species. *PLoS ONE* **6:** e14622

Horn, S., Prost, S., Stiller, M., Makowiecki, D., Kuznetsova, T., Benecke, N., Pucher, E., Hufthammer, A. K., Schouwenburg, C., Shapiro, B., Hofreiter, M. (2014). Ancient mitochondrial DNA and the genetic history of the Eurasian beaver (*Castor fiber*) in Europe. *Molecular Ecology* **23**: 1717-1729

Horsburgh, K. A. (2008). Wild or domesticated? An ancient DNA approach to canid species identified in South Africa's Western Cape Province. *Journal of Archaeological Science* **35**: 1474-1480

Höss, M., Pääbo, S., Vereshchagin, N. K. (1996). Mammoth DNA sequences. *Nature* **370**: 333

Hu, Y., Hu, S., Wang, W., Wu. X., Marshall, F. B., Chen, X., Hou, L., Wang, C. (2014). Earliest evidence for commensal processes of cat domestication. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 116-120

Hubbard, A. (1999). High-resolution modelling of the advance of the Younger Dryas ice sheet and its climate in Scotland. *Quaternary Research* **52**: 2743

Hubbard, A. L., McOris, S., Jones, T. W., Boid, R., Scott, R, Easterbee, N. (1992). Is survival of European wildcats Felis silvestris in Britain threatened by interbreeding with domestic cats? *Biological Conservation* **61**: 203-208

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754-5

Hugueney, M & Escuillie, F. (1996). Fossil evidence for the origin of behavioural strategies in early Miocene Castoridae and their role in the evolution of the family. *Paleobiology* **22**: 507-513

Hung, C.-M., Lin, R-C., Chu, J-H., Yeh, C-F., Yao, C-J., Li, S-H. (2013). The *De Novo* assembly of mitochondrial genomes of the extinct passenger pigeon (*Ectopistes migratorius*) with Next Generation Sequencing. *PLoS ONE* **8**: e56301

Hurst, G. D. D & Jiggins, F. (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B* **272**: 1525-1534

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary biology. *Molecular Biology and Evolution* **23**: 254-267

Huynen, L., Millar, C. D., Scofield, R. P., Lambert, D. M. (2003). Nuclear DNA sequences detect species limits in ancient Moa. *Nature* **425**: 175-178

International Union for the Conservation of Nature (IUCN). (2007). *European Mammal assessment*. IUCN, Gland, Switzerland and Cambridge, UK

Irwin, D., Kocher, T., Wilson, A. (1991) Evolution of the cytochrome *b* gene of mammals. *Journal of Molecular Evolution* **32**: 128-144.

Isarin, R. F. B. and Bohncke, S. J. P. (1999). Mean July temperatures during the Younger Dryas in Northwestern and Central Europe as inferred from climate indicator plant species. *Quaternary Research* **51**: 158-173

IUCN. (1998). IUCN/SSC *Guidelines for re-introductions*. IUCN Species Survival Commission, Gland, Switzerland.

IUCN/SSC. (2013). *Guidelines for reintroductions and other conservation translocations.* Version 1.0. IUCN Species Survival Commission, Gland, Switzerland, viiii + 57pp

Jaarola, M., Martíková, N., Gündüz, I., Brunhoff, C., Zima, J., Nadachowski, A., Aori, G., Bulatova, N. S., Chondropoulos, B., Fraguedakis-Tsolis, S., González-Esteban, J., López-Fuster, J. M., Kandaurov, A. S., Kefelioğlu, H., Mathias, M. da Luz., Villate, I., Searle, J. B. (2004). Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* **33**: 647-663

Jacob, J., Manson, P., Barfknecht, R., Fredricks, T. (2014). Common vole (*Microtus arvalis*) ecology and management: implications for risk assessment of plant protection products. Pest management. *Science* **70**: 869-878

Jacobi, R. M., & Higham, T. F. G. (2009). The early Lateglacial re-colonisation of Britain: new radiocarbon evidence from Gough's Cave, southwest England. *Quaternary Science Reviews* **28**: 1895-1913

Jacobi, R. M., Higham, T. F. G & Ramsey, C. B. (2006). AMS radiocarbon dating of Middle and Upper Palaeolithic bone in the British Isles: improved reliability using ultrafiltration. *Journal of Quaternary Science* **21**: 557-573

Jakobsson, M., Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**: 1801-1806

Jia, F., Lo, N., Ho, S. Y. W. (2014). The impact of modelling rate heterogeneity among sites on phylogenetic estimates of intraspecific evolutionary rates and timescales. *PLoS ONE* **9**: e95722

Jollife, I. T. (2002). *Principal components analysis*. 2nd Edition. Springer: New York

Jones, S., Campell-Palmer, R. (2014). *The Scottish Beaver Trial: The Story of Britain's First Licenced Release into the Wild.* Scottish Natural Heritage Final Report

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimation of ancient DNA damage parameters. *Bioinformatics* DOI: 10.1093

Jope, M. and Grigson, C. (1965). 'Faunal Remains', 141-167 in Keller (Ed), *Windmill Hill and Avebury*. Claredon Press: Oxford

Jouzel, J., Stievenard, M., Johnsen, S. J., Landias, A., Masson-Delmotte, V., Sveinbjornsdottir, A., Vimeux, F., Grafenstein, U von., White, J. W. C. (2007). The GRIP deuterium-excess record**.** *Quaternary Science Reviews* **26:** 1-17

Jullien, R. (1981). La faune des vertébres á l'excusion de l'homme, des oiseaux, des ronguers et des poisons. In: *La Grotte Préhistorique de Kitsos (Attique). Missions 1968-78. Recherches sur les grands civilisations:* 569-590. Lamberts, N (Ed). Paris: ADPF

Jurdíkova, N., Žiak, D., Kocian, L. (2000). Habitat requirements of *Microtus tatricus*: macrohabitat and microhabitat. In: Urban P. (ed.). *Výskum a ochrana cicavcov na Slovensku 4*. SOP COPK: Banská Bystrica

Kaagan, L. M. (2000). *The horse in late Pleistocene and Holocene Britain.* PhD dissertation, University College London

Kahlke, R. D. (1999). *The history of the Origin, Evolution and Dispersal of the Late Pleistocene Mammuthus-Coelodonta Faunal Comples in Eurasia (Large Mammals).* Fenske Companies: Rapis City)

Kalis, A. J., Merkt, J & Wunderlich, J. (2003). Environmental changes during the Holocene climatic optimum in central Europe – human impact and natural causes. *Quaternary Science Reviews* **22**: 33-79

Kass, R. E and Raferty, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**: 773-795

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, ***28:*** 1647-1649

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, ***28:*** 1647-1649

Kefi, R. (2011). Ancient DNA investigations: A review of their significance in different fields. *International Journal of Modern Anthropology* **4:**61-76

Keis, M., Remm, J., Ho, S. Y. W., Davison, J., Tammaleht, E., Tumanov, I. L., Saveljev, A. P., Männil, P., Kojola, I., Abramov, A. V., Margus, T., Saarma, U. (2013). Complete mitochondrial genomes and a novel spatial genetic method reveal cryptic phylogeographical structure and migration patterns among brown bears in north-western Eurasia. *Journal of Biogeography* **40**: 915-927

Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology and Evolution* **58**: 870-879

Kelly, A., Charman, D. J., Newnham, R. M. (2010). A Last Glacial Maximum pollen record from Bodmin Moor showing a possible cryptic northern refugium in southwest England. *Journal of Quaternary Science* **25**: 296-308

Kendall, D. G. (1977). The diffusion of shape. *Advances in Applied Probability* 9: 428-430

Kendall, D. G. (1984). Shape-manifolds, Procrustean metrics and complex projective spaces. *Bulletin of the London Mathematics Society* **16**: 81–21

Ki, J-S., Hwang, S-S., Park, T-J., Han, S-H., Lee, J-S. (2010). A comparison analysis of the complete mitochondrial genome of the Eurasian otter *Lutra Lutra* (Carnivora; Mustelidae). *Molecular Biology Reports* **37**: 1943-1955

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**: 111-120

Kitchener, A. C and Rees, E. E. (2009). Modelling the dynamic biogeography of the wildcat: implications for taxonomy and conservation. *Journal of Zoology* **279**: 144-155

Kitchener, A. C. & O'Connor, T. (2010). *Wildcats, Domestic and Feral Cats. In: Extinctions and Invasions: A social history of British Fauna. Eds:* O'Connor, T & Sykes, N. pp 83-94. Oxbow Books: Oxford

Kitchener, A. C. (1998). The Scottish wildcat – a cat with an identity crises? *British Wildlife* **9**: 232-242

Kitchener, A. C., Yamaguchi, N., Ward, J. M., MacDonald, D. W. (2005). A diagnosis for the Scottish wildcat (*Felis silvestris*): a tool for conservation action for a critically-endangered felid. *Animal Conservation* **8**: 223-237

Kitchener, A., & Lynch, J. M. (2000). *A morphometric comparison of the skulls of fossil British and extant European beavers, Castor fiber*. SNH Review 127, Scottish Natural Heritage, Edinburgh.

Kleiven, H. F., Kissel, C., Laj, C. *et al.* (2008). Reduced North Atlantic deep water coeval with the glacial lake Agassiz freshwater outburst. *Science* **319**: 60-64

Klingenberg, C. P., Barluenga, M., Meyer, A. (2002). Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. *Evolution* **56**: 1909-1920

Klingenberg, C. P. (2011). MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources* **11**: 353-357

Klingenberg, C. P. (2016). Size, shape and form: concepts of allometry in geometric morphometrics. *Development, Genes and Evolution* **226**: 113-137

Klingenberg, C. P., Barluenga, M., Meyer, A. (2002). Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. *Evolution* **56**: 1909-1920.

Knaus, B. J., Cronn, R., Liston, A., Pilgrim, K., Schwartz, M. K. (2011). Mitochondrial genome sequences illuminate maternal lineages of conservation concern in a rare carnivore. *BMC Ecology* **11**:10

Kobashi, T., Severinghaus, J. P., Brook, E. J., Barnola, J-M., Grachev, A. M. (2007). Precise timing and characterisation of abrupt climate change 8200 years ago from air trapped in polar ice. *Quaternary Science Reviews* **26**: 1212-1222

Koch, P. L., Barnosky, A. D. (2006) Late Quaternary Extinctions: State of the Debate. *Annual Review of Ecology, Evolution, and Systematics* **37**: 215-250

Koenigswald, W. von. (2002). Migrations and extinctions in the Quaternary faunas of Central and Western Europe. *Annales Géologiques Pays Helléniques* **39**: 327-335

Koenigswald, W. **von**. (2003). Mode and causes for the Pleistocene turnover in the mammalian fauna of Central Europe. *In: Distribution and migration of Tertiary mammals in Eurasia. A volume in honour of Hans de Bruijin. Eds:* Reumer, J. W. F and Wessels, W. Vol 10, pp 305-312

Kowalski, K. *(2001).* Pleistocene rodents of Europe. *Folia Quaternaria* **72***:*1–289

Kumar, S., Stecher, G., Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**: doi: 10.1093/molbev/msw054

Kurtén, B. (1965). 'On the evolution of the wild cat, *Felis silvestris* Schreber'. *Acta Zoologica Fennica III* 1-29

Laffont, R., Renvoisé, E., Navarro, N., Alibert, P., Montuire, S. (2009). Morphological modularity and assessment of developmental processes within the vole dental row (Microtus arvalis, Arvicolinae, Rodentia). *Evolution and Development* **11**: 302-311

Lang, B., Bedford, A., Brooks, S. J. *et al.* (2010). Early-Holocene temperature variability inferred from chironomid assemblages at Hawes water, northwest England. *The Holocene* **20**: 943-954

Lang, G. (1994). *Quartäre Vegetationsgeschichte Europas*. Jena: Gustav Fischer-Verlag

Larson, G. and Burger, J. (2013). A population genetic view of animal domestication. *Trends in Genetics* **29**: 197-205

Lavrov, L. S. (1979). Species of beavers (*Castor*) of the Palearctic. *Zoologicheskii Zhurnal* **58**: 88-96

Lavrov, L. S. (1983). Evolutionary developments of the genus *Castor fiber* and taxonomy of the contemporary beavers of Eurasia. *Acta Zoologica Fennica* **174**: 87-90

Lawson, T. J. (1981). 'The 1926-7 excavations of the Creag nan Uamh bone caves, near Inchnadamph, Sutherland'. *Proceedings of the Royal Society of Antiquaries of Scotland* **III**: 7-20

Lebreton, L., Desclaux, E., Hanquet, C., Moigne, A. M., Perrenoud, C. (2015). Environmental context of the Caune de l'Arago Acheulean occupations (Tautavel, France), new insights from microvertebrates in Q-R levels. *Quaternary International* http://dx.doi.org/10.1016/j.quaint.2015.12.001

Lecis, R., Pierpaoli, M., Biro, Z. S., Szemethy, L., Ragni, B., Vercillo, F., Randi, E. (2006). Bayesian analyses of admixture of wild and domestic cats (*Felis silvestris*) using linked microsatellite loci. *Molecular Ecology* **15**: 119-131

Ledevin, R., Michaux, J. R., Deffontaine, V., Henttonen, H., Renaud, S. (2010). Evolutionary history of the bank vole *Myodes glareolus*: a morphometric perspective. *Biological Journal of the Linnaean Society* **100**: 681-694

Leonard, J. A., Vilà, C., Fox-Dobbs, K., Koch, P. L., Wayne, R. K., van Valkenburg, B. (2007). Megafaunal extinctions and the disappearance of a specialized wolf ecomorph. *Current Biology* **17:** 1146-1150

Leonard, J. A., Echegaray, J., Randi, E., Vilà, C. (2014). Impact of hybridisation with domestic dogs on the conservation of wild canids. In: *Free ranging dogs and wildlife conservation* (Ed: Gompper, M. E.). Oxford University Press: Oxford

Li, G., Davis, B. W., Eizirik, E., Murohy, W. J. (2015). Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Resources* DOI:10.1101/gr.186668.114

Li, G., Hillier, L. W., Grahn, R. A. *et al.* (2016). A high-resolution SNP array-based linkage map anchors a new domestic cat draft genome assembly and provides detailed patterns of recombination. *G3* **6:** 1607-1613

Li, H., Handsaker B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup. (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, **25**, 2078-9. [PMID: 19505943]

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

Li, M., Schroeder, R., Ko, A., Stoneking, M. (2012). Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTS. *Nucleic Acids Research* **40**: 1-8

Lightowlers, R. N., Chinnery, P. F., Turnbull, D. M., Howell, N. (1997). Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. *Trends in Genetics* **11**: 450-455

Lima-Ribeiro, M., S., Hortal, J., Varela, S., Diniz-Filho, J. A. F. (2014). Constraint envelope analyses of macroecological patterns reveal climatic effects on Pleistocene mammal extinctions. *Quaternary Research* **82**: 260-269

Lisiecki, L. E & Raymo, M. E. (2005). A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. *Paleoceanography* **20**: 1003.

Lister, A. M. (1992). Mammalian fossils and Quaternary biostratigraphy. *Quaternary Science Reviews* **11**:329-344

Lohmann, G. P. (1983). Eigenshape analysis of micro-fossils – a general morphometric procedure for describing changes in shape. *Journal of the International Association for Mathematical Geology* **15**: 659–672.

Longin, R. (1971). New method of collagen extraction for radiocarbon dating. *Nature* **230**: 241-2

Lopez, J. V., Cevario, S., O'Brian, S. J. (1996). Complete nucleotide sequences of the domestic cat (Felis catus) mitochondrial genome and a transposed mtDNA tandem repeat (Numt) in the nuclear genome. *Genomics* **33**: 229-246

Lopez-Gutierrez, B., MacLeod, M., Edgecombe, G. (2001). Detecting taxonomic signal in an under-utilised character system: Geometric morphometrics of the forcipular coxae of Scutigeromorpha (Chilopoda). ZooKeys **156**: 49-66

Lord, T. C., Thorp, J. A., Wilson, P. (2015). A wild boar dominated ungulate assemblage from an early Holocene natural pit fall trap: Cave shaft

sediments in northwest England associated with the 9.3 ka BP cold event. *The Holocene*: 1-7

Lowe, J. J. & Walker, M. J. C. (1997). *Reconstructing Quaternary Environments 2nd eds*. Longman Group Ltd: Essex

Lozano, J., Virgós, E., Malo, A. F., Huertas, D. L., Casanovas, J. G. (2003). Importance of scrub-pastureland mosaics for wild-living cats occurrence in a Mediterranean area: implications for the conservation of the wildcat (Felis silvestris). *Biodiversity and Conservation* **12**: 921-935

Lozano, J., Virgós., Cabezas-Díaz, S. (2013). Monitoring European wildcat *Felis silvestris* populations using scat surveys in central Spain: are population trends related to wild rabbit dynamics or to landscape features? *Zoological Studies* **52**: **DOI:** 10.1186

Lugowski, A & Ogg, J. (2012). TimeScale Creator v.6.1.2. https://engineering.purdue.edu/Stratigraphy/tscreator/index/index.php

Lüthi, D., Le Floch, M., Bereiter, B., Blunier, T., Barnola, J-M., Siegenthaler, U., Raynaud, D., Jouzel, J., Fischer, H., Kawamura, K., Stocker, T. F. (2008). High-resolution carbon dioxide concentration record 650,000-800,000 years before present. *Nature* **453**: 379-382

Lyons, S. K., Smith, F. A., Brown, J. H. (2004). Of mice, mastodons and men: human-mediated extinctions on four continents. E*volutionary Ecology Research* **6**: 339-358

MacDonald, D. W., Yamaguchi, N., Kitchener, A. C., Daniels, M., Driscoll, C. (2010). *Reversing cryptic extinction: the history, present and future of the Scottish Wildcat*. In: MacDonald D. W., Loveridge, A. J. (Eds). Biology and conservation of wild felids. Oxford University press: Oxford

MacFadden, B. J. (2000). Middle Pleistocene climate change recorded in fossil mammal teeth from Tarija, Bolivia, and upper limit of the Ensenadan land-mammal age. *Quaternary Research* **54**: 121-131

Macleod, N & Forey, P. (2002) *Morphology, shape and phylogenetics*. Taylor & Francis: London

MacLeod, N & Rose, K. D. (1993). Inferring locomotor behaviour in Palaeogene mammals via eigenshape analysis. *American Journal of Science* **293-A**: 300-355

Macleod, N. (1999). Generalizing and extending the eigenshape method of shape space visualization and analysis. *Paleobiology* **25**:107-138

MacLeod, N. (2001). Landmarks, localization and the use of morphometrics in phylogenetic analysis. In: *Fossils, phylogeny and form* (eds: Adrian, J. M., Edgecombe, G. D., Leiberman, B. S). Kluwer Academic: New York

Macleod, N. (2005). Principal components analysis (eigenanalysis & regression 5). *Palaeontological Association Newsletter* **59**: 42-54

MacLeod, N. (2007a). Groups I. *Palaeontological Association Newsletter* **64**: 35-45

Macleod, N. (2007b). Groups II. *Palaeontological Association Newsletter* **65:** 36-49

Macleod, N. (2009a). Who is Procrustes and what has he done with my data? *Palaeontological Newsletter* **70**: 21-36

Macleod, N. (2009b). Shape Theory. *Palaeontological Association Newsletter* **71**: 34-47

Macleod, N. (2009c). Form & shape models. *Palaeontological Association Newsletter* **72**: 14-27

MacLeod, N. (2012). Going round the bend II: Extended eigenshape analysis. *Palaeontological Association Newsletter* **81**: 23-39

MacLeod, N. (2013). A picture may be worth 1000 landmarks. *Palaeontological Association Newsletter* **84**: 20-30

MacLeod, N. (2016). On the use of machine learning in morphometric analysis. *In press*

Maddison, W. P. and D.R. Maddison. (2016). Mesquite: a modular system for evolutionary analysis. Version 3.10 http://mesquiteproject.org

Manley, B. F. J. (1994). *Multivariate statistical methods: a primer*. Chapman & Hall: Suffolk

Manning, A. D., Coles, B. J., Lunn, A. G., Halley, D. J., Ashmole, P., Fallon, S. J. (2014). New evidence of late survival of beaver in Britain. *The Holocene* **24:** 1849-1855

Markova, E., Beeren, Z., van Kolfschoten, T., Strukova, T., Vrieling, K. (2012). Differentiating sibling species in the Quaternary fossil record: a comparison of morphological and molecular methods to identify *Microtus arvalis* and *Microtus rossiaemeridionalis* (Arvicolinae, Rodentia). *Journal of Systematic Palaeontology* **10**: 585-597

Martíková, N & Moravec, J. (2012). Multilocus phylogeny of arvicoline voles (Arvicolini, Rodentia) shows small tree terrace size. *Folia Zoologica* **61**:254-267

Martíková, N., Barnett, R., Cucchi, T., Struhen, R., Pascal, M., *et al.* (2013). Divergent evolutionary processes associated with colonization of offshore islands. *Molecular Ecology* **20**: 5205-5220

Martin, P. S., Wright, E. J (eds). (1967). *Pleistocene extinctions*: *the search for a cause*. Proc. 7th congress of the Int. Association for Quaternary Research, vol. 6. New Haven, CT: Yale University Press

Martin, R.A. (1993). Patterns of variation and speciation in Quaternary rodents. In: R.A. Martin, A.D. Barnosky (eds.), *Morphological change in Quaternary mammals of North America*, 227–391. Cambridge University Press

Martínková, N., Barnett, R., Cucchi, T., Struhen, R., Pascal, M., et al. (2013). Divergent evolutionary processes associated with colonization of offshore islands. Molecular *Ecology* **20**: 5205-5220

Maslin, M. **A & Ridgwell, A. J**. (2005). Mid-Pleistocene revolution and the 'eccentricity myth'. *In:* Head, M. J and Gibbard, P. L. (Eds), *Early-Middle Pleistocene transitions: the land-ocean evidence: Geological Society of London, Special Publication* **247**, pp. 19-34

Mattucchi, F., Oliveira, R., Lyons, L. A., Alves, P. C., Randi, E. (2016). European wildcat populations are subdivided into five main biogeographic groups: consequences of Pleistocene climate changes or recent anthropogenic fragmentation? *Ecology and Evolution* **6**: 3-22

Mattucci, F. (2014). Conservation genetics of European Wildcat (*Felis silvestris silvestris*): a wide and integrating analysis protocol for admixture inferences and population structure. PhD Thesis: University of Bologna Digital Library

Meyer, M. & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbour Protocols* **2010**: 1-7

McDevitt, A.D., Zub, K., Kawalko, A., Oliver, M. K., Herman, J. S., *Wójcik, J. M.* *(2012).* Climate and refugial origin influence the mitochondrial lineage distribution of weasels Mustela nivalis in a phylogenetic suture zone. *Biological Journal of the Linnaean Society* **106**: 57-69

McEwing, R., Kitchener, A. C., Holleley, C., Kilshaw, K., O'Donoghue, P. (2012). An allelic discrimination SNP assay for distinguishing the mitochondrial lineages of European wildcats and domestic cats. *Conservation Genetics Resources* **4**: 163-165

McEwing, R., Senn, H., Campbell-Palmer, R. (2015). Genetic assessment of free-living beavers in and around the River Tay catchment, east Scotland. *Scottish Natural Heritage Commissioned Report No. 682*

McGuire, J. A., Linkem, C. W., Koo, M. S., Hutchison, D. W., Lappin, A. K., Orange, D. I., Lemos-Espinol, J., Riddle, B. R., Jaeger, J. R. (2007). Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotaphytid lizards. *Evolution* **61**: 2879-2897

McGuire, J. L. (2010). Geometric morphometrics of vole (*Microtus californicus*) dentition as a new palaeoclimate proxy: Shape change along geographic and climate clines. *Quaternary International* **212**:198-205

McGuire, J. L. (2011). Identifying Californian *Microtus* species using geometric morphometrics documents Quaternary geographic range contractions. *Journal of Mammalogy* **92**: 1383-1394

McNab, B. K. (1971). On the ecological significance of Bergmann's Rule. *Ecology* **52**: 845-845

Meyer, M. & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbour Protocols* **2010**: 1-7

Meyer, M. N., Golenishchev, F. N., Bulativa, N. S., Artobolevsky, G. V. (1997). On distribution of two *Microtus arvalis* chromosomal forms in European Russia. *Zoologicheskii zhurnal* **76**: 487-493

Michaux, J. R., Magnanou, E., Paradis, E., Neiberding, C., Libois, R. (2003). Mitochondrial phylogeography of the woodmouse (*Apodemus sylvaticus*) in the Western Palearctic region. *Molecular Ecology* **12**: 685-697

Milankovitch, M. (1941). 'Kanon der Erdbestrahlung und seine Anwendung auf das Eiszeiten problem (Canon of Insolation and the Ice-Age problem).' *Royal Serbian Sciences, Special Publications* **132**: 484

Milishnikov, A. N. (2004). Population-genetic structure of beaver (*Castor fiber L.*, 1758) communities and estimation of effective reproductive size N-e of an elementary population. *Russian Journal of Genetics* **40**:772–781.

Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010) "Creating the CIPRES Science Gateway for inference of large phylogenetic trees" in *Proceedings of the Gateway Computing Environments Workshop* (GCE), 14 Nov. 2010, New Orleans, LA pp 1 - 8

Miller, W., Drautz, D. I., Ratan, A. *et al.* (2008). Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* **456**: 387-390

Mishmar, D., Ruiz-pesini, E., Golik, P. *et al.* (1999). Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Sciences of the USA* **100**: 171-176

Mitchell-Jones, A. J., Amori, G., Bogdanowicz, W., Krystufek, B., Reijnders, P. J. H., Spitzenberger, F., Stubbe, M., Thissen, J. B. M., Vohralík, V., Zima, J: *The atlas of European mammals*. London: Academic Press; 1999.

Mix, A. C., Bard, E., Schneider, R. (2001). Environmental processes of the ice age: land, oceans, glaciers (EPILOG). *Quaternary Science Reviews* **20**: 627-657

Moen, D. S., Irschick, D. J., Wiens, J. J. (2013). Evolutionary conservatism and convergence both lead to striking similarity in ecology, morphology and performance across continents in frogs. *Proceedings of the Royal Society B* **280**: 20132156

Morin, P. A., Archer, F. I., Foote, A. D., Vilstrup, J., Allen, E. E., Wade, P., Durban, J., Parsons, K., Pitman, R., Li, L., Bouffard, P., Abel-Neilsen, S. C., Rasmussen, M., Willerslev, E., Gilbert, M. T., Harkins, T. (2010). Conplete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Research* **20**: 908-916

Moritz, C. (1994). Applications of mitochondrial DNA analysis in conservation – a critical review. *Molecular Ecology* **3**: 401-411

Mosimann, J. E. (1970). Size allometry: size and shape variables with characterizations of lognormal and generalized gamma distributions. *Journal of the American Statistical Association* **65**: 930-945

Müller-Schwarze, D and Sun, Lixing. (2003). *The beaver: Natural History of a Wetlands Engineer.* Cornell University Press: US

Muñoz-Muñoz, F & Perpiñán, D. (2010). Measurement error in morphometric studies: comparison between manual and computerized methods. *Annales Zoologici Fennici* **47**: 46-56

Musser, G. G., Carleton, M. D. (2005). Superfamily Muroidea. In: *Mammal species of the world: a taxonomic and geographic reference, 3rd Ed.* (Eds Wilson, D. E. & Reeder, A. M). pp. 894-1531. John Hopkins University Press: Baltimore

Nadachowski, A & Zagorodnyuk, I. (1996). Recent *Allophaiomys*-like species in the Palaearctic: Pleistocene relicts or a return to an initial type? *Acta Zoologica Cracoviensia* **39**: 387-394

Nadachowski, A. (1984). Taxonomic value of anteroconid measurements of M_1 in common and field voles. *Acta Theriologica* **10**:123-143

Navarro, N., Zatarain, X., Montuire, S. (2004). Effects of morphometric descriptor changes on statistical classification and morphospaces. *Biological Journal of the Linnaean Society* **83**: 243-260

Nielsen, R and Weinreich, D. M. (1999). The age of nonsynonymous and synonymous mutations in animal mtDNA and implications for the mildly deleterious theory. *Genetics* **153**: 497-560

Nolet, B. A. & Baveco, J. M. (1996). Development and viability of a translocated beaver *Castor fiber* population in the Netherlands. *Biological Conservation* **75**: 125-137

Nolet, B. A., Dijkstra, A. V. V., Heidecke, D. (1994). Cadmium in beavers translocated from the Elbe River to the Rhine/Meuse estuary and the possible effect on population growth rate. *Archives of Environmental Contamination and Toxicology* **27**: 154-161

Nolet, B., & Rosell, F. (1998). Comeback of the beaver *Castor fiber*: an overview of old and new conservation problems. *Biological Conservation* **83**: 165-173

Nowell, K and Jackson, P. (1996). Wildcats. Status, survey and conservation plan. Gland: IUCN

O'Brian, J., Devillard, S., Say, L., Vanthomme, H., Léger, F., Ruette, S., Pontier, D. (2009). Preserving genetic integrity in a hybridising world: are European Wildcats (*Felis silvestris silvestris*) in eastern France distinct from sympatric feral domestic cats? *Biodiversity Conservation* **18**: 2351-2360

Oliveira, R., Godinho, R., Randi, E., Alves, P. C. (2008b). Hybridization versus conservation: are domestic cats threatening the genetic integrity of wildcats

(*Felis silvestris silvestris*) in Iberian Peninsula? *Philosophical Transactions of the Royal Society B* **363**: DOI: 10.1098

Oliveira, R., Godinho, R., Randi, E., Ferrand, N., Alves, P. C. (2008a). Molecular analysis of hybridisation between wild and domestic cats (*Felis silvestris*) in Portugal: implications for conservation. *Conservation Genetics* **9**: 1-11

Oliveira, R., Randi, E., Mattucci, F., Kurushima, J. D., Lyons, L. A., Alves, P. C. (2015). Toward a genome-wide approach for detecting hybrids: informative SNPs to detect introgression between domestic cats and European wildcats (*Felis silvestris*). *Heredity* **115**: 195-205

Ollivier, M., Tresset, A., Hitte, C., *et. al.* (2013). Evidence of coat colour variation sheds new light on ancient canids. PLOSone http://dx.doi.org/10.1371/journal.pone.0075110

Oppenheimer, S. (2006). *The origins of the British: a genetic detective story*. London, UK: Constable and Robinson

Ord, T. J. and Klump, D. A. (2014). Habitat partitioning and morphological differentiation: the Southeast Asian Draco lizards and Caribbean Anolis lizards compared. *Oecologia* **175**: 651-666

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

Orlando, L., Bonjean, D., Bocherens, H., *et al.* (2002). Ancient DNA and the population genetics of cave bears (*Ursus spelaeus*) through space and time. Molecular *Biology and Evolution* **19**: 1920-1933

Orlando, L., Ginolhac, A., Zhang, G., *et al*. (2013). Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**: 74-78

Pääbo, S. (1985). Molecular cloning of Ancient Egyptian mummy DNA. *Nature* **341**: 644-645

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

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., Hofreiter, M. (2004). Genetic analysis from ancient DNA. *Annual Review of Genetics* **38**: 645-679

Palkopoulou, E., Baca, M., Abramson, N. I., *et al*. (2016). Synchronous genetic turnovers across Western Eurasia in late Pleistocene collared lemmings. *Global Change Biology* **22**: 1710-1721

Palkopoulou, E., Mallick, S., Skoglund, P., *et al.* (2015). Complete genomes reveal signatures of demographic genetic declines in the Woolly Mammoth. *Current Biology* **25**: 1395-1400

Pearson, P. G. & Dawson, T. P. (2003). Predicting the effects of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography* **12**: 361-371

Peltier, W. R & Fairbanks, R. G. 2006. Global glacial ice volume and last Glacial Maximum duration from an extended Barbados sea level record. *Quaternary Science Reviews* **25**: 3322-3337

Perez, S. I., Bernal., V., Gonzalez, P. N. (2006). Differences between sliding semi-landmark methods in geometric morphometrics, with an application to human craniofacial and dental variation. *Journal of Anatomy* **208**: 769-784

Pfunder, M., Holzgang, O., Frey, J. E. (2004). Development of microarray-based diagnostics of voles and shrews for use in biodiversity monitoring studies, and evaluation of mitochondrial cytochrome oxidase I vs. cytochrome b as genetic markers. *Molecular Ecology* **13**: 1277-1286

Phillips, F. M., Bowen, D. Q., Elmore, D. (1994). Surface exposure dating of glacial features in Great Britain using cosomogenic chlorine-36: preliminary results. *Mineralogical Magazine* **58A**: 722-723

Pickrell, J and Reich, D. (2014). Towards a new history and geography of human genes informed by ancient DNA. *Trends in Genetics* **30**: 377-389

Piechocki, R. (1977). Ökologische Todesursachenforschung am Elbebiber (*Castor fiber albicus*). *Beiträge zur Jagd- und Wildforschung* **10**: 332-341

Pierpaoli, M., Biró, Z. S., Herrmann, M., Hupe, K., Fernandes, M., Ragni, B., Szemethy, L., Randi, E. (2003). Genetic distinction of wildcat (Felis silvestris) populations in Europe and hybridization with domestic cats in Hungary. *Molecular Ecology* **12**: 2585-2598

Piertney S. B., Stewart W. A., Lambin X., Telfer S., Aars J., Dallas J. F. (2005) Phylogeographic structure and postglacial evolutionary history of water voles (*Arvicola terrestris*) in the United Kingdom. *Molecular Ecology* **14**: 1435–1444.

Pinhasi, R., Fernandes, D., Sirak, K., Novak, *et al.* (2015). Optimal ancient DNA yields from the inner ear part of the human petrous bone. PLoSONE: http://dx.doi.org/10.1371/journal.pone. 0129102

Piras, P., Marcolini, F., Raia, P., Curicio, M., Kotsakis, T. (2010). Ecophenotypic variation and phylogenetic inheritance in first lower molar shape of extant Italian populations of *Microtus* (Terricola) *savii* (Rodentia). *Biological Journal of the Linnaean Society* **99**: 632-647

Pitti, C and Tozzi, C. (1971). La Grotto del Capriolo e la Buca della Lenna Presso Mommio (Camaiore, Lucca). *Rivista di Scienze Preistoriche* **262**: 213-258

Plummer, T. W., Bishop, L. C., Hertel, F. (2008). Habitat preference of extant African bovids based on astragalus morphology: operationalizing ecomorphology for palaeoenvironmental reconstruction. *Journal of Archaeological Science* **35:**3016-3027

Polak, M & Starmer, W. T. (2001). The quantitative genetics of fluctuating asymmetry. *Evolution* **55**: 498-511

Polly, D. P., Killick, L., Ruddy, M. (2011). Using left-right asymmetry to estimate non-genetic variation in vole teeth (Arvicolinae, Muridae, Rodentia). *Palaeontologica Electronica* **14**: 1-12

Polly, P. D, Lawing, A. M., Fabre. A-C., Goswami, A. (2013). Phylogenetic principal components analysis and geometric morphometrics. *Hystrix Italian Journal of Mammalology* **24**: 33–41.

Polly, P. D. & Eronen, J. T. (2011). Mammal associations in the Pleistocene of Britain: implications of ecological niche modelling and a method for reconstructing palaeo-climate. In: *The Ancient Human Occupation of Britain* pp. 410 – 420. (Eds. Ashton, N. M., Lewis, S. G., Stringer, C. B). *Developments in Quaternary Science* **14***:* 165 - 180

Pontius, J. U., Mullikin, J. C., Smith, D. R., Lindblah_Toh, K., Gnerre, S, *et al.* (2007). Initial sequence and comparative analysis of the cat genome. *Genome Resources* **17**: 1675-1689

Posada, D., & Crandall, K. A. (2001). Intraspecific gene geneologies: trees grafting into networks. *Trends in Ecology and Evolution* **16:** 37-45

Pozdnyakov, A. A. (1993). Morphotypic variability of grinding surfaces of molars of voles of the Maximoviszi group (Rodentia, Arvicolidae: experience of quantitative statistical analysis. *Zhurnal Obschchei Biologii* **72**: 127-139

Pritchard, J., Stephens, M., Donelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945-959

Prűfer, K., Racimo, F., Patterson, N. *et al.* (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**: 43-49

Pruvost, M., Bellone, R., Benecke, N., *et al*. (2011). Genotypes of predomestic horses match phenotypes painted in Paleolithic works of cave art. *Proceedings of National Academy of Sciences of the United States of America* **108**: 18626-18630

Ragni, B. (1993). The crucial problem of *in vivo* identification of wildcat and recognition of hybrids with domestic cats. **In:** *Seminar on biology and conservation of the wildcat* (*Felis silvestris*). Council of Europe, Environmental Encounters No. 16, Strasbourg

Rambaut, A. & Drummond, A. (2010). *FigTree v1. 3.1. Institute of Evolutionary Biology, University of Edinburgh*

Rambaut. A., Suchard, M. A, Xie, D., Drummond, A. J. (2014). Tracer v1.6, Available from http://beast.bio.ed.ac.uk/Tracer

Ramos-Onsins, S. E & Rozas, J. (2002). Statistical properties of new neutrality tests against populations growth. *Molecular Biology* **19**: 2092-2100

Rand, D. M. (2001). The units of selection on mitochondrial DNA. *Annual Review of Ecology and Systematics* **32**: 415-448

Rasmussen, S. O., Vinther, B. M., Clausen, H. B. *et al.* (2007). Early Holocene climatic oscillations recorded in three Greenland ice cores. *Quaternary Science Reviews* **26**: 1907-1914

Ravdonikas, V. J. (1936). *Naskal'nye Izobrazhenia Onezhskogo Ozera*. I. Moskva-Leningrad: Moscow-Leningrad

Reed, D. H., & Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conservation Biology* **17**: 230-237

Reich, D., Green, E. R., Kircher, M., *et al.* (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**: 1053-1060

Reimer, P., Bard, E., Bayliss A.*, et al.* (2013). Intcal13 and Marine13 radiocarbon age calibration curves 0-50,000 years cal BP. *Radiocarbon* **55**: 1869-1887

Reisch, L. (1976). Beobachtungen un Vogelknochen aus dem Spätpleistozän der Höhle Kephalari (Argolis, Griechenland). *Archäologisches Korrespondenzblatt* **6**: 261-265

Reiss, R. A. 2006. Ancient DNA from ice age insects: proceed with caution. *Quaternary Science Reviews* **25**:1877-1893

Renaud, S., Michaux, J., Schmidt, D. N., Aguilar, J-P., Mein, P., Auffray, J-C. (2005). Morphological evolution, ecological diversification and climate change in rodents. *Proceedings of the Royal Society B* **272**: 609-617

Repenning, C. A. (1968). Mandibular musculature and the origin of the subfamily Arvicolinae (Rodentia). *Acta Zoological Cracoviensia* **113**:29-72

Reyment, R. A. (1996). An idiosyncratic history of early morphometrics. In: Marcus L.F., Corti M., Loy A., Naylor G.J.P., Slice D.E. (Eds.). *Advances in Morphometrics*. Plenum Press, New York, pp. 15–22

Řičánková, V. P., Robovsky, J., Riegert, J., Zrzavy, J. (2015). Regional patterns of postglacial changes in the Palearctic mammalian diversity indicate retreat to Siberian steppes rather than extinction. *Scientific Reports* **5**: 12682

Richard, P. B. (1985). Peculiarities on the ecology and management of the Rhodanian Beaver (*Castor fiber* L.). *Zeitschrift fuer Angewandte Zoologie* **72**: 143-152

Ritchie, J. (1920). *The Influence of Man on Animal Life in Scotland: A Study of Faunal Evolution.* Cambridge University Press: Cambridge

Roberts, M. B & Parfitt, S. A. (1999). *Boxgrove: A middle Pleistocene hominid site at Eartham Quarry, Boxgrove, West Sussex*. English Heritage: London

Rodrigues, M., Bos, A. R., Hoath, R., Fernandez, C. (2016). taxonomic status and origin of the Egyptian weasel (*Mustela subpalmata*) inferred from mitochondrial DNA. *Genetica* **144**: DOI: 10.1007

Rodríguez, F., Oliver, J. L., Marín, A., Medina, J. R. (1990). The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology* **142**: 485-501

Rodríguez, R., Ramírez, O., Valdiosera, C. E. *et al.* (2011). 50,000 years of genetic uniformity in the critically endangered Iberian lynx. *Molecular Ecology* **20**: 3785-3795

Rogaev, E. I., Moliaka, Y. K., Malyarchuk, B. A., Kondrashov, F. A, Deremko, M. V. *et al.* (2006). Complete mitochondrial genome and phylogeny of Pleistocene mammoth Mammuthus primigenius. *PloS Biology* **4**: e73

Rohland, N & Hofreiter, M. (2007). Ancient DNA extraction from bones and teeth. *Nature Protocols* **2**: 1756-1762

Rohlf, F. J and Corti, M. (2000). The use of partial least-squares to study covariation in shape. *Systematic Biology* **49**: 740-753.

Rohlf, F. J and Marcus L. F. (1993). A revolution in morphometrics. *Trends in Ecology and Evolution* **8**: 129-132.

Rolf, F. J & Slice, D. E. (1990). Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology* **39**:40-59

Römpler, H., Rohland, N., Lalueza-Fox, C. *et al.* (2006). Nuclear gene indicates coat-colour polymorphism in Mammoth. *Science* **313**: DOI: 10.1126/science.1128994

Rosell, F., & Pederson, K. V. (1999). *Bever. Lendbruksforlaget*: Oslo, Norway

Rosell, F., Campbell-Palmer, R., Parker, H. (2012). More genetic data are needed before populations are mixed: response to 'Sourcing Eurasian beaver *Castor fiber* stock for re-introductions in Great Britain and Western Europe'. *Mammal Review* **42**: 319-324

Rozas, J. (2009). DNA Sequence Polymorphism Analysis using DnaSP. Pp. 337-350.*In* Posada, D. (ed.) *Bioinformatics for DNA Sequence Analysis; Methods in Molecular Biology Series* Vol. **537**

Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V., Wallace, D. C. (2004). Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* **303**: 223-226

Saccone, C., Pesole, G., Sbisá, E. (1991). The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. *Journal of Molecular Evolution* **33**: 83-91

Saitou, N & Nei, M. (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406-425

Sanchez, J. J & Endicott, P. (2006). Developing multiplexed SNP assays with special reference to degraded DNA templates. *Nature Protocols* **1**: 1370-1378

Sandom, C., Faurby, S., Sandel, B., Svenning, J-C. (2014). Global late Quaternary megafauna extinctions linked to humans, not climate change. *Proceedings of the Royal Society of London, Series B, Biological Sciences* **281**: 20133254. http://dx.doi.org/10.1098/rspb.2013.3254

Santos, S. M., da Luz Mathias, M., Mira, A. P. (2011). The influence of local, landscape and spatial factors on the distribution of the Lusitanian and the Mediterranean pine voles in a Mediterranean landscape. *Mammal Biology* **76**: 133-1425

Saveljev, A. P. (2001). Rettung des Bibers (*Castor fiber*) in Russland: offensichtlicher jagdwirtschaftlicher Erfolg mit zoologischen Problemen nach 70 Jahren. *Beiträge zur Jagd- und Wildforschung* **26**: 309-315

Savelijev, A. P. (2003). *Biological peculiarities of aboriginal and artificially created beaver populations in Eurasian and their significance for the resource management strategy.* Dissertation of Doctorate in Biological Science, Russian Research Institute of Game Management and Fur farming, Kirov.

Saveljev, A. P., & Milishnikov, A. (2002). Biological and genetic peculiarities of cross-composed and aboriginal beaver populations in Russia. *Acta Zoologica Lituanica* **12**: 397-402

Schindelin, J., Arganda-Carreras, I., Frise, E. *et al.* *(2012).* Fiji: an open-source platform for biological-image analysis. *Nature Methods **9**: 676-682*

Schreve, D. C. (1996). The mammalian fauna from the Waechter excavations. In: B. Conway, N. M. Ashton., & J. McNabb (eds). Excavations at Barnfield Pit, Swanscombe 1968-72. *Occasional Paper of the British Museum* **94**: 149-162

Schreve, D. (2001). Differentiation of the British Late Middle Pleistocene interglacials: The evidence from mammalian biostratigraphy**.** *Quaternary Science Reviews* **20**: 1693-1705

Schreve, D. C. (2013). Seventh Interim Report on Excavations at Gully Cave, Ebbor Gorge, Somerset. Department of Geography: Royal Holloway University of London (unpublished)

Schreve, D. **C**. (2014), Eighth Interim Report on Excavations at Gully Cave, Ebbor Gorge, Somerset. Department of Geography: Royal Holloway University of London (unpublished)

Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J. F., AL-Rasheid, K. AS., Willerslev, E., Krigh, A., Orlando, L. (2012). Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**: 178: **DOI:** 10.1186

Schubert, M., Jónsson, H., Chang, D., *et al*. (2014). Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proceedings of National Academy of Sciences of the United States of America* **111**: E5661 – E5669

Searle, J. B. and Wilkinson P. J. (1987). Karyotypic variation in the common shrew (*Sorex araneus*) in Britain—a 'Celtic fringe'. *Heredity* **59**: 345–351

Searle, J. B., Kotlík, P., Rambau, R. V., Marková, S., Herman, J. S., McDevitt, A. D. (2009). The Celtic fringe of Britain: insights from small mammal phylogeny. *Proceedings of the Royal Society B* **276**: DOI: 10.1098

Seifert, B. (2008). Removal of allopatric variation improves species separation in multicharacter discriminant functions when species are strongly allometric and exposes diagnostic characters. *Myrmecological News* **11**: 91-105

Senn, H. V and Ogden, R. (2015). *Wildcat hybrid scoring for conservation breeding under the Scottish Wildcat Conservation Action Plan (2015).* Royal Zoological Society of Scotland

Senn, H., Ogden, R., Frosch, C., Syrůčková, A., Campbell-Palmer, R., Munclinger, P., Durka, W., Kraus, R. H. S., Saveljev, A. P., Nowak, C., Stubbe, A., Stubbe, M., Michaux, J., Lavrov, V., Samiya, R., Ulevicius, A., Rosell, F. (2014). Nuclear and mitochondrial genetic structure in the Eurasian beaver (*Castor fiber*) – implications for future reintroductions. *Evolutionary Application* **7**: 645-662

Shackleton, N. J & Opdyke, N. D. (1973). Oxygen isotope and palaeomagnetic stratigraphy of equatorial pacific core V28 – 238: Oxygen isotope temperatures and ice volumes on a 10^5 year and 10^6 year scale. *Quaternary Research* **3**:39-55

Sievwright, H & MacLeod, N. (2012). Eigensurface analysis, ecology and modelling of morphological adaptation in the falconiform humerus (Falconiformes: Aves). *Zoological Journal of the Linnean Society* **165:**390-419

Skoglund, P., Ersmark, E., Palkopoulou, E., Dalén, L. (2015). Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Current Biology* **25**: 1515-1519

Slatkin., M & Hudson, R. R. (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*. **129**:555–562

Slice, D. E. (2001). Landmark coordinates aligned by Procrustes analysis do not lie in Kendall's shape space. *Systematic Biology* **50**: 141–149

Smith, F. A., Tomé, C. P., Smith, E. A. E., Lyons, S. K., Newsome, S. D., Stafford, T. W. (2016). Unraveling the consequences of the terminal Pleistocene megafauna extinction on mammal community assembly. *Ecogeography* **39**: 223-239

Smith, G. R. (1990). Homology in morphometrics and phylogenetics. In: *Proceedings of the Michigan Morphometrics Workshop* (eds: Rohlf, F. J; Bookstein, F. L.). The University of Michigan Museum of Zoology: Michigan

Sneath, P. H. A. (1967). Trend surface analysis of transformation grids. *Journal of Zoology* 151: 65–122.

Soberón. J. (2007). Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters* **10**: 1115-1123

Sommer, R. S., and Benecke, N. (2006). Late Pleistocene and Holocene development of the felid fauns (Felidae) of Europe: a review. *Journal of Zoology* **269**: 7-19

Spitz, R. (1985). Further development of the forecasting model for *Microtus arvalis*. *Acta Zoologica Fennica* **173**: 89-93

Springer, M. S., Debry, R. W., Douady, C., Amrine, H. M., Madsen, O., de-Jong, W. W., Stamhope, M. J. (2001). Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogenetic reconstruction. *Molecular Biology and Evolution* **18**: 132-143

Staesche, U. (1994). Die terreste aud den Buntsandsteinabris im Leinebergland bie Göttinggen. *Veröffentlichungen der Urgeschichtlichen Sammlungen des landesmuseums zu Hannover* **43**: 101-126

Stahl, P and Arois, M. (1991). *Status and conservation of the wild cat 9Felis silvestris) in Europe and around the Mediterranean rim.* Strasbourg, France: Council of Europe; 1991:61

Stewart, J. R and Lister, A. M. (2001). Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology and Evolution* **16**: 608-613

Stewart, J. R. (2008). The progressive effect of the individualistic response of species to Quaternary climate change: an analysis of British mammalian faunas. *Quaternary Science Reviews* **27**: 2499-2508

Stewart, J. R., Lister, A. M., Barnes, I., Dalèn, L. (2010). Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B* **277**: 661-671

Steyer, K., Kraus, R. H. S., Mölich, T. *et al.* (2016). Large-scale genetic census of an elusive carnivore, the European wildcat (*Felis. s. silvestris*). *Conservation Genetics* **17**: 1183-1199

Stojak, J., McDevitt, A. D., Herman, J. S., Searle, J. B., Wójcik, J. M. (2015). Post-glacial colonization of Eastern Europe from the Carpathian refugium: evidence from mitochondrial DNA of the common vole *Microtus arvalis*. *Biological Journal of the Linnaean Society* **115**: 927-939

Stott. L., Timmermann. A., Thunell. R. (2007). Southern Hemisphere and Deep-Sea Warming Led Deglacial Atmospheric CO2 Rise and Tropical Warming. *Science* **318**: 435-438

Sunquist, M. E and Sunquist, F. (2002). *Wildcats of the World*. University of Chicago Press: Chicago

Sutcliffe, A. J. (1960). 'Joint Minor Cave, Buckfastleigh', *Transactions and proceedings of the Torquay Natural History Society* **13**: 1-26

Sutcliffe, A. J. (1964). 'The mammalian fauna,' 85-111 in C. D. Ovey (Ed). *The Swanscombe Skull: A Survey of Research on a Pleistocene Site*. Royal Anthropological Institute of Great Britain and Ireland, London

Sutcliffe, A. J. and Zeuner, F. E. (1962). 'Excavations in the Torbryan Caves, Devonshire. I. Tornewton Cave'. *Proceedings of the Devon Archaeological and Exploration Society* **5**: 127-145

Svendsen, J. I., Alexanderson, H., Astakhov, V. I., Demidov, I., Dowdeswell, J. A., Funder, S., Gataullin, V., Henriksen, M., Hjort, C., Houmark-Nielsen, M., Hubberten, H. W., Ingólfsson, Ó., Jakobsson, M., Kjaer, K. H., Larsen, E., Lokrantz, H., Lunkka, J. P., Lyså, A., Mangerud, J. (2004). Late Quaternary ice sheet history of northern Eurasia. *Quaternary Science Reviews* **23**:1229-1271

Svendsen, J. I., Alexanderson, H., Astakhov, V. I., *et al*. (2004). Late Quaternary ice sheet history of northern Eurasia. *Quaternary Science Reviews* **23**: 1229-1271

Svensson, A., Andersen, K. K., Bigler, M., Clausen, H. B., Dahl-Jenson, D., Davies, S. M., Johnsen, S. J., Muscheler, R., Parrenin, F., Rasmussen, S. O., Röthlisberger, R., Seierstad, I., Steffensen, J. P., Vinther, B. M. (2008). A 60 000 year Greenland stratigraphic ice core chronology. *Climate of the Past* **4:** 47-57

Szulkin, M., Bierne, N., David, P. (2010). Heterozygosity-fitness correlations: a time for reappraisal. *Evolution* **64**: 1201-1217

Tamazian, G., Simonov, S., Dobrynin, P *et al.* (2014). Annotated features of the domestic cat – *Felis catus*. *GigaScience* **3**: 13-14

Tanner, M. A and Wong, W. H (2010). From EM to data augmentation: the emergence of MCMC bayesian computation in the 1980s. *Statistical Science* **25**:506-516

Tarailo-Graovac, M and Chen, N. (2009). Using RepeatMasker to Identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics* 4.10.1-4.10.14

Taylor, R. E. (2012). *Radiocarbon dating: an archaeological perspective*. Harcourt Brace Jovanovich: London

Taylor, R. E and Bar-Yosef, O. (2014). *Radiocarbon dating: an archaeological perspective*. Routledge: New York

Teacher, A. G. F., Thomas, J. A., Barnes, I. (2011). Modern and ancient red fox (*Vulpes vulpes*) in Europe show an unusual lack of geographical and temporal structuring and differing responses within the carnivores to historical climatic change. *BMC Evolutionary Biology* **11**:214

Teaford, M. F., Smith, M. M., Ferguson, M. W. J. (2007). *Development, function and evolution of teeth*. Cambridge University Press: Cambridge

Templeton, J. E. L., Brotherton, P. M., Llamas, B., Soubrier, J., Haak, W., Cooper, A., Austin, J. J. (2013). DNA capture and next-generation sequencing can recover whole mitochondrial genomes from highly degraded samples for human identification. *Investigative Genetics* **4**: 26

Thalmann, O., Shapiro, B., Cui, P. *et al.* (2013). Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science* **342**: 871-874

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

Thomas, E. R., Wolff, E., Mulvaney, R., Steffensen, J. P., Johnson, S. J., Arrowsmith, C., White, J. W. C., Vaugn, B., Popp, T. (2007). The 8.2 ka BP event from Greenland ice cores. *Quaternary Science Reviews* **26**: 70-81

Tobe, S. S., Kitchener, A., Linacre, A. (2009). Cytochrome b or cytochrome c oxidase subunit I for mammalian species identification - An answer to the debate. *Forensic Science International: Genetics Supplement Series* **2**: 306-307

Tougard, C., Renvoisé, E., Petitjean, A., Quéré, J. P. (2008). New insight into the colonization processes of common voles: inferences from molecular and fossil evidence. *PLoS ONE* **3**: e3532

Ungar, P. (2010). *Mammal teeth: Origin, evolution and diversity*. The John Hopkins University Press: Baltimore

Van der Meulen, A. J. (1973). Middle Pleistocene smaller mammals from the Monte Peglia, (Orvieto, Italy) with special reference to the phylogeny of *Microtus* (Arvicolinae, Rodentia). *Quaternaria* **17**:1-144

van Dyke, D.A and Meng, X-L. (2001) The art of data augmentation. *Journal of Computational and Graphical Statistics* **10**: 1–50.

van Klinken, G. J. (1999). Bone collagen quality indicators for palaeodietary and radiocarbon measurements. *Journal of Archaeological Science* **26:** 687-695

Vereshchagin, N. K & Baryshnikov, G. F. (1982). Palaeoecology of the mammoth fauna in the Eurasian arctic. In: Hopkins, D. M., Matthews, J. V., Schweger, C. E & Young, S. B (eds): *Palaeoecology of Beringia* pp. 267-279

Veron, G. (1992). Biogeographical history of the European beaver *Castor fiber* (Rodentia, Mammalia). *Mammals* **56**: 87-108

Vigne, J-D. (2011a). The origins of animal domestication and husbandry: A major change in the history of humanity and the biosphere. *Comptes Rendus Biologies* **334**: 171-181

Vigne, J-D., Carrere, I., Briois, F., Guilaine, J. (2011b). The early process of mammal domestication in the Near East. *Current Anthropology* **52**: S255-S271

Vigne, J-D., Guilaine, J., Debue, K., Haye, L., Gérard. (2004). Early taming of the cat in Cyprus. *Science* **304**: 259

Vinther, B. M., Clausen, H. B., Fischer, D. A., Koerner, R. M., Johnsen, S. J., Andersen, K. K., Dahl-Jensen, D., Rasmussen, S. O., Steffensen, J. P., Svensson, A. M. (2006). A synchronised dating of three Greenland ice cores throughout the Holocene. *Journal of Geophysical Research* **3**: 1-11

Waelbroeck, C., Labeyrie, L., Michel. E., Duplessy. J. C., McManus, J. F., Lambeck, K., Balbon. E., Labracherie. M. (2002) Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quaternary Science Reviews* **21**:295-305

Wainwright, P. C. (1991). Ecomorphology – experimental functional anatomy for ecological problems. *American Zoologist* **31**: 680-693

Walker , M. J. C., Bohncke, S. J. P., Coope. G. R., O'Connell, M., Usinger, H., Verbruggen, C. (1994). The Devensian/Weichselian Late-glacial in northwest Europe (Ireland, Britain, north Belgium, The Netherlands, northwest Germany). *Journal of Quaternary Science* **9**: 109-118

Walker, M. (2005). *Quaternary dating Methods*. John Wiley & Sons: Chichester

Walker, M. J. C., Berkelhammer, M., Björck, S., Cwyner, L. C., Fisher, D. A., Long, A. J., Lowe, J. J., Newnham, R. M., Rasmussen, S. O., Weiss, H. (2012). Formal subdivision of the Holocene Series/Epoch: a Discussion Paper by a working group of INTIMATE (Integration of ice-core, marine and terrestrial records) and the Subcommision on Quaternary Stratigraphy (International Commission on Stratigraphy). *Journal of Quaternary Science* **27**: 649-659

Walker, M. J. C., Björck, S., Lowe, J. J., Cwynar, L. C., Johnsen, K-L., Wohlfarth, B., INTIMATE group. (1999). *Quaternary Science Review*s 18: 1143-1150

Walker, M. J. C., Coope, G. R., Sheldrick, C., Turney, C. S. M., Lowe, J. J., Blockley, S. P. E & Harkness, D. D. (2003). Devensian Lateglacial environmental changes in Britain: a multi-proxy environmental record from Llanilid, South Wales, UK. *Quaternary Science Reviews* **22**: 475520

Wallace, D. C & Chalkia, D. (2013). Mitochondrial DNA Genetics and the heteroplasmy conundrum in evolution and disease. *Cold Spring Harbour Perspectives in Biology* **5**: a021220

Wang, X., Nyman, J. S., Dong, X., Leng, H., Reyes, M. (2010). *Fundamental biomechanics in bone tissue engineering*. Morgan & Claypool: DOI 10.2200/S00246ED1V01Y200912TIS004

Ward, G. K. and Wilson S. R. (1978). Procedures for comparing and combining radiocarbon age determinations: a critique. *Archaeometry* **20**: 19-31

Warne, F. (1981). *The wild-flower key: British-Isles – N.W. Europe*. Penguin Books Limited: England

Weber, G. W and Bookstein, F. L. *(*2011). *Virtual Anthropology. Springer:* Vienna

Webster, M and Sheets, H. D. (2010). *A practical introduction to landmark-based geometric morphometrics*. In: Alroy J, Hunt G, editors. *Quantitative Methods in Paleobiology Paleontological Society Papers*. 163–188

Weninger, B & Jöris, O. (2008). A 14C age calibration curve for the last 60 ka: the Greenland-Hulu U/Th timescale and its impact on understanding the Middle to Upper Palaeolithic transition in Western Eurasia. Journal of Human Evolution 55:772-781

Weninger, B., Schulting, R., Bradtmöller, M., Clare, L., Collard, M., Edinborough, K., Hilpert, J., Jöris, O., Niekus, M., Rohling, E. J., Wagner, B. (2008). The catastrophic final flooding of Doggerland by the Stroregga Slide tsunami. *Documenta Praehistorica* **35**: 1-24

Whiteford, N., Haslam, N., Weber, G., Prügel-Bennett, A., Essex, J. W. *et al.* (2005). Analysis of the feasibility of short read sequencing. *Nucleic Acids Resources* **33**: e171

Whitehouse, N. J. (2006). The Holocene British and Irish ancient forest fossil beetle fauna: implications for forest history, biodiversity and faunal colonisation. *Quaternary Science Reviews* **25**: 1755-1789

Wicks, K and Mithen, S. (2014). The impact of the abrupt 8.2 ka cold event on the Mesolithic population of western Scotland: a Bayesian chronological analysis using 'activity events' as a population proxy. *Journal of Archaeological Science* **45**: 240-269

Wilgenbusch, J. C., Warren, D. L., Swofford, D. L. (2004). AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. http://ceb.csit.fsu.edu/awty.

Willerslev, E. and Cooper, A. (2005). *Ancient DNA. Proceedings of the Royal Society B* **272**: DOI: 10.1098/rspb.2004.2813

Willerslev, E., Cappellini, E., Boomsma, W., *et al.* (2007). Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science* **317**: 111-114

Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T. P., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D. A., Cooper, A. (2003). Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* **300**: 791-795

Williams, J. W. & Jackson S. T. (2007). Novel climates, non-analogue communities and ecological surprises. *Frontiers in Ecology and the Environment* **5**: 475-482

Wilson, D. E & Reeder, D. M. (2005). *Mammal species of the world: a taxonomic and geographic reference,* 3rd Edition. John Hopkins University Press: Baltimore

Wisz, M. S., Pottier, J., Kissling, W. D*., et al.* (2013). The role of biotic interactions in shaping distributions and realized assemblages of species: implications for species distribution modelling. *Biological Reviews of the Cambridge Philosophical Society* **88**: 15-30

Wisz, M. S., Pottier, J., Kissling, W. D., Pellissier, L., Lenoir, J., Damgaard, C. F., Dormann, C. F., Forchhammer, M. C., Grytnes, J-A., Guisan, A., Heikkinen, R. K., Høye, T. T., Kühn, I., Luoto, M., Maiorana, L., Nilsson, M-C., Norman, S., Öckinger, E., Schmidt, N. M., Termansen, M., Timmermann, A., Wardle, D. A., Aastrup, P., Svenning, J-C. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews of the Cambridge Philosophical Society* **88**: 15-30

Wolfram Research, Inc. (2014). *Mathematica, Version 10.0*. Champaign: Illinois

Woods, R. (in press). *Using ancient DNA to understand the evolution of the Caribbean land mammal fauna.* PhD Thesis. Royal Holloway, University of London: Egham

Wójcik, J. M., Kawalko, A., Marková, S., Searle, J. B., Kotlík, P. (2010). Phylogeographic signatures of northward post-glacial colonization from high-latitude refugia: a case of bank voles using museum specimens. *Journal of Zoology* **281**: 249-262

Wolff, E. W, Chappellaz, J., Blunier, T., Rasmussen, S. O., & Svensson, A. (2010). Millennial-scale variability during the last glacial: The ice core record. *Quaternary Science Reviews* **29**: 2828-2838

Wolsan, M. (1993). Évolution des carnivores Quaternaires en Europe centrale dans leur context stratigraphe et paléoclimatique. *L'Anthropologie* **7**: 203-222

Woodward, S. R., Weyand, N. J., Bunell, M. (1994). DNA sequence from Cretaceous period bone fragments. *Science* **266**: 1229-1232

Xu, X. (1994). Evolution of the Chinese Castoridae. *National Science Museum Monographs* **8**: 77-97

Xue, Y., Prado-Martinez, J., Sudmant, P. H., *et al.* (2015). Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science* **348**: 242-245

Yalden, D. (1999). *The history of British mammals*. T. & A. D. Poyser, London

Yalden, D. W. (1982). When did the mammal fauna of the British Isles arrive? *Mammal Review* **12**: 1-57

Yamaguchi, N., Driscoll, C. A., Kitchener, A. C., Ward, J. M., MacDanold, D. W. (2004). Craniological differentiation between European wildcats (Felis silvestris silvestris) and African wildcats (*F. s. lybica*): implications for their evolution and conservation. *Biologigfnlgflkpmcal Journal of the Linnaean Society* **83**: 47-63

Yamaguchi, N., Kitchener, A, C., Driscoll, C., Nussberger, B. (2015). *Felis silvestris*. The IUCN Red List of Threatened Species 2015: eT60354712A50652361

Yang, Z. (1996). Among-site rate variation and its impact on phylogenetic analyses. *Trends in Ecology and Evolution* **11**: 367-372

Yang, Z. (2006). *Computational molecular evolution*. Oxford (UK): Oxford University Press

Zachos, J., Pagani, M., Sloan, L., Thomas, E., Billups, K. (2001). Trends, rhythms and aberations in global climate 65 Ma to present. *Science* 292:686-693

Zeder, M. A. (2012). *Pathways to animal domestication. In Harlan II: Biodiversity in Agriculture: Domestication, Evolution & Sustainability*. (Eds Damania, A and Gepts, P. University of California: Davis: California pp 227-259

Zelditch, M. L., Swiderski, D. L., Sheets, H. D., Fink, W. L. (2004). *Geometric Morphometrics for Biologists: A Primer*. Elsevier Academic Press: London

Zeuner, F. E. (1953). The chronology of the Mousterian at Gorham's Cave, Gibraltar. *Proceedings of the Prehistric Society* **19**: 180-189

Zischer, H., Geisert, H., von Haeseler, A., Pääbo, S. (2003). A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. *Nature* **378**: 489-492

Appendix A (Chapter 2)

Table AT1. Collagen yields, uncalibrated and calibrated ^{14}C dates for all dated specimens in this study. Specimens were dated at the Oxford Radiocarbon Accelerator Unit (ORAU) and calibrated via OxCal (Bronk-Ramsey *et al.* 2013). Sorted by age.

Genus	Species	ID	OxA Number	Site	Material	% Yield	% C	$\delta^{13}C$	$\delta^{15}N$	C:N	Uncalibrated Date	cal BP (95.4%)
Felis	*silvestris*	MM091	33,576	Dunagoil Cave	Bone	14	44.7	-22.3	7.3	3.2	1783 ± 28	1,814 - 1,618
Castor	*fiber*	MM033	33,685	Swaffham Fen	Bone	11.9	44.3	-22.7	4.3	3.2	3,870 ± 33	4,415 - 4,160
Castor	*fiber*	MM061	33,529	Burwell Fen	Tooth	5.2	45	-22.5	5.3	3.2	3,899 ± 29	4,418 - 4,247
Castor	*fiber*	MM027	33,686	Cambridgeshire Fens	Tooth	4.7	44	-22.7	7.8	3.2	3,907 ± 32	4,422 - 4,246
Castor	*fiber*	MM015	33,524	Somerset Levels	Bone	3.4	43.2	-22.8	3.4	3.3	4,726 ± 29	5,583 - 5,326
Felis	*silvestris*	MM093	33,527	Robin Hood Cave	Bone	5.7	45.4	-20	6.4	3.2	6,200 ± 33	7,239 - 6,999
Felis	*silvestris*	MM092	33,528	Pin Hole Cave	Bone	9.3	45.4	-19.2	6.4	3.2	7,820 ± 36	8,715 - 8,484
Felis	*silvestris*	MM098	34,902	Gough's Cave	Bone	10.7	42.3	-20.5	5.6	3.3	9305 ± 45	10,655 - 10,299
Castor	*fiber*	MM004	33,525	Gough's Cave	Bone	7.1	45.7	-21.5	2.8	3.2	10,110 ± 45	11,989 - 11,406
Castor	*fiber*	MM005	33,526	Gough's Cave	Bone	7.4	46	-21.7	3.2	3.2	10,300 ± 50	12,386 - 11,836

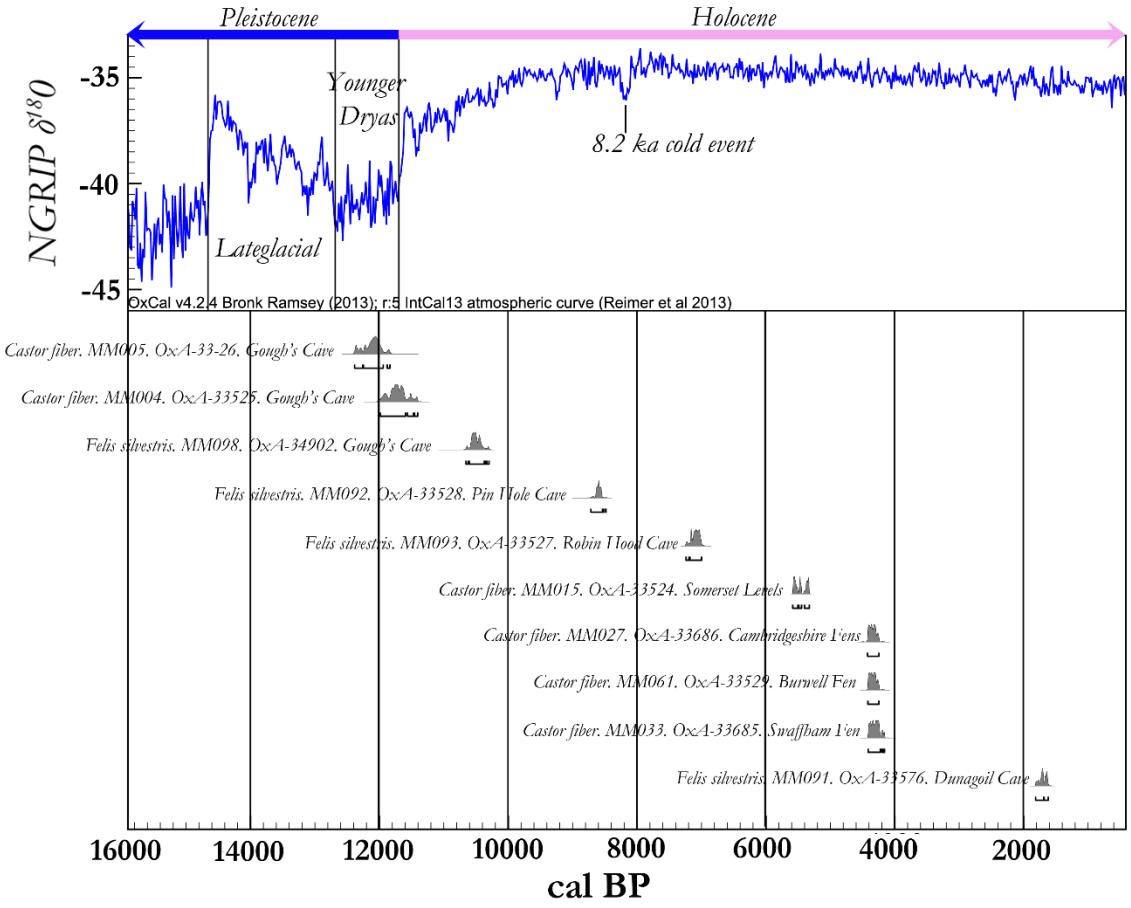


Figure AF1. All ten AMS ^{14}C radiocarbon dates obtained as part of this study. Specimens were dated at the Oxford Radiocarbon Accelerator Unit (ORAU) and calibrated via OxCal (Bronk-Ramsey *et al.* 2013). Displayed by decreasing age.

Protocol A1. Ancient DNA Extraction. From Dabney *et al.* (2013)

Between 10 and 50mg of finely powdered bone or tooth was digested for 18 hours in 1mL extraction buffer with a concentration of proteinase K at 0.25mg/mL and EDTA at 0.45M. The digestion was carried out at 37°C with constant rotation. The solution was centrifuged at 17,500RPM for two minutes to pellet the non-digested power. The supernatant was then removed and added to 13mL of binding buffer containing guanidine hydrochloride at 5M, isopropanol at 40%, Tween-20 at 0.05% and sodium acetate at 90mM. DNA was bound to a silica mesh membrane over two centrifuge steps of 2,500RMP for 4 min and 3,000RMP for 2 min. Silica particles with the conjugated DNA were then subjected to an ethanol wash before the DNA was eluted in 2 x 50-µl volumes in a low-salt TET buffer (Tris-HCL at 10mM, EDTA at 1mM and Tween-20 at 0.05%) to give a final volume of 100-µl. DNA extracts were stored at -20°C.

Protocol A2. Polymerase Chain Reaction (PCR).

PCR reactions were performed in 25µl volumes with final volumes of 0.5 U of HotStart Taq DNA polymerase, 1x Taq polymerase buffer with $(NH_4)_2SO_4$, 0.5µM $MgCI_2$, 0.1mg/mL Bovine Serum Albumin (BSA), 0.2µM of deoxynucleoside triphosphates (dNTPs) and 0.4µM of each primer. Cycling conditions were: initial denaturation at 95°C for 5 mins followed by 50 cycles of denaturation at 94°C for 1 min, annealing at between 49°C to 51°C (depending on optimal annealing temperature for each primer pair) for 1 min and extension at 72°C for 1 min with a final extension at 72°C for 10 mins. Negative extraction controls and negative PCR controls (with H_2O) were performed with each round of PCR. Amplicons were visualised on 2% agarose gels and products of successful amplifications were purified using 0.15x Ampure Beads to remove primer dimer and residual contaminants of the PCR process. Sequencing was performed commercially by Macrogen on an AB1 Prism 310 Sequencer with primers at 5pM.

Protocol A3. DNA Library Build. From Meyer & Kircher (2010).

First, isolated DNA fragments were converted to blunt end DNA. A 50µl starting template containing 20µl DNA extract and 30µl of ddH_20 was added to a master mix with final concentrations of 1x Buffer Tango, 100µM dNTPs, 1mM ATP, 0.5U/µL T4 polynucleotide kinase (PNK) and 0.1U/µL T4 DNA polymerase at a final volume of 70µl. This was incubated at 25°C for 15 mins followed by a phosphorylation step at 12°C for 5 mins. Samples were purified via the Qiaquick PCR Purification Kit before elution in 22µl warm Tris-CL 10mM, pH 8.5 (EB buffer). 20µl of eluted sample was added to 19µl of adapter ligation master mix containing final concentrations of 1x T4 DNA ligase buffer, 5% PEG-400, 0.125U/µL T4 DNA ligase and adapter mix at 2.5µM each and incubated at 22°C for 30 mins. Samples were purified as above and eluted in 22µL EB buffer. 20µL of eluted sample was added to 20µL of adapter fill-in master mix containing final concentrations of 1x ThermoPol Reaction Buffer, dNTPs at 250µM each and Large Fragment *Bst* DNA Polymerase at 0.3U/µL. Reaction conditions were 37°C for 20 mins followed by a purification step as above. Samples were eluted in 33µL. 3µL of index was added to each sample and 11µl of sample was added to 39µL (to give three 50µL reactions per sample) of an index PCR master mix containing reagents at final concentrations of 1x Amplitaq Buffer, 2.4mM $MgCI_2$, dNTPs at 200µM each, 0.4mg/mL BSA, 0.2µM primer IS4 and 0.5U/µL AmpliTaq Gold. Final concentration of index primer was 0.2µM. Index PCR conditions were: initial denaturation at 94°C for 2 mins, 20 cycles of denaturation at 94°C for 40s - annealing at 60°C for 40s – extension at 72°C for 40s followed by a final extension at 72°C. A final PCR purification step was performed as above giving in EB Buffer to a final volume of 80µl. Sequencing was performed on an Illumina NextSeq500.

Protocol A4. In-solution Hybridisation Capture Enrichment.

A hybridisation mix containing (per reaction): 9µl HYB1 (20X SSPE for a final concentration of 9X), 0.5 µl HYB2 (0.5M EDTA for final conc. 0.0125M), 3.5 µl HYB3 (50X Denhardt's Solutions for final conc. *c.* 9X), 0.5µl HYB4 (10% SDS for final conc. 0.25%), 1µl 20U/µl RNase Block and 5.5µl baits (diluted 1:10) were combined to give a volume of 20µl of 18.5µl was used as the final reaction volume. A blockers mix containing (per reaction): 2.5µl Block1 (1µg/ul human C_0t-1 DNA for a final conc. 0.2µg/ul), 2.5µl Block2 (1µg/µl salmon sperm DNA for a final conc. of 0.2µg/ul), 0.5µl Block 3 (library-specific adapter blockers) and 7ul of DNA library was combined to give a volume of 12.5µl of which 12µl was used as the final reaction volume. Blockers-library mix was incubated in the thermal cycler for 5 mins @ 95°C then at the hybridisation temp of 55°C. Hybridisation mix was heated to the hybridisation temperature for 5 mins in the thermal cycler before 18µl was added to the Blocker-libraries mix and incubated for a minimum of 20 hours. 30µl of magnetic beads were washed three times in 200µl Binding Buffer (1M NaCI: 10mM Tris-HCl, 1mM EDTA) before resuspension in 70µl Binding Buffer and warming to the hybridisation temperature of 55°C warmed beads and hybridisation reactions were combined and incubated with occasional agitation for 30 mins. Beads were pelleted, supernatant removed and washed three time at hybridisation temp with 500µl warmed Wash Buffer (400µl HYB4, 39.6 mL ddH^2O and 10ml Wash Buffer 2 (0.1X SSC, 0.1% SDS). Beads were re-suspended in 30µl of 10mM Tris-Cl, 0.05% Tween-20. This mixture was then split into two 15µl reactions and amplified with KAPA HiFi DNA Polymerase at 1X conc. Reactions times were activation @ 98°C for 20s, 20 cycles of denaturation @ 98°C for 30s, annealing @ 55°C for 30s, extension @ 72°C for 30s before a final extension for 5mins @ 72°C. Beads were then pelleted and the supernatant pooled, purified (as in Protocol 2.3) and eluted in Buffer EB to a final volume of 40µl. Elution volumes was kept as low as possibly to ensure sufficient DNA concentration for downstream sequencing.

Appendix B (Chapter 3).

Table BT1. Specimen list of all modern *Microtus* species used in this study. Sourced from The Natural History Museum, London.

Genus	Species	Collection ID	Country	Latitude	Longitude
Microtus	*agrestis*	51.742	Austria	47.1696	12.81379
Microtus	*agrestis*	51.743	Austria	47.1696	12.81379
Microtus	*agrestis*	70.1649	Austria	47.67585	12.25314
Microtus	*agrestis*	98.6.7.21	Denmark	55.927909	12.300806
Microtus	*agrestis*	20.11.6.2	Estonia	58.595272	25.013607
Microtus	*agrestis*	25.6.1.13	Estonia	58.595272	25.013607
Microtus	*agrestis*	25.6.1.15	Estonia	58.595272	25.013607
Microtus	*agrestis*	25.6.1.16	Estonia	58.595272	25.013607
Microtus	*agrestis*	1.11.7.11	Finland	60.12378	24.441484
Microtus	*agrestis*	70.1642	Finland	67.92223	26.504644
Microtus	*agrestis*	70.1643	Finland	67.92223	26.504644
Microtus	*agrestis*	70.1645	Finland	67.92223	26.504644
Microtus	*agrestis*	70.1646	Finland	67.92223	26.504644
Microtus	*agrestis*	70.1647	Finland	67.92223	26.504644
Microtus	*agrestis*	98.11.4.3	Finland	60.12378	24.441484
Microtus	*agrestis*	98.11.4.4	Finland	60.12378	24.441484
Microtus	*agrestis*	98.11.4.5	Finland	60.12378	24.441484
Microtus	*agrestis*	98.11.4.6	Finland	60.12378	24.441484
Microtus	*agrestis*	1.11.7.12	France	48.829924	5.912225
Microtus	*agrestis*	45.7.5.7	France	42.666667	1
Microtus	*agrestis*	5.11.18.25	France	46.18744	6.292736
Microtus	*agrestis*	5.11.18.26	France	46.18744	6.292736
Microtus	*agrestis*	5.11.18.32	France	46.18744	6.292736
Microtus	*agrestis*	62.1547	France	47.373797	7.061744
Microtus	*agrestis*	6.4.2.8	France	46.190206	6.317927
Microtus	*agrestis*	63.136	France	50.870706	1.584027
Microtus	*agrestis*	63.1366	France	50.870706	1.584027
Microtus	*agrestis*	63.1367	France	50.870706	1.584027
Microtus	*agrestis*	67.516	France	50.870706	1.584027
Microtus	*agrestis*	8.8.4.228	France	44.359532	1.406208
Microtus	*agrestis*	8.8.4.229	France	45.76043	3.957103
Microtus	*agrestis*	8.8.4.232	France	48.447193	2.603497
Microtus	*agrestis*	11.9.10.6	Germany	48.190776	8.336815
Microtus	*agrestis*	1987.23	Germany	48.790447	11.497889
Microtus	*agrestis*	47.1048	Germany	52.206212	13.551638
Microtus	*agrestis*	47.811	Germany	52.206212	13.551638
Microtus	*agrestis*	13.7.27.2	Hungary	46.469478	19.980436
Microtus	*agrestis*	11.8.17.3	Hungary	46.694616	22.478352
Microtus	*agrestis*	165	Hungary	47.162494	19.503304
Microtus	*agrestis*	166	Hungary	47.162494	19.503304
Microtus	*agrestis*	167	Hungary	47.162494	19.503304
Microtus	*agrestis*	169	Hungary	47.162494	19.503304
Microtus	*agrestis*	170	Hungary	47.162494	19.503304
Microtus	*agrestis*	171	Hungary	47.162494	19.503304

Table BT1 cont. Specimen list of all modern *Microtus* species used in this study. Sourced from The Natural History Museum, London.

Genus	Species	Collection ID	Country	Latitude	Longitude
Microtus	*agrestis*	175	Hungary	47.162494	19.503304
Microtus	*agrestis*	177	Hungary	47.162494	19.503304
Microtus	*agrestis*	47.692	Italy	46.577833	12.24694
Microtus	*agrestis*	31.11.9.12	Italy	43.94236	12.457777
Microtus	*agrestis*	31.11.9.13	Italy	43.94236	12.457777
Microtus	*agrestis*	31.11.9.14	Italy	43.94236	12.457777
Microtus	*agrestis*	11.1.2.46	Norway	60.197538	24.744463
Microtus	*agrestis*	26.11.21.10	Norway	62.145255	6.074798
Microtus	*agrestis*	26.11.21.11	Norway	62.145255	6.074798
Microtus	*agrestis*	26.11.21.12	Norway	62.145255	6.074798
Microtus	*agrestis*	65.3951	Norway	58.735793	5.647735
Microtus	*agrestis*	65.4032	Norway	69.726919	30.045043
Microtus	*agrestis*	65.4033	Norway	69.726919	30.045043
Microtus	*agrestis*	65.4035	Norway	69.726919	30.045043
Microtus	*agrestis*	98.5.2.8	Norway	62.230538	8.350119
Microtus	*agrestis*	98.5.2.9	Norway	62.230538	8.350119
Microtus	*agrestis*	58.548	Poland	52.700713	23.867672
Microtus	*agrestis*	58.549	Poland	52.700713	23.867672
Microtus	*agrestis*	58.55	Poland	52.700713	23.867672
Microtus	*agrestis*	58.551	Poland	52.700713	23.867672
Microtus	*agrestis*	73.1279	Romania	47.5736000	25.3482060
Microtus	*agrestis*	14.11.1.87	Russia	55.557923	98.656935
Microtus	*agrestis*	14.11.1.88	Russia	55.557923	98.656935
Microtus	*agrestis*	14.11.1.89	Russia	55.557923	98.656935
Microtus	*agrestis*	14.6.12.1	Russia	50.618192	86.219931
Microtus	*agrestis*	2.2.3.1	Russia	55.340749	38.293984
Microtus	*agrestis*	20.7.4.36	Russia	61.52401	105.318756
Microtus	*agrestis*	28.6.19.24	Russia	65.571	63.056
Microtus	*agrestis*	28.6.19.25	Russia	65.571	63.056
Microtus	*agrestis*	94.1.1.14	Spain	43.29749	-7.680772
Microtus	*agrestis*	95.4.29.3	Spain	42.575055	-8.133856
Microtus	*agrestis*	0.5.15.4	Sweden	59.858564	17.638927
Microtus	*agrestis*	0.5.15.5	Sweden	59.858564	17.638927
Microtus	*agrestis*	45.10.25.6	Sweden	60.128161	18.643501
Microtus	*agrestis*	50.88	Sweden	63.214844	14.314527
Microtus	*agrestis*	50.89	Sweden	63.214844	14.314527
Microtus	*agrestis*	69.768	Sweden	63.171192	14.95918
Microtus	*agrestis*	62.1548	Switzerland	44.86809	3.693144
Microtus	*agrestis*	62.1552	Switzerland	46.241556	6.218586
Microtus	*agrestis*	62.1553	Switzerland	46.241556	6.218586
Microtus	*agrestis*	62.1554	Switzerland	46.251453	5.546253
Microtus	*agrestis*	62.1558	Switzerland	45.020533	5.305351
Microtus	*agrestis*	62.1559	Switzerland	45.020533	5.305351
Microtus	*agrestis*	62.1563	Switzerland	47.195456	7.1811

Table BT1 cont. Specimen list of all modern *Microtus* species used in this study. Sourced from The Natural History Museum, London.

Genus	Species	Collection ID	Country	Latitude	Longitude
Microtus	*agrestis*	11.1.2.52	UK	53.567471	-0.080784
Microtus	*agrestis*	11.1.2.53	UK	53.567471	-0.080784
Microtus	*agrestis*	11.1.3.246	UK	53.567471	-0.080784
Microtus	*agrestis*	11.1.3.247	UK	53.567471	-0.080784
Microtus	*agrestis*	11.1.3.249	UK	53.567471	-0.080784
Microtus	*agrestis*	11.1.3.275	UK	52.111111	0.03971
Microtus	*agrestis*	11.1.3.274	UK	52.111111	0.03971
Microtus	*agrestis*	1990.48	UK	53.576865	-2.428219
Microtus	*agrestis*	65.378	UK	51.525026	-0.3415
Microtus	*agrestis*	7.7.7.2907	UK	50.875875	-1.632772
Microtus	*agrestis*	7.7.7.2916	UK	50.875875	-1.632772
Microtus	*agrestis*	7.7.7.2918	UK	50.875875	-1.632772
Microtus	*agrestis*	7.7.7.2919	UK	50.875875	-1.632772
Microtus	*agrestis*	7.7.7.2923	UK	50.875875	-1.632772
Microtus	*arvalis*	47.407	Afganistan	34.907472	68.260074
Microtus	*arvalis*	47.409	Afganistan	34.907472	68.260074
Microtus	*arvalis*	19.7.7.2934	Austria	47.253741	11.601487
Microtus	*arvalis*	31.11.9.15	Austria	47.202702	12.716846
Microtus	*arvalis*	90.1.30.1	Austria	47.253741	11.601487
Microtus	*arvalis*	3.3.30.3	Belguim	50.3	4.76667
Microtus	*arvalis*	3.3.30.4	Belguim	50.3	4.76667
Microtus	*arvalis*	20.7.4.35	Caucasus	43.585181	45.000000
Microtus	*arvalis*	26.2.2.19	Caucasus	43.585181	45
Microtus	*arvalis*	36.12.6.25	Caucasus	43.585181	45
Microtus	*arvalis*	36.12.6.26	Caucasus	43.585181	45
Microtus	*arvalis*	12.4.1.101	China	43.209621	81.391983
Microtus	*arvalis*	12.4.1.93	China	45.0000000	85.0000000
Microtus	*arvalis*	12.4.1.94	China	45.0000000	85.0000000
Microtus	*arvalis*	12.4.1.96	China	45.0000000	85.0000000
Microtus	*arvalis*	5.12.4.6	China*	42.000000	80.000000
Microtus	*arvalis*	5.12.4.7	China*	42.000000	80.000000
Microtus	*arvalis*	21.12.1.10	Croatia	45.258156	17.383963
Microtus	*arvalis*	21.12.1.13	Croatia	45.258156	17.383963
Microtus	*arvalis*	21.12.1.11	Croatia	45.258156	17.383963
Microtus	*arvalis*	21.12.1.12	Croatia	45.258156	17.383963
Microtus	*arvalis*	21.12.1.14	Croatia	45.258156	17.383963
Microtus	*arvalis*	21.12.1.8	Croatia	45.258156	17.383963
Microtus	*arvalis*	21.12.1.9	Croatia	45.258156	17.383963
Microtus	*arvalis*	25.6.1.18	Estonia	49.634137	8.350718
Microtus	*arvalis*	25.6.1.19	Estonia	49.634137	8.350718
Microtus	*arvalis*	25.6.1.21	Estonia	49.634137	8.350718
Microtus	*arvalis*	5.4.9.10	France	46.18744	6.292736
Microtus	*arvalis*	6.6.4.23	France	43.483152	-1.558626
Microtus	*arvalis*	6.6.4.24	France	43.483152	-1.558626

Table BT1 cont. Specimen list of all modern *Microtus* species used in this study. Sourced from The Natural History Museum, London.

Genus	Species	Collection ID	Country	Latitude	Longitude
Microtus	*arvalis*	6.6.4.27	France	43.483152	-1.558626
Microtus	*arvalis*	6.6.4.28	France	43.483152	-1.558626
Microtus	*arvalis*	62.163	France	47.507009	6.862954
Microtus	*arvalis*	62.1631	France	47.507009	6.862954
Microtus	*arvalis*	62.1635	France	46.113183	6.332688
Microtus	*arvalis*	62.1636	France	46.113183	6.332688
Microtus	*arvalis*	62.1637	France	44.520518	6.761623
Microtus	*arvalis*	8.8.10.106	France	47.507009	6.862954
Microtus	*arvalis*	8.8.10.110	France	47.507009	6.862954
Microtus	*arvalis*	8.8.10.114	France	48.573405	7.752111
Microtus	*arvalis*	8.9.1.70	France	45.76043	3.957103
Microtus	*arvalis*	21.4.27.22	Georgia	41.650601	44.012677
Microtus	*arvalis*	21.4.27.23	Georgia	41.650601	44.012677
Microtus	*arvalis*	19.7.7.2073	Germany	51.104541	13.201738
Microtus	*arvalis*	2011.778	Germany	54.086546	13.392341
Microtus	*arvalis*	97.12.4.28	Germany	51.29315	14.831172
Microtus	*arvalis*	95.4.18.15	Germany	50.642196	11.195119
Microtus	*arvalis*	95.4.18.16	Germany	50.642196	11.195119
Microtus	*arvalis*	97.12.4.30	Germany	51.29315	14.831172
Microtus	*arvalis*	31.11.11.57	Greece	40.0856	22.3586
Microtus	*arvalis*	31.11.11.61	Greece	38.44722	22.3586
Microtus	*arvalis*	31.11.11.62a	Greece	38.44722	22.3586
Microtus	*arvalis*	31.11.11.66	Greece	38.44722	22.3586
Microtus	*arvalis*	10.10.2.15	Hungary	47.259995	22.389519
Microtus	*arvalis*	10.10.2.16	Hungary	47.259995	22.389519
Microtus	*arvalis*	10.10.2.20	Hungary	47.259995	22.389519
Microtus	*arvalis*	10.10.2.21	Hungary	47.259995	22.389519
Microtus	*arvalis*	10.10.2.22	Hungary	47.259995	22.389519
Microtus	*arvalis*	10.10.2.23	Hungary	47.259995	22.389519
Microtus	*arvalis*	10.9.14.27	Hungary	47.259995	22.389519
Microtus	*arvalis*	11.1.14.6	Hungary	47.259995	22.389519
Microtus	*arvalis*	27.10.26.40	Iran	31.883333	59.8
Microtus	*arvalis*	27.10.26.44	Iran	31.883333	59.8
Microtus	*arvalis*	27.10.26.38	Iran	31.883333	59.8
Microtus	*arvalis*	66.4632	Italy	46.540471	12.135652
Microtus	*arvalis*	8.7.18.8	Italy	44.494887	11.342616
Microtus	*arvalis*	8.8.2.4	Italy	44.494887	11.342616
Microtus	*arvalis*	90.3.5.15	Italy	47.253741	11.601487
Microtus	*arvalis*	12.4.1.100	Kazakhstan	47.259995	22.389519
Microtus	*arvalis*	14.5.10.165	Kazakhstan	44.162778	80.000000
Microtus	*arvalis*	14.5.10.166	Kazakhstan	44.162778	80.000000
Microtus	*arvalis*	14.5.10.168	Kazakhstan	44.162778	80.000000
Microtus	*arvalis*	14.5.10.169	Kazakhstan	44.162778	80.000000
Microtus	*arvalis*	14.5.10.170	Kazakhstan	44.162778	80.000000

Table BT1 cont. Specimen list of all modern *Microtus* species used in this study. Sourced from The Natural History Museum, London.

Genus	Species	Collection ID	Country	Latitude	Longitude
Microtus	*arvalis*	68.429	Macedonia	41.123098	20.801648
Microtus	*arvalis*	68.43	Macedonia	41.123098	20.801648
Microtus	*arvalis*	68.431	Macedonia	41.123098	20.801648
Microtus	*arvalis*	28.4.4.29	Russia	52.049708	47.369227
Microtus	*arvalis*	34.2.11.20	Russia	51.556379	45.979817
Microtus	*arvalis*	36.12.6.24	Russia	54.204836	37.618492
Microtus	*arvalis*	42.11.11.4	Russia	43.585181	45
Microtus	*arvalis*	23.12.1.55	Russia	50.618192	86.219931
Microtus	*arvalis*	23.12.1.57	Russia	50.618192	86.219931
Microtus	*arvalis*	23.12.1.58	Russia	50.618192	86.219931
Microtus	*arvalis*	34.11.7.23	Serbia	44.738964	20.378606
Microtus	*arvalis*	34.7.1.22	Serbia	44.738964	20.378606
Microtus	*arvalis*	34.8.14.7	Serbia	44.892606	20.533991
Microtus	*arvalis*	55.12.24.352	Siberia	61.01371	99.196656
Microtus	*arvalis*	34.11.26.30	Slovakia	49.112374	20.129861
Microtus	*arvalis*	34.11.26.33	Slovakia	49.112374	20.129861
Microtus	*arvalis*	34.11.26.34	Slovakia	49.112374	20.129861
Microtus	*arvalis*	1951.8	Spain	42.999529	-4.139262
Microtus	*arvalis*	1951.9	Spain	42.999529	-4.139262
Microtus	*arvalis*	61.155	Spain	39.770232	-0.201144
Microtus	*arvalis*	61.156	Spain	39.770232	-0.201144
Microtus	*arvalis*	8.2.9.205	Spain	41.715521	-5.694451
Microtus	*arvalis*	8.7.30.10	Spain	40.240961	-5.996218
Microtus	*arvalis*	5.8.3.21	Switzerland	47.010202	8.483356
Microtus	*arvalis*	62.1924	Switzerland	44.86809	3.693144
Microtus	*arvalis*	79.9.25.51	Switzerland	46.204391	6.143158
Microtus	*arvalis*	79.9.25.52	Switzerland	46.204391	6.143158
Microtus	*arvalis*	1.5.8.4	Turkey	56.620833	84.709167
Microtus	*arvalis*	6.5.1.61	Turkey	38.963745	35.243322
Microtus	*arvalis*	6.5.1.62	Turkey	38.963745	35.243322
Microtus	*arvalis*	6.5.1.63	Turkey	38.963745	35.243322
Microtus	*arvalis*	62.245	Turkey	40.604104	40.604104
Microtus	*arvalis*	25.9.2.59	UK	49.424199	-2.532351
Microtus	*arvalis*	25.9.2.62	UK	49.424199	-2.532351
Microtus	*arvalis*	8.9.2.28	UK	49.433607	-2.559365
Microtus	*arvalis*	25.9.2.52	UK	49.424199	-2.532351
Microtus	*arvalis*	25.9.2.55	UK	49.424199	-2.532351
Microtus	*arvalis*	25.9.2.56	UK	49.424199	-2.532351
Microtus	*arvalis*	25.9.2.58	UK	49.424199	-2.532351
Microtus	*arvalis*	25.9.2.61	UK	49.424199	-2.532351
Microtus	*arvalis*	25.9.2.64	UK	49.424199	-2.532351
Microtus	*arvalis*	61.116	UK	55.378051	-3.435973
Microtus	*arvalis*	62.649	UK	49.463302	-2.596639
Microtus	*arvalis*	62.651_29.3.61	UK	-	-

Table BT1 cont. Specimen list of all modern *Microtus* species used in this study. Sourced from The Natural History Museum, London.

Genus	Species	Collection ID	Country	Latitude	Longitude
Microtus	*arvalis*	70.277	UK	55.378051	-3.435973
Microtus	*arvalis*	8.9.2.26	UK	49.433607	-2.559365
Microtus	*arvalis*	25.9.2.57	UK	49.424199	-2.532351
Microtus	*arvalis*	17.4.25.1	Ukraine	49.839683	24.029717
Microtus	*arvalis*	17.4.25.2	Ukraine	49.839683	24.029717
Microtus	*arvalis*	17.4.25.3	Ukraine	49.839683	24.029717

Table BT2. Distance measurements for all left lower first molars from Gully Cave specimens. Distances are derived from the 2D landmark analysis.

Distance Table			Distance Table		
Object ID	*Microtus agrestis*	*Microtus arvalis*	Object ID	*Microtus agrestis*	*Microtus arvalis*
MM001L	0.03	**0.003**	MM045L	0.036	**0.004**
MM002L	0.022	**0.011**	MM046L	**0.004**	0.028
MM003L	0.036	**0.003**	MM047L	0.037	**0.005**
MM004L	0.04	**0.008**	MM048L	**0.001**	0.033
MM005L	0.025	**0.007**	MM049L	**0.014**	0.018
MM006L	0.041	**0.009**	MM050L	0.023	**0.009**
MM007L	0.044	**0.012**	MM051L	0.022	**0.01**
MM008L	0.032	**0**	MM052L	**0.007**	0.039
MM009L	0.032	**0**	MM053L	0.031	**0.001**
MM010L	0.034	**0.002**	MM054L	**0.012**	0.02
MM011L	0.028	**0.005**	MM055L	0.018	**0.015**
MM012L	0.028	**0.004**	MM056L	0.035	**0.003**
MM013L	0.026	**0.006**	MM057L	0.026	**0.006**
MM014L	0.042	**0.009**	MM059L	0.033	**0.001**
MM015L	0.033	**0.001**	MM0139L	0.025	**0.007**
MM016L	0.038	**0.006**	MM0140L	0.02	**0.013**
MM017L	0.04	**0.008**	MM0141L	0.026	**0.006**
MM018L	0.028	**0.004**	MM0142L	0.028	**0.004**
MM019L	0.021	**0.011**	MM0144L	0.035	**0.002**
MM021L	**0.006**	0.026	MM0146L	0.026	**0.006**
MM022L	0.036	**0.004**	MM0151L	0.041	**0.009**
MM024L	0.028	**0.004**	MM0152L	0.047	**0.015**
MM028L	0.039	**0.006**	MM0153L	0.029	**0.003**
MM029L	0.02	**0.012**	MM0157L	**0.011**	0.022
MM030L	0.023	**0.009**	MM0158L	0.034	**0.002**
MM031L	0.034	**0.002**	MM0159L	0.025	**0.007**
MM032L	**0.008**	0.025	MM0161L	0.022	**0.01**
MM033L	0.033	**0.001**	MM0162L	0.031	**0.001**
MM035L	**0.009**	0.024	MM0163L	0.029	**0.003**
MM036L	0.02	**0.012**	MM0164L	0.023	**0.01**
MM037L	0.022	**0.01**	MM0165L	0.035	**0.002**
MM038L	0.02	**0.012**	MM0167L	0.036	**0.004**
MM039L	0.024	**0.008**	MM0175L	**0.004**	0.028
MM040L	0.027	**0.006**	MM0176L	0.04	**0.008**
MM041L	0.022	**0.011**	MM0180L	0.022	**0.01**
MM042L	0.029	**0.004**	MM0187L	0.026	**0.006**
MM043L	0.028	**0.004**	MM0192L	0.034	**0.002**
MM044L	0.062	**0.03**			

Table BT2 cont. Distance measurements for all right lower first molars from Gully Cave specimens. Distances are derived from the 2D landmark analysis.

Distance Table			Distance Table		
Object ID	*Microtus agrestis*	*Microtus arvalis*	Object ID	*Microtus agrestis*	*Microtus arvalis*
MM027R	0.025	**0.003**	MM106R	0.02	**0.008**
MM060R	0.023	**0.005**	MM108R	0.027	**0.001**
MM061R	0.024	**0.003**	MM109R	**0.011**	0.016
MM062R	0.03	**0.003**	MM110R	0.022	**0.005**
MM064R	0.024	**0.004**	MM112R	0.015	**0.012**
MM065AR	0.021	**0.007**	MM113R	**0.001**	0.027
MM065BR	0.026	**0.001**	MM114R	0.021	**0.007**
MM066R	0.033	**0.006**	MM115R	0.038	**0.011**
MM067R	0.019	**0.009**	MM116R	0.019	**0.009**
MM069R	0.023	**0.004**	MM117R	0.022	**0.006**
MM070R	0.022	**0.005**	MM118R	**0.01**	0.017
MM071R	0.022	**0.005**	MM119R	**0.001**	0.026
MM072R	0.016	**0.012**	MM120R	**0.013**	0.014
MM073R	**0.01**	0.017	MM121R	0.021	**0.006**
MM074R	0.041	**0.013**	MM122R	0.018	**0.01**
MM075R	**0.006**	0.021	MM123R	0.03	**0.002**
MM076R	0.032	**0.004**	MM124R	0.029	**0.001**
MM077R	0.028	**0.001**	MM125R	0.018	**0.01**
MM078R	0.017	**0.011**	MM126R	0.034	**0.007**
MM079R	0.026	**0.001**	MM127R	0.023	**0.005**
MM080R	**0.002**	0.026	MM128R	0.025	**0.003**
MM081R	**0.003**	0.024	MM129R	0.031	**0.004**
MM082R	0.017	**0.01**	MM130R	**0.001**	0.028
MM083R	**0.004**	0.023	MM131R	0.024	**0.003**
MM084R	0.034	**0.007**	MM132R	0.027	**0**
MM085R	0.019	**0.009**	MM133R	0.026	**0.002**
MM086R	0.041	**0.013**	MM134R	0.024	**0.004**
MM088R	0.019	**0.008**	MM135R	0.022	**0.005**
MM089R	0.031	**0.003**	MM136R	0.02	**0.007**
MM090R	0.022	**0.006**	MM137R	0.034	**0.007**
MM091R	0.014	**0.014**	MM143R	**0**	0.027
MM092R	0.019	**0.009**	MM147R	**0.004**	0.031
MM093R	0.015	**0.013**	MM148R	0.015	**0.013**
MM094R	**0.012**	0.016	MM149R	0.026	**0.001**
MM095R	0.028	**0.001**	MM152R	**0.013**	0.014
MM096R	0.022	**0.005**	MM155R	0.019	**0.008**
MM097R	**0.013**	0.015	MM156R	0.031	**0.003**
MM100R	**0.009**	0.037	MM160R	0.015	**0.012**
MM102R	**0.002**	0.026	MM166R	0.024	**0.004**
MM103R	**0.002**	0.025	MM169R	0.02	**0.008**
MM104R	0.017	**0.011**			

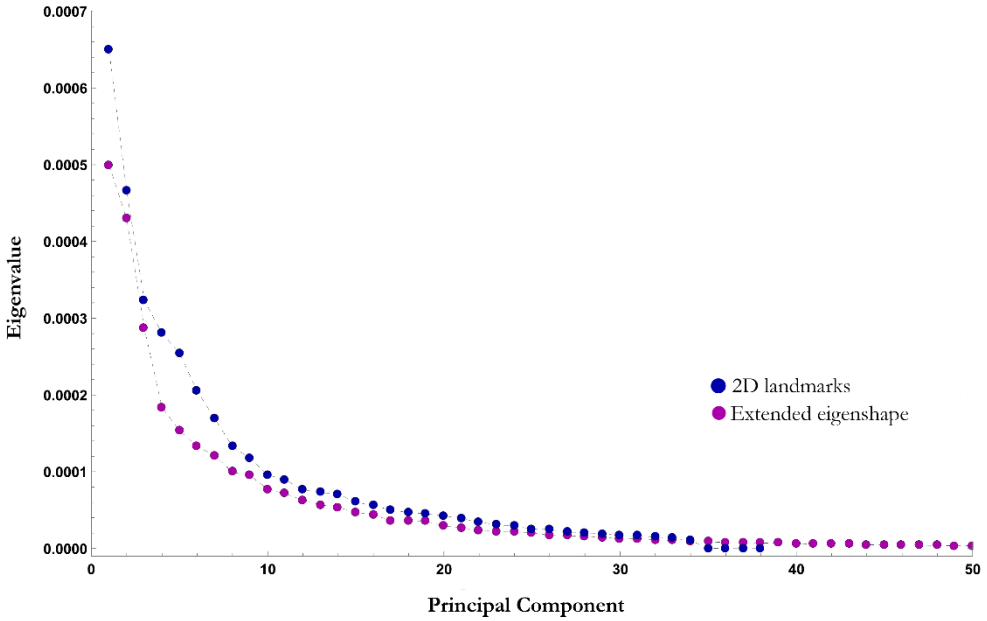


Figure BF1. Screeplot of eigenvalues for each principal component in the 2D and extended eigenshape analysis of *Microtus agrestis* and *Microtus arvalis* M_1 shape. Only the first 50 PC's are shown; 38 were generated in the 2D landmark dataset and 224 were generated in the extended eigenshape analysis.

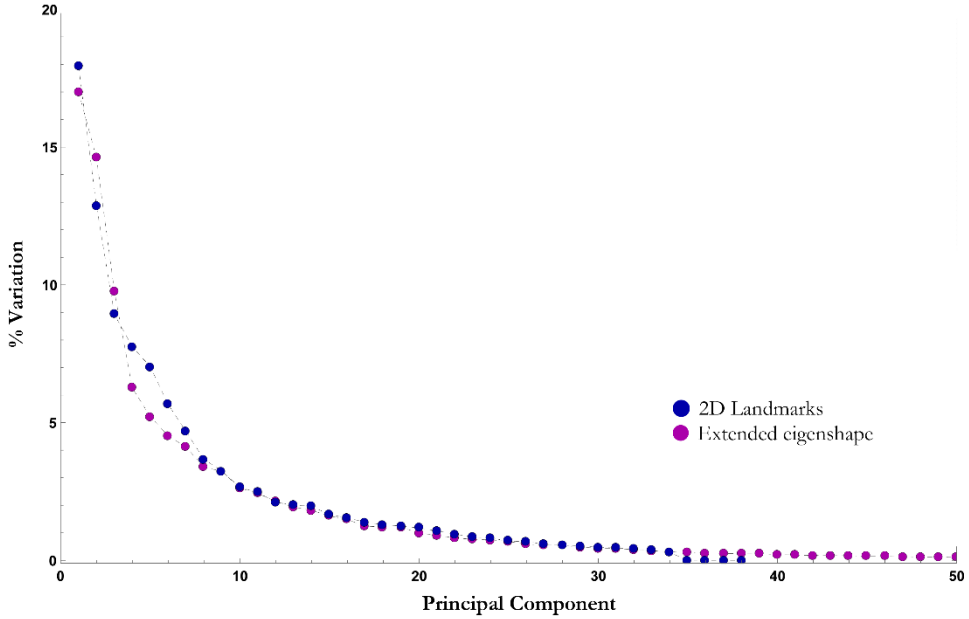


Figure BF2. Screeplot of each principal component in the 2D and extended eigenshape analysis of *Microtus agrestis* and *Microtus arvalis* M_1 shape and the percentage of variation they account for. Only the first 50 PC's are shown; 38 were generated in the 2D landmark dataset and 224 were generated in the extended eigenshape analysis.

Appendix C (Chapter 4).

Table CT1. Accession numbers, clades and citations for unique *Microtus arvalis* haplotypes used in this study.

Genus	Species	Accession No	Country	Reference	Clade
Microtus	*arvalis*	GU187364	Balkan	Buzan *et al* (2010)	Balkan
Microtus	*arvalis*	GU187365	Balkan	Buzan *et al* (2010)	Balkan
Microtus	*arvalis*	GU187366	Balkan	Buzan *et al* (2010)	Balkan
Microtus	*arvalis*	GU187367	Balkan	Buzan *et al* (2010)	Balkan
Microtus	*arvalis*	GU187369	Balkan	Buzan *et al* (2010)	Balkan
Microtus	*arvalis*	GU187370	Balkan	Buzan *et al* (2010)	Balkan
Microtus	*arvalis*	GU187371	Balkan	Buzan *et al* (2010)	Balkan
Microtus	*arvalis*	GU187372	Balkan	Buzan *et al* (2010)	Balkan
Microtus	*arvalis*	GU187373	Balkan	Buzan *et al* (2010)	Balkan
Microtus	*arvalis*	GU187374	Balkan	Buzan *et al* (2010)	Balkan
Microtus	*arvalis*	GU187375	Balkan	Buzan *et al* (2010)	Balkan
Microtus	*arvalis*	GU187376	Balkan	Buzan *et al* (2010)	Balkan
Microtus	*arvalis*	GU187377	Balkan	Buzan *et al* (2010)	Balkan
Microtus	*arvalis*	GU187378	Balkan	Buzan *et al* (2010)	Balkan
Microtus	*arvalis*	GU187379	Balkan	Buzan *et al* (2010)	Balkan
Microtus	*arvalis*	GU187380	Balkan	Buzan *et al* (2010)	Balkan
Microtus	*arvalis*	AY220770	Germnay	Haynes *et al.* (2003)	Central
Microtus	*arvalis*	AY220776	Denmark	Haynes *et al.* (2003)	Central
Microtus	*arvalis*	AY708463	Switzerland	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708467	Switzerland	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708469	Switzerland	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708470	Switzerland	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708474	Germany	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708475	Germany	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708476	Germany	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708477	Germany	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708478	Germany	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708479	Germany	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708480	Germany	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708482	Switzerland	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708483	France	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708486	Switzerland	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708486	Switzerland	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708491	Germany	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708494	Germany	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708495	Germany	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708496	Germany	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708498	Germany	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708514	Switzerland	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708517	Switzerland	Fink *et al.* (2004)	Central
Microtus	*arvalis*	AY708519	Switzerland	Fink *et al.* (2004)	Central
Microtus	*arvalis*	GU190390	France	Martinkova *et al.* (2013)	Central
Microtus	*arvalis*	GU190499	Netherlands	Martinkova *et al.* (2013)	Central
Microtus	*arvalis*	GU190507	Netherlands	Martinkova *et al.* (2013)	Central
Microtus	*arvalis*	GU190507	Netherlands	Martinkova *et al.* (2013)	Central
Microtus	*arvalis*	GU190508	Netherlands	Martinkova *et al.* (2013)	Central
Microtus	*arvalis*	GU190512	Netherlands	Martinkova *et al.* (2013)	Central
Microtus	*arvalis*	GU190522	Netherlands	Martinkova *et al.* (2013)	Central
Microtus	*arvalis*	GU190523	Netherlands	Martinkova *et al.* (2013)	Central

Table CT1 cont. Accession numbers, clades and citations for unique *Microtus arvalis* haplotypes used in this study.

Genus	Species	Accession No	Country	Reference	Clade
Microtus	*arvalis*	GU190526	Netherlands	Martinkova *et al.* (2013)	Central
Microtus	*arvalis*	GU190660	Denmark	Martinkova *et al.* (2013)	Central
Microtus	*arvalis*	GU190662	Germany	Martinkova *et al.* (2013)	Central
Microtus	*arvalis*	GU190663	Germany	Martinkova *et al.* (2013)	Central
Microtus	*arvalis*	KP255541	Poland	Stojak *et al.* (2015)	Central
Microtus	*arvalis*	KP255542	Poland	Stojak *et al.* (2015)	Central
Microtus	*arvalis*	KP255543	Poland	Stojak *et al.* (2015)	Central
Microtus	*arvalis*	KP255544	Poland	Stojak *et al.* (2015)	Central
Microtus	*arvalis*	KP255547	Poland	Stojak *et al.* (2015)	Central
Microtus	*arvalis*	KP255547	Poland	Stojak *et al.* (2015)	Central
Microtus	*arvalis*	KP255549	Poland	Stojak *et al.* (2015)	Central
Microtus	*arvalis*	KP255552	Poland	Stojak *et al.* (2015)	Central
Microtus	*arvalis*	KP255552	Poland	Stojak *et al.* (2015)	Central
Microtus	*arvalis*	KP255554	Poland	Stojak *et al.* (2015)	Central
Microtus	*arvalis*	KP255560	Poland	Stojak *et al.* (2015)	Central
Microtus	*arvalis*	AY220767	Slovakia	Haynes *et al.* (2003)	Eastern
Microtus	*arvalis*	AY220768	Slovakia	Haynes *et al.* (2003)	Eastern
Microtus	*arvalis*	AY220769	Hungary	Haynes *et al.* (2003)	Eastern
Microtus	*arvalis*	AY220770	Finland	Haynes *et al.* (2003)	Eastern
Microtus	*arvalis*	AY220771	Russia	Haynes *et al.* (2003)	Eastern
Microtus	*arvalis*	AY220772	Poland	Haynes *et al.* (2003)	Eastern
Microtus	*arvalis*	AY220773	Poland	Haynes *et al.* (2003)	Eastern
Microtus	*arvalis*	AY220775	Poland	Haynes *et al.* (2003)	Eastern
Microtus	*arvalis*	AY708461	Austria	Fink *et al.* (2004)	Eastern
Microtus	*arvalis*	AY708472	Czech Republic	Fink *et al.* (2004)	Eastern
Microtus	*arvalis*	KP255486	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255487	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255488	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255490	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255491	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255493	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255495	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255496	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255498	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255504	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255506	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255512	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255517	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255518	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255519	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255525	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255528	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255533	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255534	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255536	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255538	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255546	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255557	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255558	Poland	Stojak *et al.* (2015)	Eastern

Table CT1 cont. Accession numbers, clades and citations for unique *Microtus arvalis* haplotypes used in this study.

Genus	Species	Accession No	Country	Reference	Clade
Microtus	*arvalis*	KP255569	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255571	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255575	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255580	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255581	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255582	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255584	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255587	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255588	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255605	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255608	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255611	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255612	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255614	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255615	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255617	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255618	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255619	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	KP255620	Poland	Stojak *et al.* (2015)	Eastern
Microtus	*arvalis*	AY708492	Germany	Fink *et al.* (2004)	Germany Freiberg
Microtus	*arvalis*	AY708493	Germany	Fink *et al.* (2004)	Germany Freiberg
Microtus	*arvalis*	AY708507	Germany	Fink *et al.* (2004)	Germany Freiberg
Microtus	*arvalis*	AY220766	Italy	Haynes *et al.* (2003)	Italy
Microtus	*arvalis*	AY708465	Italy	Fink *et al.* (2004)	Italy
Microtus	*arvalis*	AY708466	Italy	Fink *et al.* (2004)	Italy
Microtus	*arvalis*	AY708513	Italy	Fink *et al.* (2004)	Italy
Microtus	*arvalis*	AM991028	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991032	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991033	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991034	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991037	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991038	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991039	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991041	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991043	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991050	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991052	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991052	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991054	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991055	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991056	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991059	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991061	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991077	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991078	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991079	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991081	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991082	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991083	France	Tougard *et al*. (2008)	Western-North

Table CT1 cont. Accession numbers, clades and citations for unique *Microtus arvalis* haplotypes used in this study.

Genus	Species	Accession No	Country	Reference	Clade
Microtus	*arvalis*	AM991084	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991086	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991088	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991090	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991096	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AM991098	France	Tougard *et al*. (2008)	Western-North
Microtus	*arvalis*	AY220779	UK	Haynes *et al.* (2003)	Western-North
Microtus	*arvalis*	AY220780	UK	Haynes *et al.* (2003)	Western-North
Microtus	*arvalis*	AY220781	UK	Haynes *et al.* (2003)	Western-North
Microtus	*arvalis*	AY220782	UK	Haynes *et al.* (2003)	Western-North
Microtus	*arvalis*	AY220783	UK	Haynes *et al.* (2003)	Western-North
Microtus	*arvalis*	AY220784	UK	Haynes *et al.* (2003)	Western-North
Microtus	*arvalis*	AY220785	UK	Haynes *et al.* (2003)	Western-North
Microtus	*arvalis*	AY220786	UK	Haynes *et al.* (2003)	Western-North
Microtus	*arvalis*	AY220788	Netherlands	Haynes *et al.* (2003)	Western-North
Microtus	*arvalis*	AY708462	Belgium	Fink *et al.* (2004)	Western-North
Microtus	*arvalis*	AY708481	Switzerland	Fink *et al.* (2004)	Western-North
Microtus	*arvalis*	AY708490	Germany	Fink *et al.* (2004)	Western-North
Microtus	*arvalis*	AY708504	France	Fink *et al.* (2004)	Western-North
Microtus	*arvalis*	AY708506	France	Fink *et al.* (2004)	Western-North
Microtus	*arvalis*	AY708521	France	Fink *et al.* (2004)	Western-North
Microtus	*arvalis*	GU190391	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190392	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190396	Luxembourg	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190398	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190406	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190407	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190412	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190413	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190415	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190422	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190422	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190424	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190425	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190427	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190432	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190434	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190436	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190439	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190441	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190442	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190447	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190449	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190456	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190456	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190463	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190477	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190488	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190491	UK	Martinkova *et al.* (2013)	Western-North

Table CT1 cont. Accession numbers, clades and citations for unique *Microtus arvalis* haplotypes used in this study.

Genus	Species	Accession No	Country	Reference	Clade
Microtus	*arvalis*	GU190531	Netherlands	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190531	Netherlands	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190536	Belgium	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190536	Belgium	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190546	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190552	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190554	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190556	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190570	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190589	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190595	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190597	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190604	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190606	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190616	Germany	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190628	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190635	France	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190653	Guernsey	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU190654	Guernsey	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU197787	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU197789	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU197790	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU197797	Uk	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU197797	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU197799	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU197800	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU197808	UK	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU197817	Belgium	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU197820	Belgium	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	GU197820	Belgium	Martinkova *et al.* (2013)	Western-North
Microtus	*arvalis*	AM991024	France	Tougard *et al*. (2008)	Western-South
Microtus	*arvalis*	AM991026	France	Tougard *et al*. (2008)	Western-South
Microtus	*arvalis*	AM991027	France	Tougard *et al*. (2008)	Western-South
Microtus	*arvalis*	AM991031	France	Tougard *et al*. (2008)	Western-South
Microtus	*arvalis*	AM991036	France	Tougard *et al*. (2008)	Western-South
Microtus	*arvalis*	AM991044	France	Tougard *et al*. (2008)	Western-South
Microtus	*arvalis*	AM991046	France	Tougard *et al*. (2008)	Western-South
Microtus	*arvalis*	AM991047	France	Tougard *et al*. (2008)	Western-South
Microtus	*arvalis*	AM991049	France	Tougard *et al*. (2008)	Western-South
Microtus	*arvalis*	AM991062	France	Tougard *et al*. (2008)	Western-South
Microtus	*arvalis*	AM991065	France	Tougard *et al*. (2008)	Western-South
Microtus	*arvalis*	AM991069	France	Tougard *et al*. (2008)	Western-South
Microtus	*arvalis*	AM991070	France	Tougard *et al*. (2008)	Western-South
Microtus	*arvalis*	AM991072	France	Tougard *et al*. (2008)	Western-South
Microtus	*arvalis*	AM991073	France	Tougard *et al*. (2008)	Western-South
Microtus	*arvalis*	AM991076	France	Tougard *et al*. (2008)	Western-South
Microtus	*arvalis*	AM991087	France	Tougard *et al*. (2008)	Western-South
Microtus	*arvalis*	AY220788	Spain	Haynes *et al.* (2003)	Western-South
Microtus	*arvalis*	AY220789	Spain	Haynes *et al.* (2003)	Western-South

Table CT1 cont. Accession numbers, clades and citations for unique *Microtus arvalis* haplotypes used in this study.

Genus	Species	Accession No	Country	Reference	Clade
Microtus	*arvalis*	AY708485	France	Fink *et al.* (2004)	Western-South
Microtus	*arvalis*	AY708488	France	Fink *et al.* (2004)	Western-South
Microtus	*arvalis*	AY708497	Spain	Fink *et al.* (2004)	Western-South
Microtus	*arvalis*	AY708499	Spain	Fink *et al.* (2004)	Western-South
Microtus	*arvalis*	AY708499	Spain	Fink *et al.* (2004)	Western-South
Microtus	*arvalis*	AY708500	Spain	Fink *et al.* (2004)	Western-South
Microtus	*arvalis*	AY708523	Spain	Fink *et al.* (2004)	Western-South
Microtus	*arvalis*	GU190385	Spain	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190387	Spain	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190421	France	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190548	France	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190557	France	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190562	France	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190564	France	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190567	France	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190569	France	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190629	France	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190630	France	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190634	France	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190638	Spain	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190638	Spain	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190639	Spain	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190641	Spain	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190641	Spain	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190646	Spain	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190646	Spain	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190647	Spain	Martinkova *et al.* (2013)	Western-South
Microtus	*arvalis*	GU190647	Spain	Martinkova *et al.* (2013)	Western-South

Appendix Table CT2. Tip dated samples used in the BEAST simulations.

*Sample MM112 is here given the date for the median age of the Bølling warm phase but placement in the phylogenetic tree did not differ whether it was given the tip date of the Bølling or of the Younger Dryas.

ID	Accession Number	Locality	Period	Median Age (cal BP)
R29	GU197795	Orkney	Late-Holocene	863.5
R130	GU197817	Belgium	Late Holocene, Medieval	1200
R131	GU197818	Belgium	Late Holocene, Medieval	1200
R133	GU197819	Belgium	Late Holocene, Medieval	1200
R134	GU197820	Belgium	Late Holocene, Medieval	1200
R138	GU197823	Belgium	Late Holocene, Medieval	1200
R62	GU197806	Orkney	Late-Holocene	1352
R59	GU197804	Orkney	Late-Holocene	1789.5
R58	GU197803	Orkney	Late-Holocene	1795
R125	GU197809	Orkney	Late-Holocene	2922.5
R16	GU197791	Orkney	Late-Holocene	2922.5
R195	GU197816	Orkney	Late-Holocene	2922.5
R25	GU197794	Orkney	Late-Holocene	2922.5
R2	GU197787	Orkney	Late-Holocene	2922.5
R30	GU197796	Orkney	Late-Holocene	2922.5
R31	GU197797	Orkney	Late-Holocene	2922.5
R39	GU197799	Orkney	Late-Holocene	2922.5
R43	GU197800	Orkney	Late-Holocene	2922.5
R60	GU197805	Orkney	Late-Holocene	2922.5
R6	GU197789	Orkney	Late-Holocene	2922.5
R124	GU197808	Orkney	Mid-Holocene	4250
R189	GU197813	Orkney	Mid-Holocene	4301.5
R191	GU197814	Orkney	Mid-Holocene	4377
R194	GU197815	Orkney	Mid-Holocene	4385
R3	GU197788	Orkney	Mid-Holocene	4404.5
R20	GU197792	Orkney	Mid-Holocene	4609.5
R23	GU197793	Orkney	Mid-Holocene	4628
R179	GU197812	Orkney	Mid-Holocene	5037
R177	GU197811	Orkney	Mid-Holocene	5063
R45	GU197802	Orkney	Mid-Holocene	5085.5
R44	GU197801	Orkney	Mid-Holocene	5243.5
MM108	-	Gully Cave	Holocene	9850
MM115	-	Gully Cave	Early Holocene	9850
MM105	-	Gully Cave	Younger Dryas	12250
MM113	-	Gully Cave	Younger Dryas	12250
MM114	-	Gully Cave	Younger Dryas	12250
MM100	-	Gully Cave	Bølling	14325
MM110	-	Gully Cave	Bølling	14325
*MM112	-	Gully Cave	Bølling	14352
MM099	-	Gully Cave	Bølling	14352
MM109	-	Gully Cave	Bølling	14352

CT3 (a-d). BEAST posterior statistics; mean likelihood, estimated sample size (ESS) and highest posterior density (HPD) values.

a.

BEAST Run 1 - Strick Clock, Constant Population Size			
Statistic	Mean	ESS	95% HPD Interval
Posterior	-4784.275	656	-4855.8283, -4716.1317
Likelihood	-2816.51	1960	-2839.7779, -2795.3397
Prior	-1967.765	609	-2039.9504, -1896.9409
Treelikelihood	-2816.51	1690	-2839.7779, -2795.3397

BEAST Run 1 - Sampled From Prior			
Statistic	Mean	ESS	95% HPD Interval
Posterior	205.14	6	-634.11, 767.0508
Likelihood	-	-	-
Prior	205.14	6	-634.11, 767.0508
Treelikelihood	-	-	-

b.

BEAST Run 2 - Relaxed Log Normal Clock, Constant Population Size			
Statistic	Mean	ESS	95% HPD Interval
Posterior	-4665.145	216	-4751.9773, -4569.7
Likelihood	-2775.621	358	-2801.5118, -2749.7874
Prior	-1889.524	199	-1976.048, -1800.9286
Treelikelihood	-2775.621	358	-2801.5118, -2749.7874

BEAST Run 2 - Sampled From Prior			
Statistic	Mean	ESS	95% HPD Interval
Posterior	-834.596	5	-1930.4271, 285.4844
Likelihood	-	-	-
Prior	-834.596	5	-1930.4271, 285.4844
Treelikelihood	-	-	-

CT3 (a-d) cont. BEAST posterior statistics; mean likelihood, estimated sample size (ESS) and highest posterior density (HPD) values.

a.

BEAST Run 3 - Strick Clock, Coalescent Exponential Population			
Statistic	Mean	ESS	95% HPD Interval
Posterior	-4778.947	584	-4848.8532, -4714.2028
Likelihood	-2818.812	976	-2842.5799, -2797.8953
Prior	-1960.135	510	-2025.9824, -1891.7261
Treelikelihood	-2818.812	976	-2842.5799, -2797.8953

BEAST Run 3 - Sampled From Prior			
Statistic	Mean	ESS	95% HPD Interval
Posterior	-1475.658	4	-2023.3309, 26.7802
Likelihood	-	-	-
Prior	-1475.658	4	-2023.3309, 26.7802
Treelikelihood	-	-	-

b.

BEAST Run 4 - Relaxed Log Normal Clock, Coalescent Exponential Population			
Statistic	Mean	ESS	95% HPD Interval
Posterior	-4692.526	218	-4782.5689, -4614.9023
Likelihood	-2782.495	302	-2809.301, -2756.723
Prior	-1910.031	262	-1987.4697, -1825.5306
Treelikelihood	-2782.495	302	-2809.301, -2756.723

BEAST Run 4 - Sampled From Prior			
Statistic	Mean	ESS	95% HPD Interval
Posterior	-813.719	5	-1691.1739, 217.2019
Likelihood	-	-	-
Prior	-813.719	5	-1691.1739, 217.2019
Treelikelihood	-	-	-

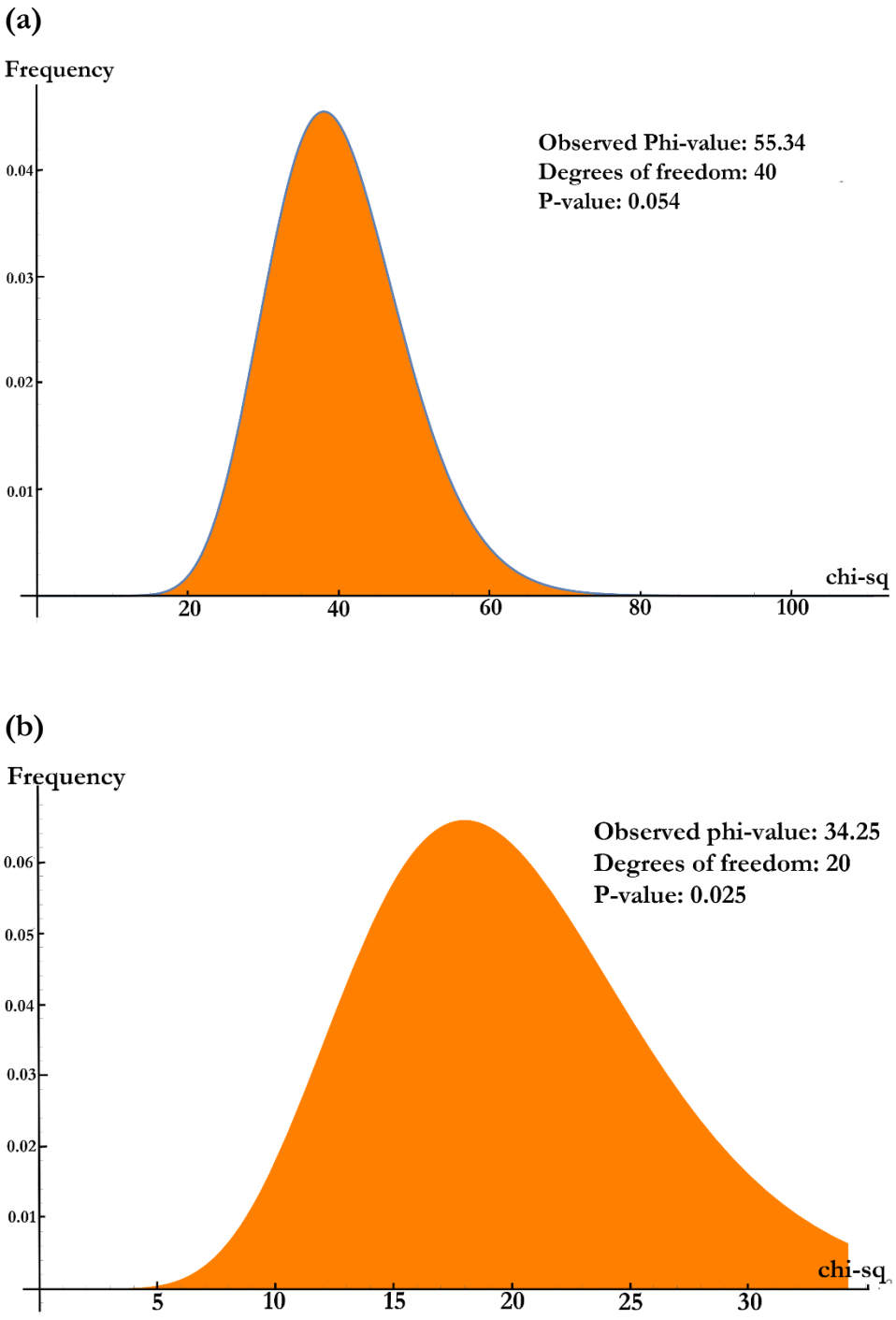
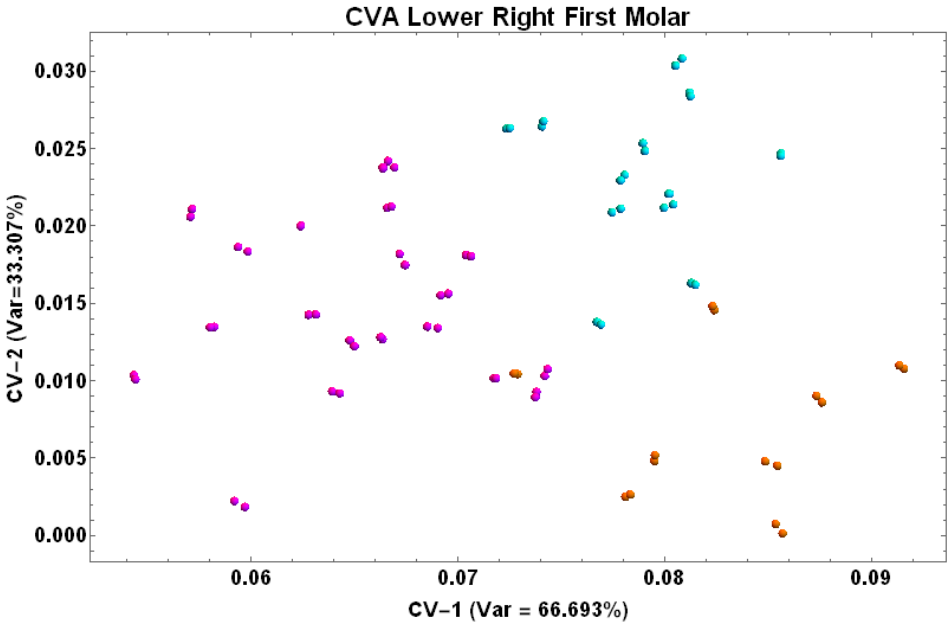


Figure CF1. Log-likelihood ratio results for the Canonical Variates analyses of the three (a) and two (b) group scenarios. This test found a marginally non-significant result for the three group scenario and a significant result for the two-group scenario.

a.



b.

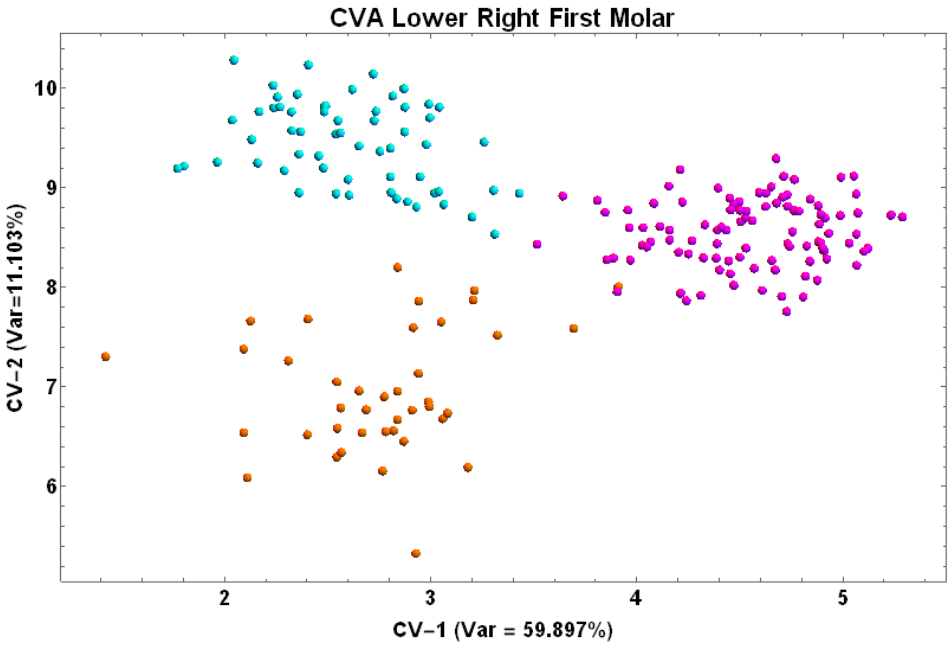


Figure CF2. Data augmentation represented in the Canonical Variates (CV) space for the three group scenario. (a) shows an augmentation where the sample size has been increased two-fold and the drift value set to 5% and (b) shows an augmentation where the sample size has been increased 5-fold and the drift value set to 100%. Orange – Holocene, Cyan – Younger Dryas, Magenta – Bølling.

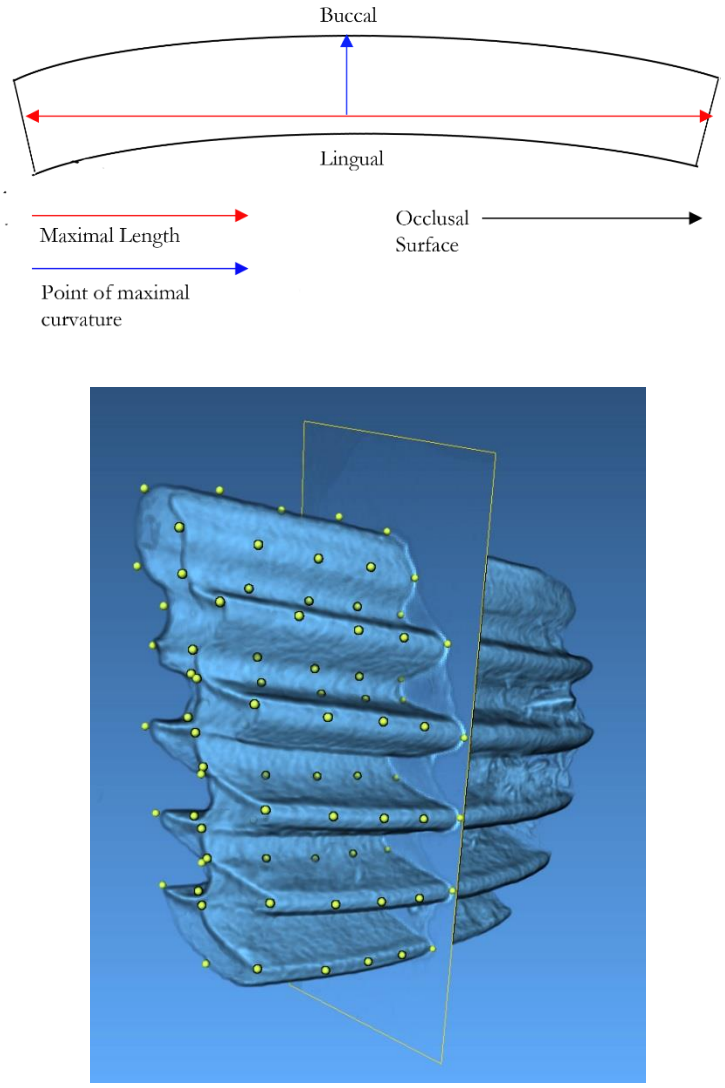


Figure CF3. Calculation of the site of termination for landmark transects in the *M. arvalis* molar. When viewed from the posterior lobe, with the buccal surface uppermost, the point of maximal curvature was determined from the line representing maximal length.

Appendix D (Chapter 5).

Table DT1. Full list of all 49 *Castor fiber* samples that were taken during the course of this study, their origin, collections and localities.

Lab ID	Museum	Site 1	Site 2	County	Region
MM001	UMZC	Holwood Farm	Chatteris	Cambridgeshire	England
MM002	UMZC	Burwell Fen	Fens	Cambridgeshire	England
MM003	UMZC	Unknown	Fens	Cambridgeshire	England
MM004	UBSSM	Gough's Old Cave	Cheddar	Mendips	England
MM005	UBSSM	Gough's Old Cave	Cheddar	Mendips	England
MM006	UBSSM	Sunhole Cave	Cheddar	Mendips	England
MM007	GCMA	Wet Sink	Forest of Dean	Gloucester	England
MM008	UMZC	Somersham Fen	Fens	Cambridgeshire	England
MM009	UMZC	Burwell Fen	Fens	Cambridgeshire	England
MM015	SCCM	Aller Moor	Chedzoy	Somerset	England
MM016	SCCM	Kings Sedgemoor Drains	Chedzoy	Somerset	England
MM022	SM	Upware	Dimmocks Cave	Cambridgeshire	England
MM023	SM	Burwell Fen	Fen	Cambridgeshire	England
MM024	SM	Swaffham Fen	Fens	Cambridgeshire	England
MM025	SM	Unknown	Fens	Cambridgeshire	England
MM026	SM	Burwell Fen	Fens	Cambridgeshire	England
MM027	SM	Unknown	Fens	Cambridgeshire	England
MM028	SM	Burwell Fen	Fens	Cambridgeshire	England
MM029	SM	Unknown	Fens	Cambridgeshire	England
MM030	SM	Reach Fen	Fens	Cambridgeshire	England
MM031	SM	Burwell Fen	Fens	Cambridgeshire	England
MM032	SM	Reach Fen	Fens	Cambridgeshire	England
MM033	SM	Swaffham Fen	Fens	Cambridgeshire	England
MM034	SM	Swaffham Fen	Fens	Cambridgeshire	England
MM035	SM	Littleport	Fens	Cambridgeshire	England
MM042	SM	Burwell Fen	Fens	Cambridge	England
MM044	SM	Burwell Fen	Fens	Cambridgeshire	England
MM046	MOL	Sanderson Factory Site	Denham	Thames Catchment	England
MM047	SM	Burwell Fen	Fens	Cambridgeshire	England
MM048	SM	Unknown	Fens	Cambridgeshire	England
MM049	SM	Swaffham Fen	Fens	Cambridgeshire	England
MM050	SM	Burwell Fen	Fens	Cambridgeshire	England
MM051	UMZC	Burwell Fen	Fens	Cambridgeshire	England
MM052	UMZC	Burwell Fen	Fens	Cambridgeshire	England
MM054	SM	Burwell Fen	Fens	Cambridgeshire	England
MM055	SM	Burwell Fen	Fens	Cambridgeshire	England
MM056	SM	Burwell Fen	Fens	Cambridgeshire	England
MM057	SM	Burwell Fen	Fens	Cambridgeshire	England
MM058A	NMSZ	Linton Loch	Fens	Roxburghshire	Scotland
MM058B	NMSZ	Middlestot's Bog	Fens	Berwickshire	Scotland
MM059	SM	Unknown	Fens	Cambridgeshire	England
MM060	GLRCM	Hills Flats	Severn Estuary	Gloucester	England
MM061	SM	Burwell Fen	Fens	Cambridgeshire	England
MM068	UoY	Star Carr	Scarborough	Yorkshire	England
MM083b	UoY	Star Carr	Scarborough	Yorkshire	England
MM084	UoY	Star Carr	Scarborough	Yorkshire	England
MM085	UoY	Star Carr	Scarborough	Yorkshire	England
MM089	NMSZ	Loch of Marlee	Coupar Angus	Perth and Kinross	Scotland

Table DT2. Primer pairs and their properties used in amplification of a 492bp sequence of the *Castor fiber* control region. Please note that where necessary these were supplemented by primers designed in Horn *et al.* (2014).

Primer	Sequence 5' - 3'	Size (bp)	GC%	Tm	Tm Delta	Product Size (bp)	Hairpin Tm	Self-dimer Tm	Pair-dimer Tm
CF1_Test_F	TGGTCGTTCATTAGCGCATT	20	45	58.3	0.1	99	None	11.5	None
CF1_Test_R	TGGTTTCACGGAGGATGGTA	20	50	58.4					
CF2_Test_F	CGAGCAGGAATAACAAGCACG	21	52	59.9	0.4	89	None	None	None
CF2_Test_R	TCCATGCTGGATTGGTGTTATGT	23	43	60.3					
CF1F	ACCAGCAATGGGGAAACACT	20	50	59.8	0	81	31.8	None	Nonre
CF1R	GAGAATGTCAGCTTTGGGTGC	21	50	59.8					
CF2F	TCAGGGGAGAAACAAAGTCC	20	50	57	0.3	85	None	None	None
CF2R	TGTACATGTATATGTTACAGGGGA	24	38	56.7					
CF3F	TCTCATAATTAAACTATCCCCTGT	24	33	54.8	0.4	85	None	None	None
CF3R	TGTGGAATATAACATAATGCACG	23	35	55.2					
CF4F	ACAACAGTCTATGTATATCGTGCA	24	38	57.9	0.3	92	43	None	1.4
CF4R	ACAGTTATGTACTATGTACGATTAGCA	27	33	58.2					
CF5F	GCTTGAGTACTTCAAATGCTAATCG	25	40	58.8	0.1	88	44.8	4.1	None
CF5R	ACGTGCTTGTTATTCCTGCTC	21	48	58.9					
CF6F	CCCGAGCAGGAATAACAAGC	20	55	59	0.1	90	45	None	None
CF6R	CCATGCTGAATTGGTGTTATGTTG	24	42	59.1					
CF7F	CAACATAACACCAATTCAGCATGG	24	42	59.1	0.2	91	None	None	None
CF7R	GCGCTAATGAACGACCAATAGA	22	45	58.9					
CF8F	TAATCTATTGGTCGTTCATTAGCGC	25	40	59.6	0.2	85	None	5.5	None
CF8R	TCAAAAGTTTGAACTGTGGTAGGG	24	42	59.4					
CF9F	GGATATCCCCTACCACAGTTCA	22	50	58.7	0	86	None	11	0.2
CF9R	AGCGAGAAGAGGGACATTCC	20	55	58.7					
CF10F	GGAATGTCCCTCTTCTCGCT	20	50	59.2	0	86	None	None	None
CF10R	CCTGAAGTAAGAACCAGATGCC	22	55	59.2					
GF1F	ACAACAGTCTATGTATATCGTGCA	24	38	57.9	2.2	130	43	None	None
GF1R	TGTTATTCCTGCTCGGGGAAC	21	52	60.1					
GF2F	AGTACATAACTGTATAATCGTACA	24	29	52.9	2.7	126	39.9	9.9	None
GF2R	ATCCATGCTGAATTGGTGTT	20	40	55.6					
GF3F	CAACATAACACCAATTCAGC	20	40	53.1	0.9	130	None	None	None
GF3R	GGGGATATCCATTTTGAAGA	20	40	52.2					
GF4F	ACAGTTCTCTTCAAAATGGA	20	35	52.3	5.6	130	None	None	0.4
GF4R	AGAAACCCCCAGAACGTATG	20	50	57.9					

Table DT3. BEAST output run parameters (mean, estimated sample size and 95% HDP interval) for posterior, likelihood and prior in all *Castor fiber* analyses.

BEAST Run 1 - Strick Clock, Constant Population Size, Clock Rate 1.0 x 10-7			
Statistic	Mean	ESS	95% HPD Interval
Posterior	-2909.9784	3703.2874	-2945.1235, -2878.2289
Likelihood	-1834.1747	1456.8491	-1850.5276, -1818.4802
Prior	-1075.8037	6082.605	-1106.04, -1048.2284

BEAST Run 1 - Sampled From Prior			
Statistic	Mean	ESS	95% HPD Interval
Posterior	-706.4925	702.3606	-777.8688, -640.3784
Likelihood	0	-	-
Prior	-706.4925	702.3606	-777.8688, -640.3784

BEAST Run 2 - Strick Clock, Constant Population Size, Clock Rate 3.0 x 10-7			
Statistic	Mean	ESS	95% HPD Interval
Posterior	-2819.3984	1636.9396	-2851.1041, -2788.5537
Likelihood	-1837.4882	1287.4533	-1853.8572, -1820.9412
Prior	-198.9102	1682.391	-1013.3962, -954.4429

BEAST Run 2 - Sampled From Prior			
Statistic	Mean	ESS	95% HPD Interval
Posterior	-706.2402	449.1562	-771.9138, -639.567
Likelihood	0	-	-
Prior	-706.2402	449.156	-771.9138, -639.567

BEAST Run 3 - Strick Clock, Constant Population Size, Clock Rate 5.0 x 10-7			
Statistic	Mean	ESS	95% HPD Interval
Posterior	-2792.0537	4400.5821	-2823.654, -2761.2143
Likelihood	-1837.8258	2564.3433	-1855.498, -1822.5854
Prior	-954.2279	4844.5327	-983.0325, -925.3181

BEAST Run 3 - Sampled From Prior			
Statistic	Mean	ESS	95% HPD Interval
Posterior	-2792.0537	4400.5821	-2823.654, -2761.2143
Likelihood	-1837.8258	2564.3433	-1855.498, -1822.5854
Prior	-954.2279	4844.5327	-983.0325, -925.3181

Table DT3 cont. BEAST output run parameters (mean, estimated sample size and 95% HDP interval) for posterior, likelihood and prior in all *Castor fiber* analyses.

BEAST Run 4 - Strick Clock, Exponential Growth, Clock Rate 1.0 x 10-7			
Statistic	Mean	ESS	95% HPD Interval
Posterior	-2904.2543	3286.3938	-2937.5337, -2871.808
Likelihood	-1835.0412	1475.734	-1852.305, -1818.9831
Prior	-1069.2131	5263.8844	-1098.1648, -1043.9718

BEAST Run 4 - Sampled From Prior			
Statistic	Mean	ESS	95% HPD Interval
Posterior	-696.8935	608.0774	-771.8382, -628.8197
Likelihood	0	-	-
Prior	-696.8935	608.0774	-771.8382, -628.8197

BEAST Run 5 - Strick Clock, Exponential Growth, Clock Rate 3.0 x 10-7			
Statistic	Mean	ESS	95% HPD Interval
Posterior	-2821.138	2396.3996	-2854.0872, -2790.0492
Likelihood	-1835.2065	1376.4022	-1852.1513, -1819.2159
Prior	-985.9315	3042.2763	-1014.2751, -960.2765

BEAST Run 5 - Sampled From Prior			
Statistic	Mean	ESS	95% HPD Interval
Posterior	-697.8594	261.1848	-774.8075, -626.2675
Likelihood	0	-	-
Prior	-697.8594	261.1848	-774.8075, -626.2675

BEAST Run 6 - Strick Clock, Exponential Growth, Clock Rate 5.0 x 10-7			
Statistic	Mean	ESS	95% HPD Interval
Posterior	-2788.187	4305	-2819.5129, -2756.9564
Likelihood	-1836.7055	3186.5319	-1853.3376, -1820.9597
Prior	-951.472	4913.2012	-979.5644, -924.687

BEAST Run 6 - Sampled From Prior			
Statistic	Mean	ESS	95% HPD Interval
Posterior	496.1445	7.7369	-312.9745, -1420.7279
Likelihood	0	-	-
Prior	496.1445	7.7369	-312.9745, -1420.7279

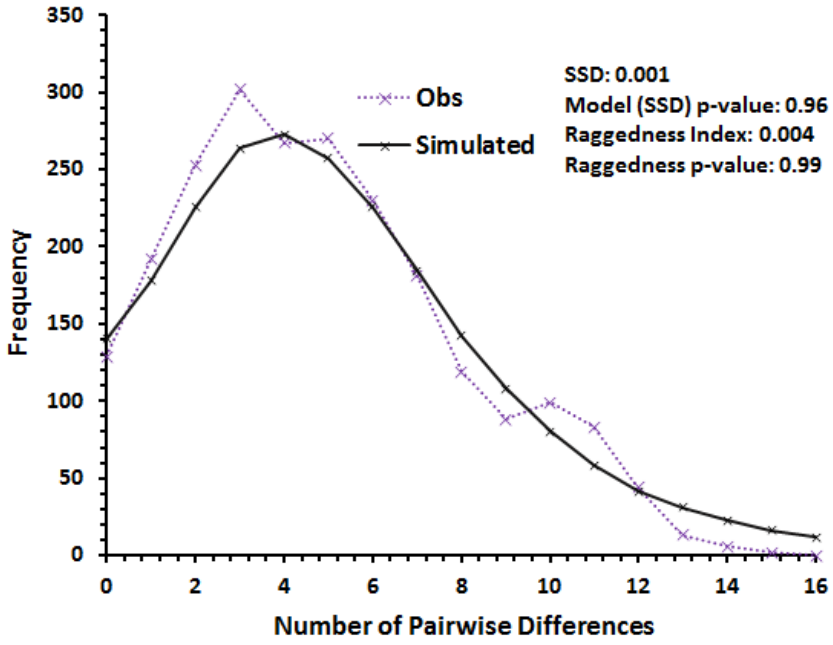
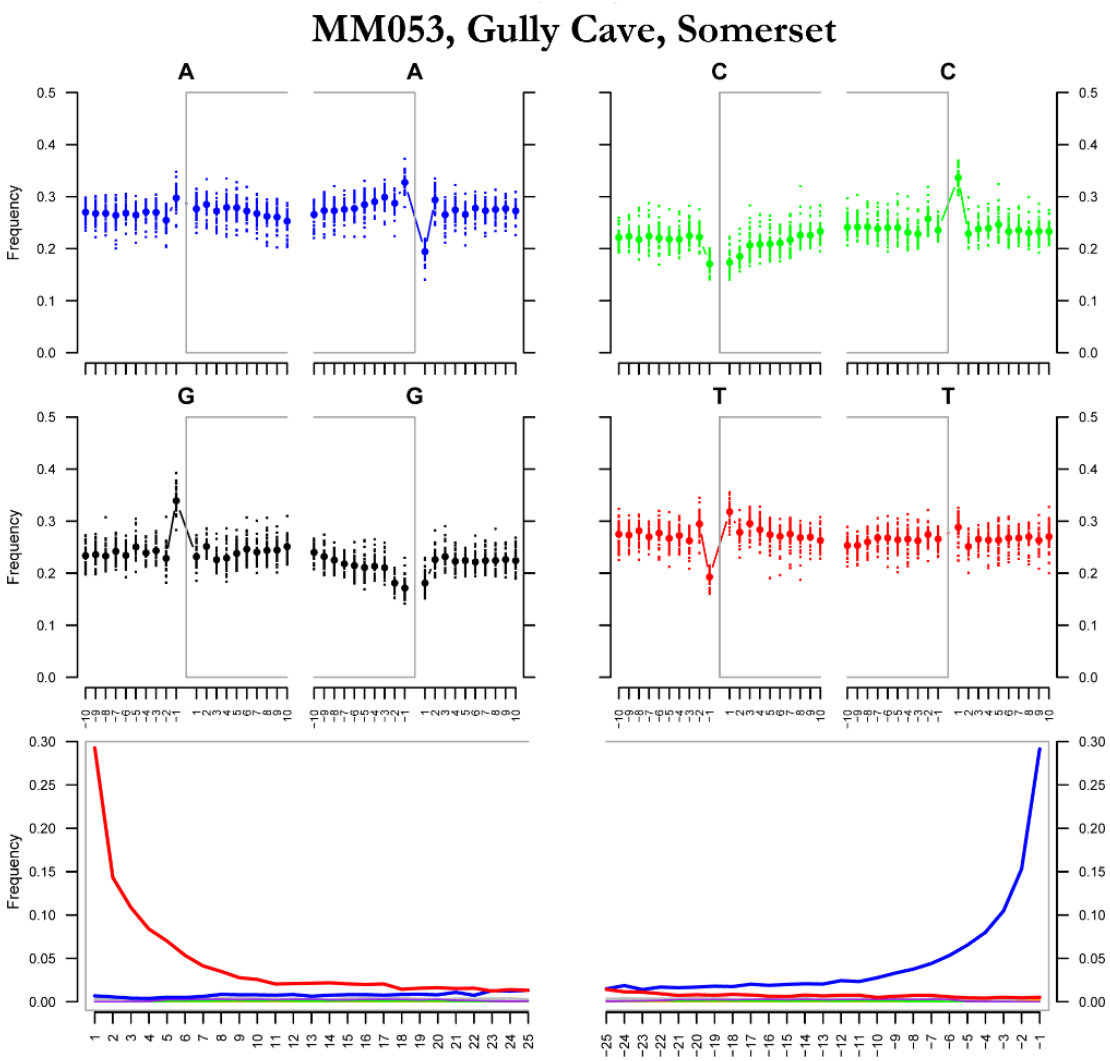


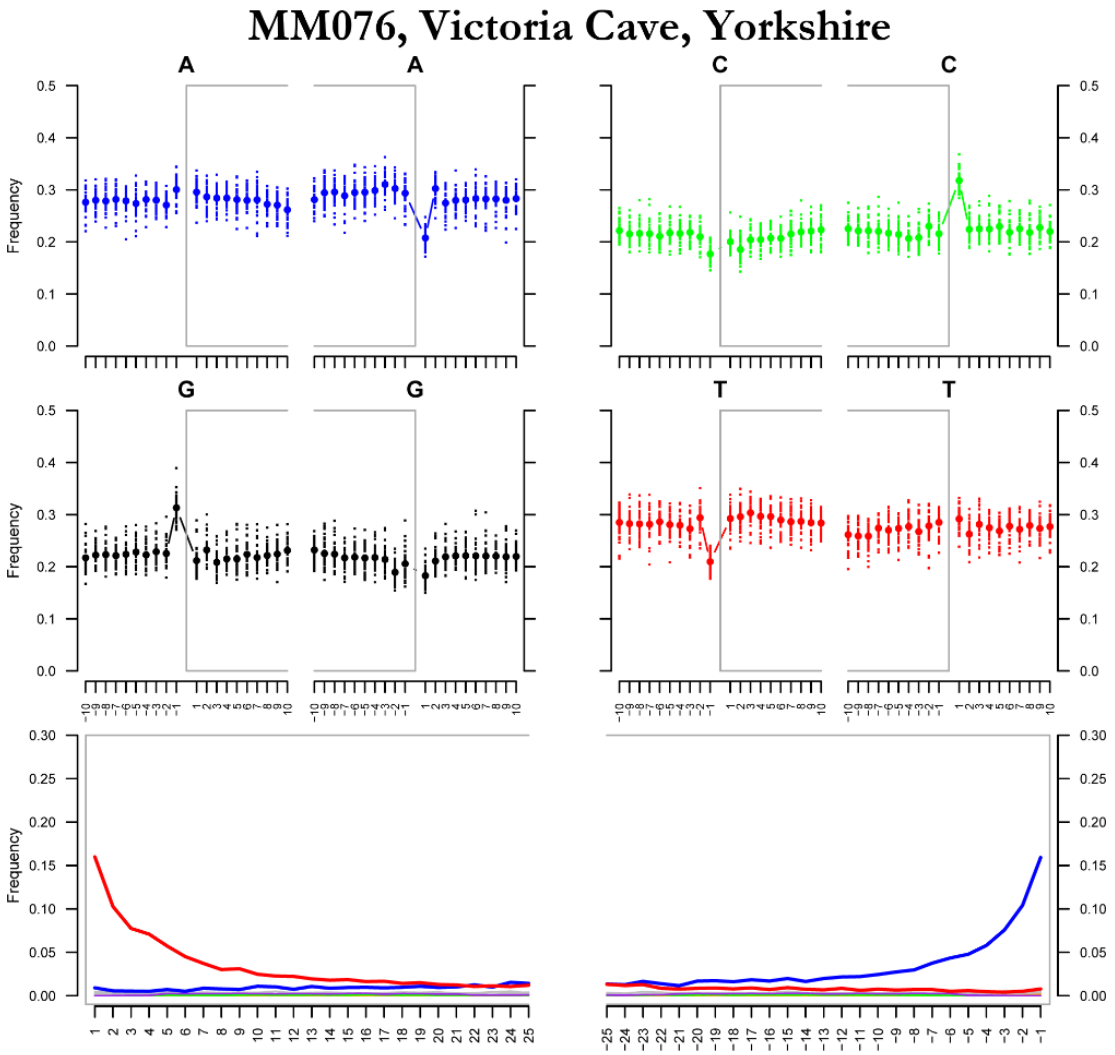
Figure DF1. Mismatch distribution of observed frequencies of pairwise distributions within the Western clade. Simulated data show expected frequencies under a hypothesis of population expansion; observed data show observed pairwise differences in the dataset.

Appendix E (Chapter 6).



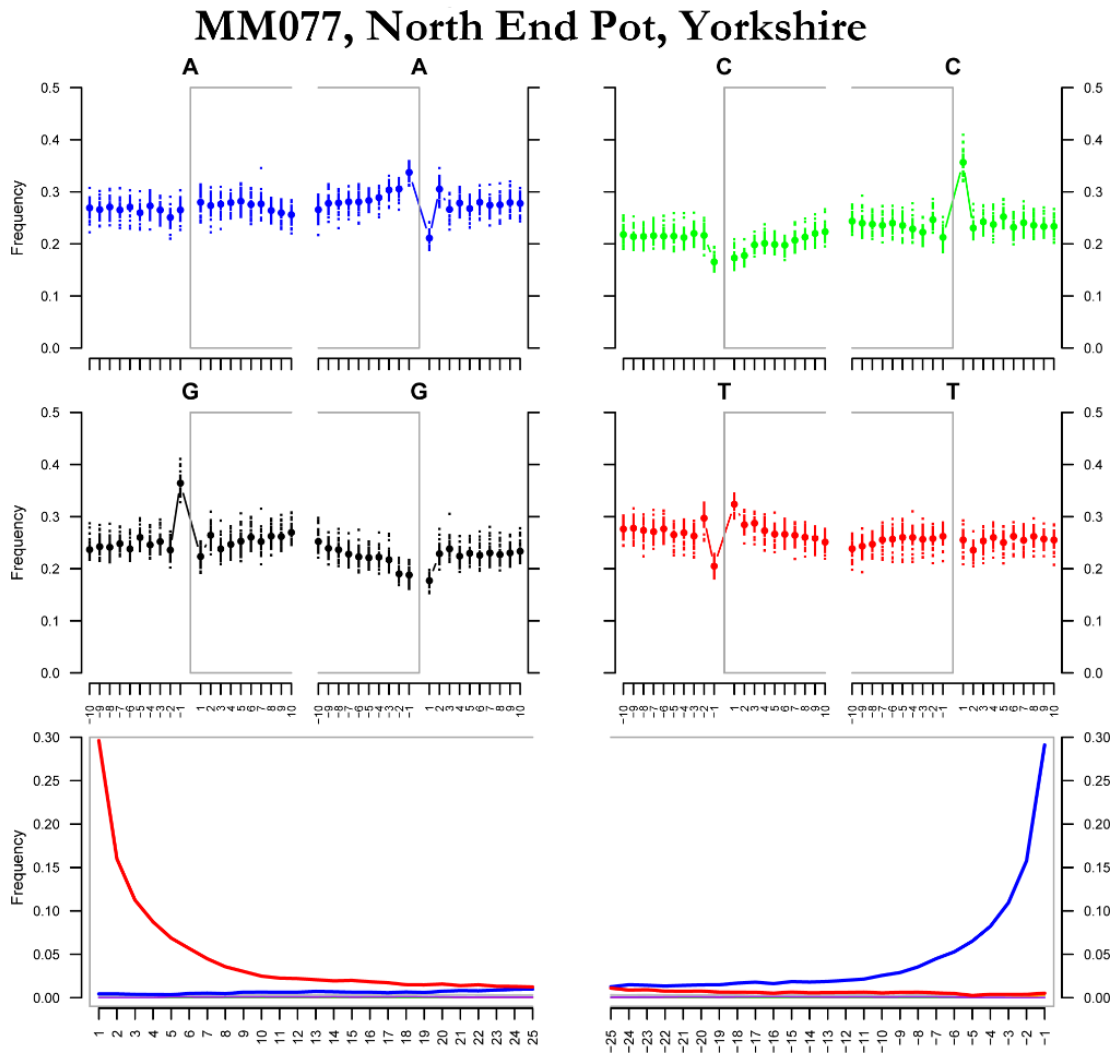
a.

Figure EF1 (a-f). Damage plots for ancient European wildcat *Felis silvestris silvestris* samples in this study estimated with mapDamage2.0 (Jónsson *et al.* 2011). Upper plot: gray brackets indicate strand breaks across which the frequency of nucleotides is shows (y-axis). Lower plot indicates transitions where red is C to T transitions and blue is G to A transitions. X-axis denotes positions along the DNA fragment and the y-axis shows the frequency of sites containing a nucleotide change from the reference sequences (*F. s. catus* genome v8.0).



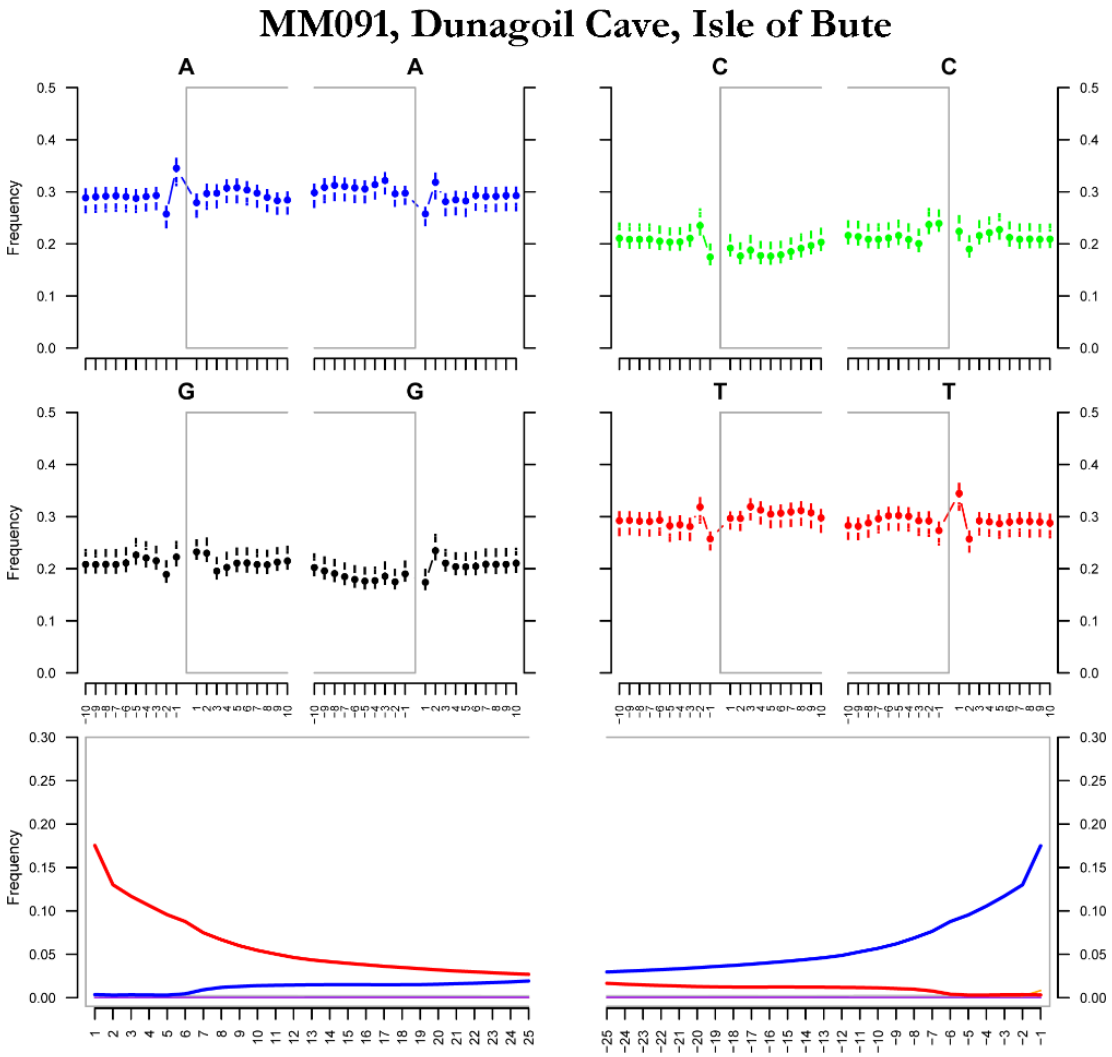
b.

Figure EF1 (a-f) cont. Damage plots for ancient European wildcat *Felis silvestris silvestris* samples in this study estimated with mapDamage2.0 (Jónsson *et al.* 2011). Upper plot: gray brackets indicate strand breaks across which the frequency of nucleotides is shows (y-axis). Lower plot indicates transitions where red is C to T transitions and blue is G to A transitions. X-axis denotes positions along the DNA fragment and the y-axis shows the frequency of sites containing a nucleotide change from the reference sequences (*F. s. catus* genome v8.0).



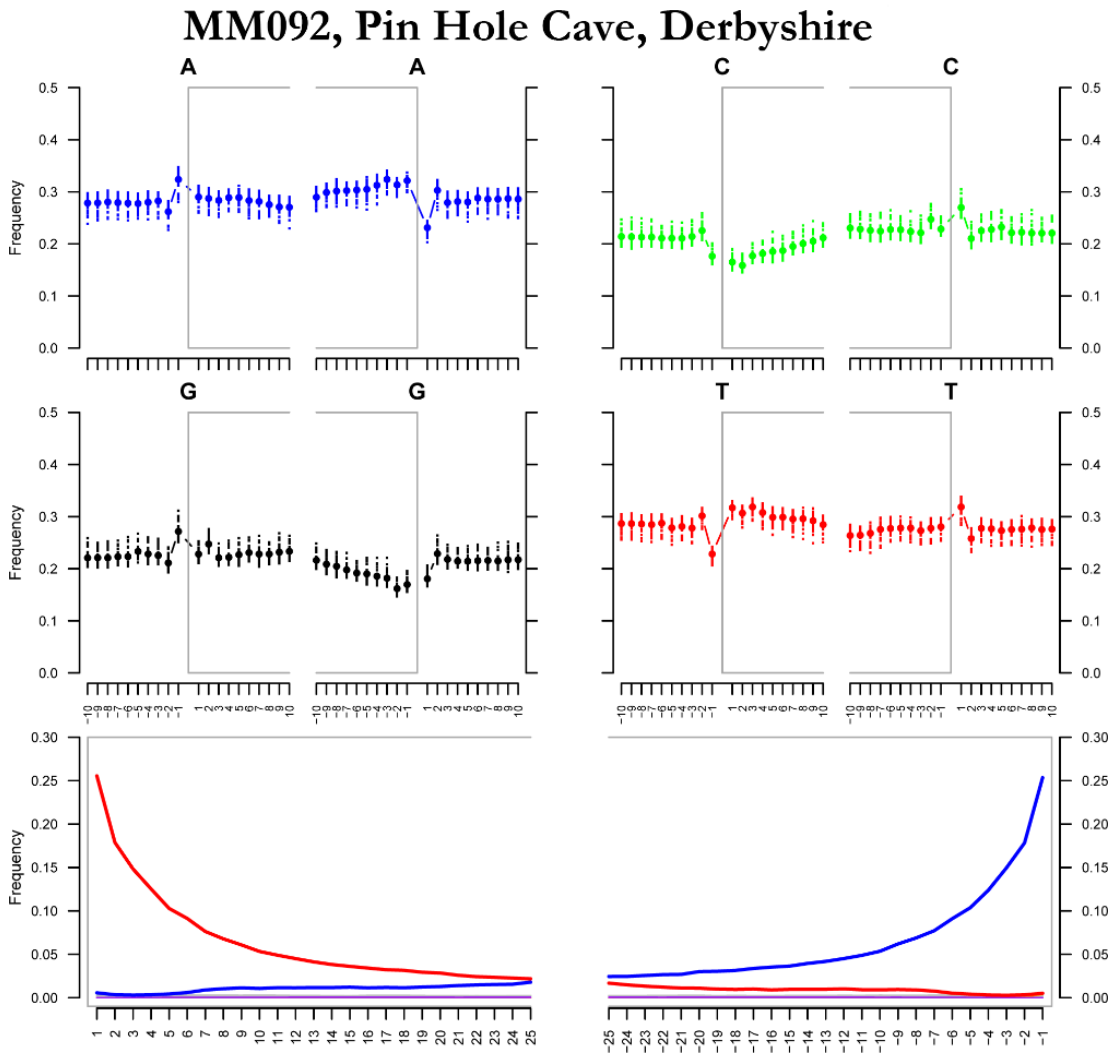
c.

Figure EF1 (a-f) cont. Damage plots for ancient European wildcat *Felis silvestris silvestris* samples in this study estimated with mapDamage2.0 (Jónsson *et al.* 2011). Upper plot: gray brackets indicate strand breaks across which the frequency of nucleotides is shows (y-axis). Lower plot indicates transitions where red is C to T transitions and blue is G to A transitions. X-axis denotes positions along the DNA fragment and the y-axis shows the frequency of sites containing a nucleotide change from the reference sequences (*F. s. catus* genome v8.0).



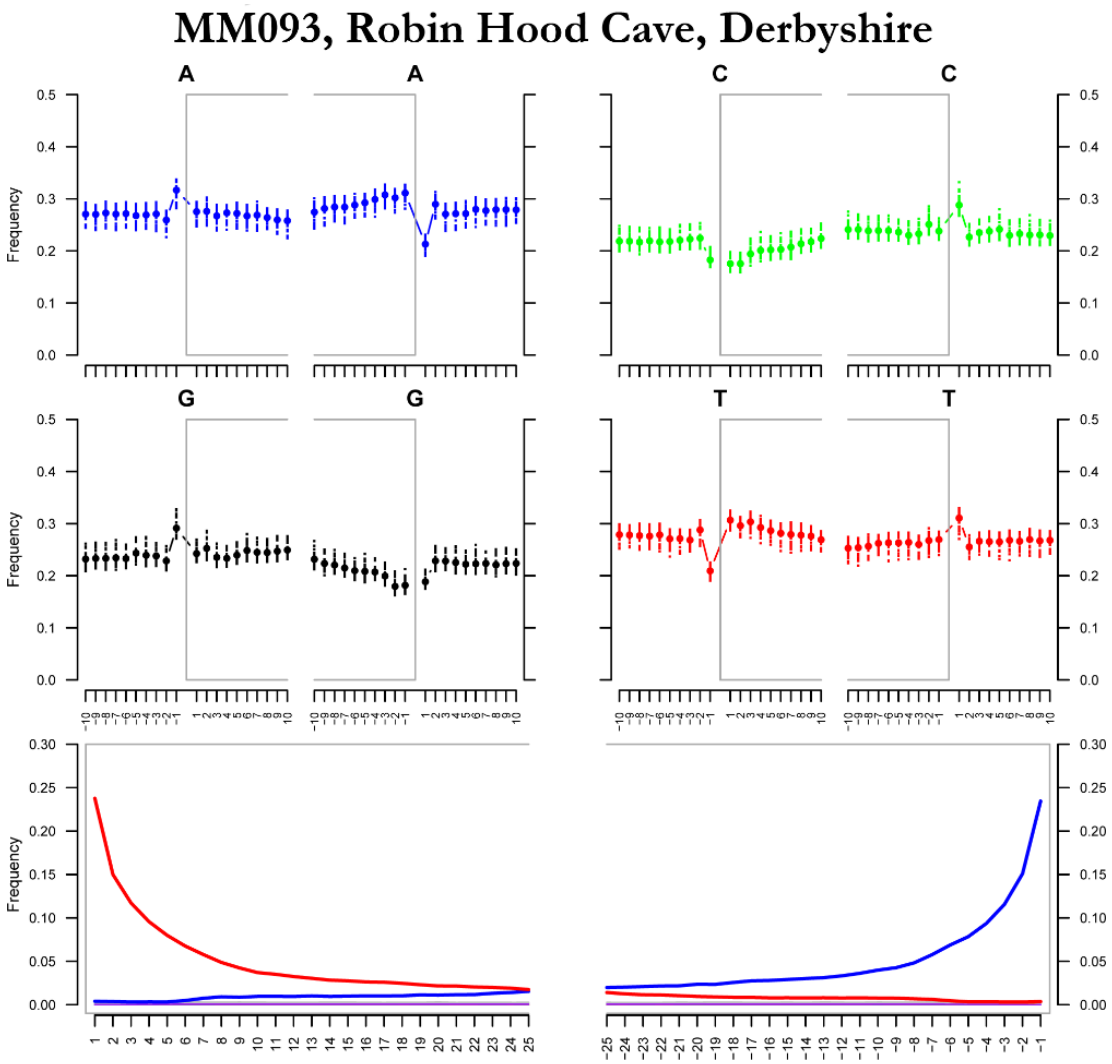
d.

Figure EF1 (a-f) cont.. Damage plots for ancient European wildcat *Felis silvestris silvestris* samples in this study estimated with mapDamage2.0 (Jónsson *et al.* 2011). Upper plot: gray brackets indicate strand breaks across which the frequency of nucleotides is shows (y-axis). Lower plot indicates transitions where red is C to T transitions and blue is G to A transitions. X-axis denotes positions along the DNA fragment and the y-axis shows the frequency of sites containing a nucleotide change from the reference sequences (*F. s. catus* genome v8.0).



e.

Figure EF1 (a-f) cont. Damage plots for ancient European wildcat *Felis silvestris silvestris* samples in this study estimated with mapDamage2.0 (Jónsson *et al.* 2011). Upper plot: gray brackets indicate strand breaks across which the frequency of nucleotides is shows (y-axis). Lower plot indicates transitions where red is C to T transitions and blue is G to A transitions. X-axis denotes positions along the DNA fragment and the y-axis shows the frequency of sites containing a nucleotide change from the reference sequences (*F. s. catus* genome v8.0).



f.

Figure EF1 (a-f) cont. Damage plots for ancient European wildcat *Felis silvestris silvestris* samples in this study estimated with mapDamage2.0 (Jónsson *et al.* 2011). Upper plot: gray brackets indicate strand breaks across which the frequency of nucleotides is shows (y-axis). Lower plot indicates transitions where red is C to T transitions and blue is G to A transitions. X-axis denotes positions along the DNA fragment and the y-axis shows the frequency of sites containing a nucleotide change from the reference sequences (*F. s. catus* genome v8.0).