University of Huddersfield Repository

Dulias, Katharina

ARCHAEOGENETICS AND PALAEOGENETICS OF THE BRITISH ISLES

Original Citation

Dulias, Katharina (2019) ARCHAEOGENETICS AND PALAEOGENETICS OF THE BRITISH ISLES. Doctoral thesis, University of Huddersfield.

This version is available at http://eprints.hud.ac.uk/id/eprint/34890/

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

http://eprints.hud.ac.uk/

UNIVERSITY OF HUDDERSFIELD

DOCTORAL THESIS

ARCHAEOGENETICS AND PALAEOGENETICS OF THE BRITISH ISLES

Author:
Katharina DULIAS

Supervisor: Dr.
Ceiridwen J. EDWARDS

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

in the

Archaeogenetics Research Group
School of Applied Sciences

University of HUDDERSFIELD

Copyright Statement©

1. The author of this thesis (including any appendices and/or schedules to this thesis) owns any copyright in it (the "Copyright") and s/he has given The University of Huddersfield the right to use such copyright for any administrative, promotional, educational and/or teaching purposes.

2. Copies of this thesis, either in full or in extracts, may be made only in accordance with the regulations of the University Library. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.

3. The ownership of any patents, designs, trademarks and any and all other intellectual property rights except for the Copyright (the "Intellectual Property Rights") and any reproductions of copyright works, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

Declaration of Authorship

I, Katharina DULIAS, declare that this thesis titled, "ARCHAEOGENETICS AND PALAEOGENETICS OF THE BRITISH ISLES" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"Only two things are infinite, the universe and human stupidity, and I'm not sure about the former."

Albert Einstein

"Do I contradict myself?
Very well, then I contradict myself.
I am large, I contain multitudes."

Walt Whitman

UNIVERSITY OF HUDDERSFIELD

Abstract

Biological and Geographical Sciences
School of Applied Sciences

Doctor of Philosophy

ARCHAEOGENETICS AND PALAEOGENETICS OF THE BRITISH ISLES

by Katharina DULIAS

Although it is well known that different waves of migration have shaped the population of the British Isles through time, little is known about these population movements over the past 12,000 years. This time period includes the recolonisation of the British Isles by the first hunter-gatherers, and the subsequent arrival of the Neolithic peoples, to the Roman conquest and Anglo-Saxon and Viking migrations, all of which have shaped the genetic diversity of the modern population. However, although these population movements are seen in the archaeological and historical record, it is mostly unknown to what extent different populations arrived into Britain, or from where exactly they originated. The timing and scale of these events are as unknown as what became of the indigenous populations. The recent advance of high-throughput DNA sequencing, especially in the field of ancient DNA, together with the increased number of genome-wide sequencing data of ancient humans from various regions and eras, makes larger scale studies of specific ancient populations, such as those of the British Isles, possible. These new techniques now make it feasible to generate considerable amounts of data from large sample sizes of archaeological human remains, allowing questions to be addressed at population level. The use of uniparental markers (mtDNA and Y-chromosome) at high resolution can show a fine-scale picture of the genetic diversity and origin of ancient and modern-day populations of the British Isles, and add to the information gained through genome-wide sequencing. My project highlights population turnover during the Neolithic to Bronze Age transition, in the Scottish Isles, identifies possible Near Eastern/North African ancestry in a Bell Beaker individual from northeastern England, answers questions about the kin relationship between individuals in an Iron Age cave burial in Scotland and within cemeteries from the Bronze

Age, Iron Age and the post-Roman era, and shows the differences in mitochondrial and Y-chromosome haplogroup frequencies across different time periods to the modern-day British population. The provided time-resolved data has allowed analysis of distinct migrations to the British Isles and the genetic composition of its populations throughout different time periods.

Acknowledgements

First of all, I would like to thank Dr. Ceiridwen J. Edwards for her marvellous supervision and support over the last years, as well as Professor Martin Richards for being my co-supervisor and all the insightful discussions. I learned a lot from you.
I also thank Dr. Maria Pala for her co-supervision and input on the modern DNA analysis of this project.
My profound gratitude goes to Dr. Jarek Bryk and Professor Dan Bradley for all their comments and suggestions.
I am very grateful for the time I could spend at the Molecular Population Genetics Group of the Smurfit Institute of Genetics at Trinity College Dublin, especially Dr. Valeria Mattiangeli for her time and dedication for ancient DNA extraction from bones.
I want to thank Steven Birch (University of Aberdeen), Prof. Dame Sue Black (Lancaster University), Prof. Andrew Chamberlain (University of Manchester), Garrard Cole (UCL), Bob Kenyon, George Kozikowski (West Coast Archaeological Services Scotland), Hazel Moore (Environment and Archaeology Services Edinburgh), Jenny Murray, Prof. Dominic Powlesland (University of York), Dr. Clive Waddington (Archaeological Research Services) and Graeme Wilson (Environment and Archaeology Services Edinburgh) for the contribution of ancient samples, as without you this work would not have been impossible.
I thank Dr. Peter Ditchfield (University of Oxford) for undertaking the dietary stable isotope analysis and radiocarbon dating.
I thank Drs. Francesca Gandini (University of Huddersfield), Pedro Soares and Teresa Rito (both University of Minho, Braga) for all their work on the phylogenetic trees, Prof. Jim Wilson (University of Edinburgh) for the provision of modern samples from the British Isles, Prof. David Reich (Harvard Medical School) for his collaboration on the post-Roman and Saxon samples, Dr. Antonia Flaquer (Ludwig-Maximilians University Munich) and Prof. Walther Parson (Innsbruck Medical University) for the provision of modern samples from Germany and the Netherlands.
I would like to thank the Leverhulme Trust Doctoral Scholarship programme to fund my doctorate degree.
I thank Dr. Rui Martiniano for his tremendous help with the bioinformatic analysis, as well as Dr. Lara Cassidy. You have made my life much easier!
My profound gratitude goes to Dr. Stefano Vanin! You have been an amazing advisor and your support, especially in the most discouraging moments,

over the last three years, was very much appreciated.
I want to thank M. George B. Foody for his incredible patience whenever I was fighting with programming and for always helping me out. Your support means a lot to me.
Dr. Cooper Grace, I cannot thank you enough for saving me from tossing my computer out of the window when I got too frustrated with silly syntax errors in LaTeX. Your support during the writing-up was incredibly helpful and kept me sane.
I want to thank Rohan Raval for always keeping me motivated and cheering me up when I got frustrated. Your help with R was much appreciated.
I thank all the members of the Archaeogenetics group and the Evolutionary Genetics groups for all the constructive discussions during our meetings: Marisa, Bernardo, Ginés, Jade, Alessandro, Marina, Bobby, Gonzalo, Rita, Simão and Pierre.
Giorgia Giordani, Fabiola Tuccia and Jennifer Pradelli, I cannot thank you enough for all your support! I will forever be grateful for your friendship.
I thank Joseph Mwamba Mwansa for being there throughout the biggest stress, for always supporting me and pushing me to be the best I can be.
Last but not least, I would like to thank my family. Without your support I would not be who I am today.

Contents

List of Figures

List of Tables

For my parents

Chapter 1

Introduction

1.1 Human population genetics

The study of the genetic composition of biological populations and their variation is called population genetics. The founders of the field in the 1920s and 1930s were Fisher, Haldane and Wright (Wright, 1949), who developed the first formal models to explore how evolutionary forces impact the genetic composition of a population (Okasha, 2016; Charlesworth et al., 2017). The first studies on population genetics used quantitative traits, such as large chromosomal polymorphisms (Brncic, 1956) and phenotypic serological variation (Cavalli-Sforza et al., 1967), enabling the assessment of the relationships between different human populations. Since the first studies, using "classical" markers such as blood groups rather than molecular markers, Cavalli-Sforza and his colleagues dominated the field of human population genetics (Cavalli-Sforza et al., 1967). The development of Sanger sequencing lifted the field to a new level by allowing direct analysis of DNA variation (Sanger et al., 1977), whilst the invention of the polymerase chain reaction (PCR) (Mullis et al., 1986), which allows amplification of specific small regions of the genome, plus the automation of Sanger sequencing (Smith et al., 1986) further aided analysis. This improvement of sequencing techniques facilitated detailed molecular studies of uniparental markers, such as mitochondrial deoxyribonucleic acid (mtDNA) and the male-specific part of the Y-chromosome (MSY). However, the expense of sequencing multiple genomic loci remained high, and was a major limiting factor for the analysis of larger genomic regions until the recent development of next-generation sequencing techniques (NGS) (Ansorge, 2009; Scholz et al., 2012). Next-generation sequencing has led to the possibility of undertaking large-scale whole-mtDNA sequencing and the generation of complete human genomes. The sequencing of the first whole human genome, the Human Genome Project (Sawicki et al., 1993), began in the 1990s using Sanger sequencing, and the results of this

incredible effort were presented in 2001 (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001) and completed in 2004. Since the development of NGS techniques, several projects have aimed to map human genome variation on a large scale, such as the Human Genome Diversity Project (HGDP), which started in the 1990s (Weiss et al., 1992), the 1,000 Genome Project (Siva, 2008; 1000 Genomes Project Consortium, 2010), and the more recent genome diversity projects of the British Isles: UK10K (The UK10K Consortium, 2015) and the sequel study UK100K, which is currently on-going. While the HGDP and the 1,000 Genomes Project were aiming to map the genetic variation across human populations worldwide, the UK10K project focused on the identification of rare genetic variation through whole genome sequencing of 10,000 genomes from the population of the United Kingdom (UK) (The UK10K Consortium, 2015). After the success of this project, a follow-up study was started by Genomics England, which is currently sequencing 100,000 more British genomes.

The analysis of human history using molecular genetic analysis was defined by Renfrew (2000) as "archaeogenetics". One of the major questions in this subject has been the spread of the Neolithic from the Near East into Europe and its impact on the European population. The focus has since expanded to include most other periods of human history, such as the first arrival of modern humans into Europe and Asia, diversity in early hunter-gatherer populations, the impact of climate change at the end of the ice age, human dispersals out of Ice Age refugia, Bronze Age migrations of the Bell Beaker culture, Roman and Viking invasions, to fine-scale population structure analysis (e.g. Richards et al. 2000; Helgason et al. 2001; Weale et al. 2002; Capelli et al. 2003; Goodacre et al. 2005; Pala et al. 2012; Leslie et al. 2015; Gandini et al. 2016; Gilbert et al. 2017a; Gilbert et al. 2017b). The discipline of archaeogenetics has widely attempted to map the demographic history of the British Isles, but, until recently, with the publication of the People of the British Isles (PoBI) study (Leslie et al., 2015) (discussed in more detail in Appendix B), and a very similar study using the same approach to investigate the fine-scale population structure of Ireland (Gilbert et al., 2017a; Gilbert et al., 2017b), no clear results were available (as summarised in Silva et al. 2019).

1.2 Phylogeography

Phylogeography is a highly integrative approach combining information from population genetics, biogeography, molecular evolution and phylogenetics

in order to describe distributions of genetic lineages across a specific geographic area of interest (Avise et al., 1987). The phylogeographic patterns identified can be used to infer demographic, evolutionary and biogeographic processes (Arbogast, 2001). Mitochondrial DNA diversity has been used to determine phylogenetic relationships among lineages, which can then be further analysed for their geographical distribution and time depth. The combination of phylogenetics with the geographic distribution of genetic lineages can illuminate the divergence and migration history of the species of interest (Hewitt, 2001). A phylogeographic analysis of the British Isles and Europe based on whole mitochondrial genomes can be found in Appendix B and C, respectively.

1.3 Uniparental markers

Uniparental markers, such as mitochondrial DNA (mtDNA) and the Y-chromosome, are used to analyse phylogeographic patterns within species.

1.3.1 Mitochondrial DNA

The human mitochondrial genome is circular, double-stranded and 16,568 base pairs (bp) long. It occurs in multiple, identical copies per cell and the genetic structure is highly conserved (Taanman, 1999). MtDNA is inherited through the maternal line of descent, without recombination, and naturally accumulates mutations over time. This mutation rate can be determined and used to calculate the "mitochondrial molecular clock", which allows for calculations of the ages of different genetic lineages (Soares et al., 2009). This time-dependent mutation rate takes into account the effect of purifying selection on more recent nodes of the complete human mitochondrial tree (Soares et al., 2009). A set of linked alleles, or single nucleotide polymorphisms (SNPs), is called a haplotype, and a cluster of haplotypes sharing a single common ancestor and defined by one or more base variants or SNPs is classified as a haplogroup. The first human mitochondrial genome sequenced (Anderson et al., 1981) was mainly of an individual from the United Kingdom. After re-sequencing, it was revised (Andrews et al., 1999) and this corrected version, known as the rCRS (revised Cambridge Reference Sequence; GenBank: NC_012920), is nowadays used as the reference sequence for mitochondrial genome studies.

1.3.2 Y-chromosome

The male counterpart to mtDNA is the Y chromosome, which is used to trace the paternal line of descent. More than 95% of the Y-chromosome is male-specific, with a total length of ∼58 million base pairs (bp) (Skaletsky et al., 2003). Unlike other chromosomes, the Y-chromosome largely escapes recombination and is usually passed on intact from generation to generation, only changing by mutation (Jobling et al., 2003). As SNPs have low mutation rates, they can provide an extremely detailed phylogenetic tree of Y-chromosomal lineages, and can thus be used to construct unique phylogenies based on the same classification of haplotypes and haplogroups as mitochondrial DNA. Local differentiation of the Y-chromosome can be enhanced as a consequence of patrilocality (Jobling et al., 2003; McEvoy et al., 2008; Gilbert et al., 2017a; Gilbert et al., 2017b).

1.4 Ancient DNA

Genetic population history was, for many years, reconstructed using primarily modern day genetic variation, with, in the case of mitochondrial genomes and Y-chromosomes, estimates of differences through time being calculated through the use of molecular clocks. Archaeogenetic analysis based on this approach, for example using founder analysis to estimate arrival times of populations in a gene pool, dominated the field for several years (see Richards et al. 2008). However, with the advance of NGS techniques, and ever-improving methods in the field of ancient DNA (aDNA), a new dimension has been added to the reconstruction of population history. We can now directly analyse the genetic diversity at specific points in time and space (Hofreiter et al., 2001).

With the possibility of extracting endogenous DNA from ancient tissues of a variety of remains, the field of palaeogenetics emerged. Svante Pääbo pioneered this field for ancient human DNA following the successful amplification of quagga DNA by Higuchi et al. (1984). Since then, studies using aDNA to reconstruct past genomes and facilitate interpretations of the human past have increased exponentially (Rizzi et al., 2012). In 2018 alone (up to September 2018), 772 new whole human genome sequences were published, as well as 102 complete mitogenomes of ancient individuals (Brace et al., 2018; de Barros Damgaard et al., 2018; Mathieson et al., 2018; Mittnik et al., 2018; Neparaczki et al., 2018; Olalde et al., 2018; Valdiosera et al.,

2018), giving incredible new insights into the human past. To date, the oldest sample of successfully amplified and verified aDNA is from a horse found in permafrost dating to 780-560 thousand years before present (kyr BP) (Orlando et al., 2013). This sample likely represents the maximum age range of retrievable ancient DNA. Most ancient sequences analysed date to the Palaeolithic and younger.
Nevertheless, there are several provisos on the use of aDNA, with preservation being the major problem. Bones have been deposited in various ways in different environments, and are thus exposed to varied environmental factors (Winther Pedersen et al., 2015). DNA degradation can be caused by bacterial and fungal DNases, UV radiation, heat, and chemical modifications, such as oxidation, deamination, depurination and other hydrolytic processes (Hofreiter et al., 2001; Willerslev et al., 2005; Winther Pedersen et al., 2015), leading to highly fragmented and modified DNA. One of the best environments for aDNA preservation is permafrost (Hofreiter et al., 2001; Willerslev et al., 2005; Winther Pedersen et al., 2015) due to the stable low temperature. Generally, the survival of aDNA in temperate climates can also be good, and there are several outstanding publications reporting complete and partial ancient genomes of hominins, animals and plants (for example Krings et al. 1997; Pääbo et al. 2004; Coolen et al. 2006; Noonan et al. 2006; Orlando et al. 2006; Gamba et al. 2014; Epp et al. 2015; Günther et al. 2015; Lazaridis et al. 2016). The most problematic depositional areas are wet and warm climates, because these conditions facilitate fast DNA degradation. Another problem is the geographic distribution and availability of ancient samples. Concerning human remains, different burial rituals can also skew the reconstruction of ancient populations due to stochastic inhumation and preservation.

1.4.1 Ancient DNA from bones and teeth

Because they are often the only physical remains surviving, bones are the main source of DNA information from human and other animals. Bone degradation has been well studied in the decades since the discovery that proteins and DNA survive in ancient tissue (Tuross, 1994; Collins et al., 2002). The basic building block of bone is the mineralised collagen fibril, which, together with bone vascularisation, determines the bone microstructure through its organisation (Collins et al., 2002; Mrevlishvili et al., 2005). The bone matrix consists of several highly mineralised discontinuity layers, called cementing lines (Campos et al., 2012), which are the separating anatomical units. Bone

is composed of hydroxyapatite (HAP) and organic material, such as collagen, osteocalcin and osteonectin (Campos et al., 2012). Collagen promotes the stabilisation of the DNA double helix and its hydration shell through hydrogen bond formation (Mrevlishvili et al., 2005), which might also prolong DNA survival (Campos et al., 2012). After death, hydrolysis starts to degrade the collagen, leaving the DNA unprotected (Collins et al., 2002), with the DNA surviving in the bioapatite crystals in the bone structure (Campos et al., 2012). This highlights that the inorganic component of bone is at least as important for the long-term survival of DNA as the organic component, with DNA being trapped in the osteoid as it forms (Collins et al., 2002; Campos et al., 2012). Four possible locations for DNA survival in bones have been postulated, which can be separated into two forms of preservation: *in vivo* (where DNA is overlain or encapsulated by HAP) and post-mortem (where DNA is released from collagen fibrils into the macroporosity of the bone during degradation or trapped by HAP during bone mineral re-crystallisation) (Campos et al., 2012). Considering the role of bone apatite for DNA preservation, a stable, slightly alkaline pH in the environment makes limestone caves a very good source for aDNA, especially as they also have an almost constant cold temperature (Collins et al., 2002; Edwards, 2012). The key mechanism for the fossilisation of bones is the slow loss of collagen by chemical hydrolysis, coupled with infilling of subsequent porosity (Collins et al., 2002). This is dependent on time, temperature and the pH of the environment, but mainly on the hydrology of the site, as mineral transformations will expose collagen to an accelerated chemical and biological degradation (Collins et al., 2002). The most common mechanism is the degradation caused by microbial activity (biodegradation), which is optimised at near-neutral pH conditions that otherwise protect the bone (Collins et al., 2002). In general, the survival of DNA in bone is strongly dependent on the differences in burial rituals and burial environments.

As well as bone, teeth can be used for DNA studies. They have a comparable preservation to that seen in the petrous bone (Hansen et al., 2017), discussed below. The tough enamel and cementum of whole teeth, or tooth roots embedded in the mandibular or maxillary bone, can provide improved protection for DNA in many environments and, additionally, DNA appears less prone to contamination from modern DNA (Adler et al., 2011).

1.4.2 Petrous bone

Several different bone elements have been tested for endogenous DNA retrieval. In general, harder, denser bone preserves endogenous DNA better (Haynes et al., 2002). The hardest bones in the human body are found in the skull. The average success rate of DNA recovery of various cranial bones varies between 52-68% for the parietal bone, occipital bone and the frontal bone (Pinhasi et al., 2015). The hardest and densest bone part is the temporal bone, which is located at the base of the skull between the sphenoid and occipital bones (Gamba et al., 2014; Pinhasi et al., 2015). The petrous part of the temporal bone comprises the otic capsule (bony labyrinth) of the inner ear (see Figure 1.1 A and B red square), and the inner petrous portion of the temporal bone has allowed extraction of endogenous DNA of up to 90% (Pinhasi et al., 2015). Since the density throughout the petrous bone is not uniform, the endogenous DNA content of different sections has been tested (Pinhasi et al., 2015); with the densest bone parts within the otic capsule providing the highest endogenous aDNA yields (Pinhasi et al., 2015).

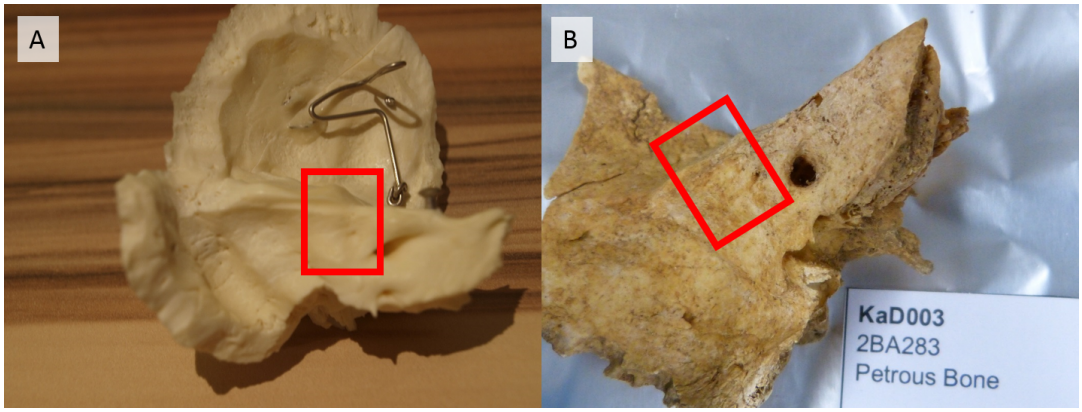

FIGURE 1.1: The petrous part of the temporal bone. A - a modern left petrous bone; B - an ancient left petrous bone. The red square indicates the position of the inner petrous portion, comprising of the otic capsule.

It has been shown that petrous bones from warmer regions of the world can contain surviving DNA, allowing analyses of samples from hot environments (Pinhasi et al., 2015). A more recent study compared the preservation of DNA in the petrous bone with that of tooth cementum, with the petrous bone generally containing higher endogenous DNA yields (Hansen et al., 2017), showing that both substrates are excellent for ancient genomic research.

1.5 Dietary stable isotopes

The analysis of carbon and nitrogen stable isotopes for dietary reconstruction is commonly undertaken on dentine or bone collagen. As collagen is mostly directly synthesised from dietary protein, its isotopic composition thus reflects the main protein source. Due to the different turnover times of collagen in different types of bones, different bones reflect the diet of an individual at different life stages (Chenery et al., 2010). During childhood, the dietary isotopic signature at the time of tooth formation is preserved in the dentine, whereas dietary isotopic signatures of an individual's adulthood are preserved within the growth of long bones during adolescence and following decades (Chenery et al., 2010). Dietary reconstruction uses carbon stable isotope ratios ($\delta^{13}C$) to distinguish between major food sources, such as C3 and C4 plants, and the products of animals feeding on these plants (Chenery et al., 2010). In temperate climates, they are mainly used as differentiation between marine food and terrestrial C3 plant-based foods. Nitrogen stable isotope ratios ($\delta^{15}N$) are used to complement $\delta^{13}C$, as they give an indication of the trophic level an organism is feeding at, which allows an estimation of the importance of animal and plant-based foods in an organism's diet (Chenery et al., 2010). The reconstruction of diets of ancient populations using carbon and nitrogen stable isotope analysis has illustrated significant differences in diet across populations (Schulting et al., 2002a; Evans et al., 2006; Müldner et al., 2007; Chenery et al., 2010).

1.6 The peopling of Britain

1.6.1 First modern humans

The settlement history of the British Isles is full of unanswered questions. Although the first modern human remains date to before the Last Glacial Maximum (LGM), ~26.5-19.5 kya (Barton, 1999), little is known regarding migration, admixture and continuous occupation of Britain during the Pleistocene and subsequent Holocene. During this time, with the relatively isolated peripheral position of the islands, facing the Atlantic at one edge of the European peninsula, the British Isles were only accessible by sea and primarily from one direction, namely the European mainland (France, Belgium, Netherlands, Germany, Denmark). It is likely that *Homo sapiens* in Britain disappeared by ~24,000 years BP, driven out by the climatic conditions of the

LGM.
The earliest dated modern human remains in the British Isles are from Goat's Hole Cave, Paviland, in Wales. Part of a ceremonial burial, and erroneously originally entitled the "Red Lady", the skeleton was of a young adult male, dating to the Early Upper Palaeolithic ($25,840 \pm 280$ cal. BP; OxA-8025) (Aldhouse-Green et al., 1998). The individual had been buried with a handaxe, discoidal cores, leaf points and carinated scrapers, making this by far the richest Aurignacian site in Britain (Aldhouse-Green et al., 1998; Dinnis, 2012). The population to whom the "Red Lady" belonged colonised Britain in the Early Upper Palaeolithic, but then most likely moved south in front of the extending ice caps (Figure 1.2),(Aldhouse-Green et al., 1998). Populations substantially declined and only temporary (possibly seasonally) occupied northern Europe during the LGM, rapidly expanding into northern areas as soon as the temperatures increased (Blockley et al., 2006).
The onset of warmer conditions after ∼13,000 BP (Windermere Interstadial) facilitated the earliest reappearance of hunter-gatherers in the British Isles (Housley et al., 1997). Human remains from this period are extremely rare, especially ones with associated radiocarbon dates. With radiocarbon ages of *ca*. 12,600 cal. BP, Gough's Cave and Sun Hole (Cheddar Gorge, Somerset), Aveline's Hole (Mendip, Somerset) and Pixie's Hole (Chudleigh Valley, Devon) have the oldest Late Glacial determinations of human presence in the British Isles (Jacobi et al., 2009). During this period the Creswellian culture (the British Late Magdalenian) emerged and most human activity was preserved in the form of modifications and cut marks on bones, worked antler and ivory (Barton, 1999), as well as stone blades and bladelets (Wygal et al., 2014). By the time of the Younger Dryas (∼12,900-11,700 BP), the climate had once again returned to much colder conditions, and there is no reliable evidence of human activities from this period, possibly because populations were driven out of Britain due to the cold (Barton et al., 2003; Cunliffe, 2013). It is as yet unclear if occupation of the British Isles continued throughout the Younger Dryas or not.

1.6.2 Mesolithic

Only at the end of the Younger Dryas can more confidence be attributed to the presence of humans in the British Isles. Star Carr, the "type site" for the British Mesolithic, has been been argued to have been a specialised site for the working of antlers (Mithen, 1999). Radiocarbon dating places the occupation of this Mesolithic site to $10,920 \pm 80$ cal. BP and $10,740 \pm 130$

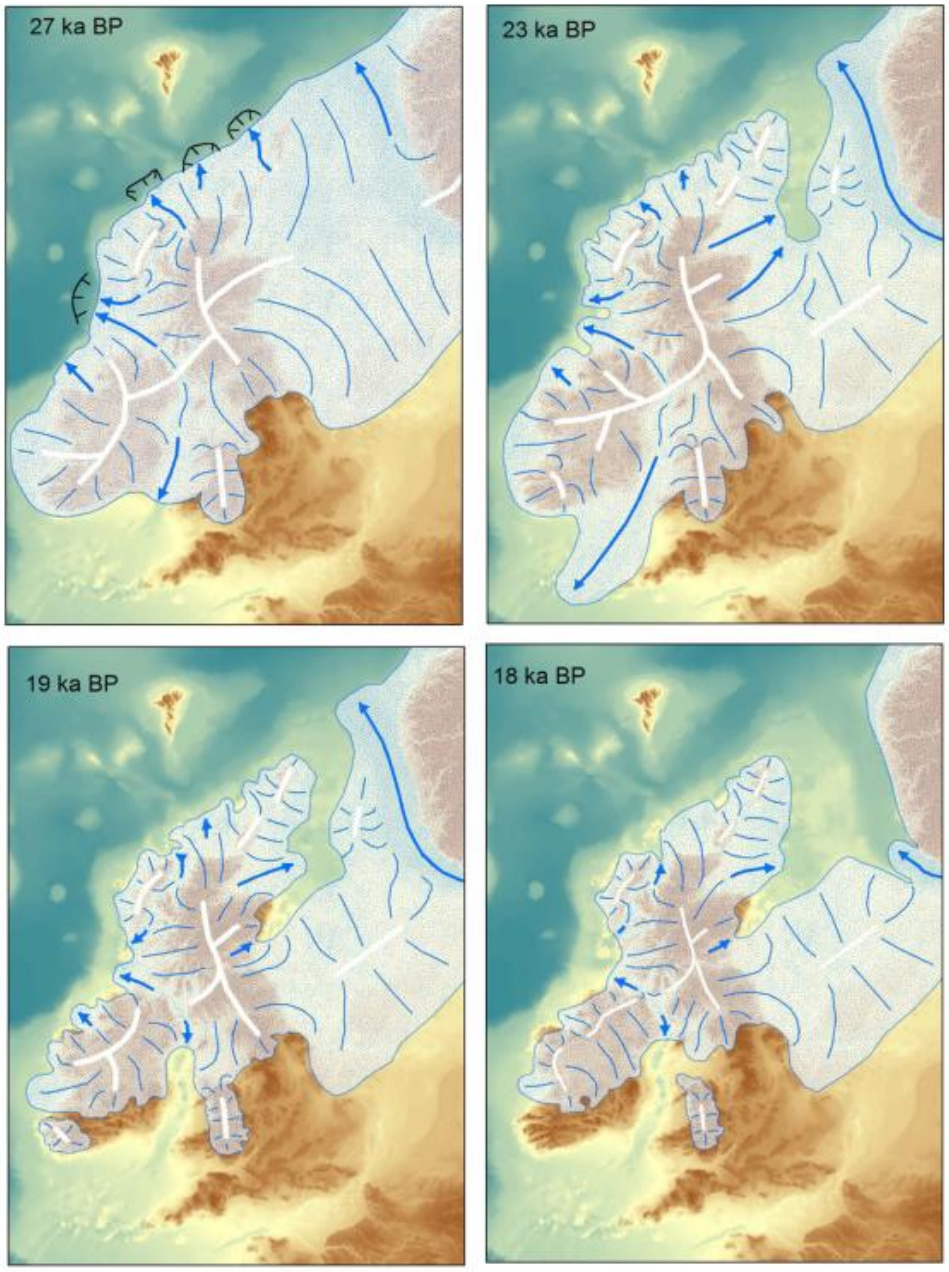

FIGURE 1.2: Ice cover Britain. Location of ice sheets in the British Isles at different time periods (adapted from Clark et al. 2012).

cal. BP (Dark, 2000). Other Early Mesolithic sites in Britain are: Thatcham in Berkshire, occupied between $10,365 \pm 170$ cal. BP and $9,480 \pm 160$ cal. BP; Oakhanger in Hampshire, occupied around $9,045 \pm 66$ cal. BP (Mithen, 1999); and a site in Northamptonshire, which was dated to $10,320 \pm 150$ cal. BP (OxA-803) based on a worked reindeer antler object (Cook et al., 1994).
During the Mesolithic, around 8,500 BP, the British peninsula became an island (White et al., 2000). The melting ice sheets caused extreme changes in the northern landscapes of Britain, and increasing temperatures supported a

rapid change in fauna and flora of the emerging island. The amount of charcoal finds increases at this period, as hunter-gatherer populations started to modify their surrounding landscapes by intensifying wild plant husbandry, forest clearances and vegetation disturbance, which was mainly accomplished by burning and ring-barking (Brown, 1997). The hunter-gatherer lifestyle continued until the onset of the Neolithic in the British Isles, ∼6,200 years ago.
Similarly to the Palaeolithic, human bones are extremely rare finds in this period. However, recently, the genome of an individual found in Gough's Cave, Somerset, named "Cheddar Man", was analysed, along with five other Mesolithic individuals from Britain (Brace et al., 2018). Interestingly, all of these people had genetic similarity to western hunter-gatherers (WHG) in Europe, such as from Loschbour in Luxembourg (Brace et al., 2018), and Cheddar Man himself was predicted to have had dark hair and a dark skin colour.

1.6.3 Neolithic

The first domestication of cereal crops and sheep, goats and pigs took place around ∼9,000 BC (11,000 BP) in the Near East (Ammerman et al., 1984). Along with this Neolithic transition, several cultures arose between 8,000-4,000 BC as people moved across Europe, spreading agriculture and introducing new patterns of settlement (Haak et al., 2010). The spread of the Neolithic was either based on the migration of people or on the adoption and spread of ideas. The traditional model, put forward by Cavalli-Sforza and his colleagues in the 1980s, suggested a large influx of people from the Near East at the start of the Neolithic and a rolling wave of advance of Near Eastern farmers across Europe, with continuous admixture at the wave-front (Ammerman et al., 1984). By contrast, mitochondrial studies in the 1990s proposed a smaller-scale pioneer colonisation for the earliest farmers in Europe and emphasised an important role for the later, gradual assimilation of indigenous hunter-gatherer lineages in the ancestry of European populations (Richards et al., 2000), whereas more recent ancient DNA studies have uncovered a discontinuity of mtDNA (Haak et al., 2005; Haak et al., 2010; Brandt et al., 2013; Deguilloux et al., 2012) and Y-chromosome (Lacan et al., 2011a; Lacan et al., 2011b; Lacan et al., 2013) lineages between Mesolithic hunter-gatherers and Neolithic farmers. The most common Y-chromosome haplogroup in western Europe is R1b, which was initially dated to have originated in the Paleolithic (Semino et al., 2000) in Iberia, but has since been

shown using aDNA to have arisen in the east, probably Ukraine/Russia, and expanded to the west into Iberia (Brandt et al., 2015).
Genome-wide studies have found three main components of autosomal genetic ancestry in Europeans. Firstly, what can be termed a "western hunter-gatherer" component, seen in the present day Saami population (who lack Neolithic ancestry), but at very low level in modern Near Eastern people (Pereira et al., 2017). A second "Neolithic" component, common today across the Near East and Mediterranean Europe, was carried by Early Neolithic individuals in these regions, as well as in Iberia. The highest levels of this Neolithic ancestry are found in modern Sardinians (>50%) (Keller et al., 2012; Sikora et al., 2014), Neolithic Anatolians ($\sim$90%) and Neolithic and Epipalaeolithic Levantine populations ($\sim$100%) (Lazaridis et al., 2016). These observations suggest the arrival of pioneer-colonists with a new genetic make-up into Europe during the Early Neolithic. The third ancestry component, the "Steppe" or "Caucasus hunter-gatherer" component, spread from the east into Europe during the Late Neolithic/Bronze Age, and originated in the Mesolithic of the Iran/Caucasus region (Jones et al., 2015; Lazaridis et al., 2016). These population movements and admixture considerably reshaped the gene pool of western Eurasia within the past 5 kya (Allentoft et al., 2015; Haak et al., 2015).
The location of the British Isles at the western outskirts of Europe delayed the arrival of the Neolithic transition compared to most other European countries, and the environmental conditions were generally more demanding for the practice of farming (Ammerman et al., 1984). The transition from hunting and gathering to farming had a strong impact on the transformation of the landscape, due to deforestation, land clearances for agricultural use, and burning (Woodbridge et al., 2014), and was extremely disruptive (Rowley-Conwy, 2004). Thus, the earliest agricultural activity in Britain is thought to date back to *c.* 4,000 BC or slightly earlier (Woodbridge et al., 2014). Conflicting scenarios for the spread of the Neolithic in Britain have been proposed. The first appearance of Neolithic activities around 4,100 cal. BC spread from the south-east to the north-west as a gradual and regional phenomenon, until it also became established in Ireland and Scotland by around 3,800 cal. BC (Whittle et al., 2011). Alternatively, it has also been postulated that the introduction of the Neolithic may have arisen through several migrations, deriving from different regions across the north of France (Sheridan, 2010).
Some of the emerging Neolithic cultures created distinct forms of pottery,

where the pottery styles found in Britain differ from the ones found in Germany and Denmark (Ammerman et al., 1984). Furthermore, distinct geographic traditions existed across the British Isles alongside a "standard" package of codes and practices as associated with the Neolithic package (Harding, 1997). Significant contrasts between North Yorkshire and lowland England are visible in geographic differences and monuments, as well as differences between Early and Late Neolithic enclosures (Harding, 1997). Neolithic settlements in Scotland had their own local pottery variations (Cowie, 1993).
An innate character of Neolithic societies was the large aggregation of people per site and the gradual development of individual burials and causewayed enclosures, which set the basis for a long-term continuity of social histories within different regions (Harding, 1997). The spread of Neolithic activities in Britain differs from the European mainland, and the genetic admixture during the colonisation phase differs from that seen in other European regions at the same time (Brace et al., 2018). Whereas in mainland Europe, where the colonisation by farmers from the east took place much earlier, slow admixture processes occurred in local populations (Brace et al., 2018). By contrast, there is overwhelming support for agriculture being introduced into Britain by incoming continental farmers, with only small and geographically structured levels of introgression from the indigenous hunter-gatherer populations. Across Britain, the evidence for admixture between hunter-gatherers and farmers is limited, with early British Neolithic farmers deriving two-third to three-quarters of their ancestry from Anatolian farmers, with similar ancestry proportions as found in the Early Neolithic in Iberian and Middle Neolithic central European individuals (Brace et al., 2018). Neolithic individuals from Wales exhibited the lowest levels of western hunter-gatherer ancestry, whilst south-east England and Scotland retained the highest western hunter-gatherer ancestral proportions (Brace et al., 2018). The substantial descent of British farmers from populations related to the Iberian Neolithic indicates that immigration of farmers from the European continent was the major factor for the appearance of the Neolithic in Britain and that the immigration was primarily from Iberia (Brace et al., 2018; Olalde et al., 2018).
Our understanding of ancient people increases in the Neolithic. This is based on the increasing number of burial sites and the growing and spreading populations at this time. The large number of archaeological finds from this period are due to the adoption of a more sedentary lifestyle and the development of new settlements and social cultures, including burial rites, which increases the possibility to study local relationships and status of people.

1.6.4 Bronze Age

The Bronze Age in the British Isles (2,500-800 BC) is identified first by the use of copper, and then bronze as a technological innovation, and both were associated with societal changes. The first evidence of copper metallurgy in the British Isles is from 2,400 BC in Ireland, with other mines established around 1,900 BC in parts of northern England and Wales (Bradley, 2007). The Bronze Age is a crucial period in the development of the British Isles, due to the changes in settlement, burial rites, pottery styles and building of monuments (Parker Pearson, 1999). Settlements enclosed by ditches and palisades appeared, as well as wheeled vehicles and evidence of thriving sea-borne trade (Champion, 1999). At the beginning of the period, people began to be buried individually but, during the Middle Bronze Age, cremation became universal and grave goods were largely limited to Deverel-Rimbury pots that contained the ashes. Pottery is the key artefact of this period as, along with burial rites, it was used in daily routine and food preparation, as well as a means of social identity (Champion, 1999). In the later Bronze Age (from $\sim$1,400 BC), people started dividing the landscape into a pattern of territories, containing access to valley land, upland and access to open moor (Champion, 1999). By $\sim$1,000 BC many of these territories had been partially abandoned, possibly due to climatic deterioration or human over-exploitation (or both) (Champion, 1999). By the end of the Bronze Age, both Ireland and Britain sustained contacts with continental Europe (Mallory, 2013). The culture with possibly the biggest impact on western Europe at this time was the Bell Beaker culture that originated in Iberia and spread into central and north-western Europe between 2,750 and 2,500 BC (Kristiansen, 2014; Brandt et al., 2015; Olalde et al., 2018). With the rise of the Bell Beakers, a network of trade and migration was installed across Europe, circulating metal goods within Ireland and Britain and beyond, and also leading to immigration from northern Britain and the Atlantic fringe to Ireland, and *vice versa* (Mallory, 2013).

Strontium and oxygen isotope analysis of British Bell Beaker remains have shown that these people must have travelled some fair distance in their lifetime. One such example is the Amesbury Archer, who was buried near Stonehenge around 2,300 BC. He is associated with the earliest Bell Beaker phase in Britain and was accompanied by a rich amount of grave goods. According to stable isotopic analysis, he may have been a long-distance traveller from the Rhine (Pearson et al., 2016). However, long-distance travel, such as this individual may have undertaken, is considered an exception. Although

analysis of >250 Bell Beaker individuals from across Britain indicated a considerable degree of mobility between their childhood and death, mostly only moved around within Britain (Pearson et al., 2016).
A recent large-scale whole-genome analysis of 400 European individuals from the Neolithic, Copper Age and Bronze Age, of which 226 were associated with the Bell Beaker culture, documents a phenomenon of almost complete population replacement in Britain with the arrival and spread of the Bell Beakers (Olalde et al., 2018). About 90% of the British gene pool was replaced within a few hundred years. British Neolithic individuals were clearly distinct from the British Beaker-associated people, as the latter brought large amounts of Steppe-ancestry into the country (Olalde et al., 2018). This study showed that the Steppe-ancestry arrived earlier into Britain (∼2,500 BC) than into Ireland (∼2,000 BC) (Cassidy et al., 2016), and that this genetic component is still predominant in the modern British population (Olalde et al., 2018). Beaker individuals from southern Britain showed closest genetic affinity to Bell Beaker individuals from the Netherlands (Olalde et al., 2018). Their level of shared genetic drift, and identical levels of Steppe-related ancestry, suggest that they likely share the same ancestral gene pool (Olalde et al., 2018). Y-chromosome haplogroup R1b represents over 90% of the sequenced British Beaker and Bronze Age men, while, so far, no British Neolithic individual belonging to this Y-chromosome haplogroup has been found (Olalde et al., 2018). The maternal lineage compositions of the indigenous Neolithic people and that of the arriving Beaker-associated individuals suggest that this movement involved both men and women, because the mtDNA haplogroups were introduced at the same time as Y-chromosome haplogroup R1b (Olalde et al., 2018), although the mtDNAs so far identified as new arrivals are a small minority, so it appears that there was a sex bias in the movement of people into Britain.

1.6.5 Iron Age

Iron-working began in Britain ∼800/700 BC. Pastoral farming was expanded and hillforts, such as Cadbury Castle (Somerset), Croft Ambrey (Herefordshire), Danebury (Hampshire) and Maiden Castle (Dorset) (Haselgrove, 1999), were increasingly used. After ∼350 BC, the majority of the hillforts were abandoned, while a smaller number were massively elaborated and heavily protected (Haselgrove, 1999). Cult centres developed and wares were exchanged over wide distances, which brought the British Isles into contact

with the Roman Empire. The increase in the number of settlements at this time has been interpreted as a sign for rising population numbers (Haselgrove, 1999), and the trading network throughout Europe was in constant use, bringing even exotic presents, such as the Barbary ape, into Ireland (Newman, 1998). The large town-like settlements of the Iron Age allowed people to change their social networks to smaller, more flexible ones compared to the strong bonds of country life (Pryor, 2004). The lack of burials, due to cremation and excarnation of the dead, has had a strong impact on the survival of Iron Age material culture.
The Arras culture, found in East Yorkshire but uncommon elsewhere in Britain (Hill, 1995), was loosely associated with the Parisi tribe of pre-Roman Britain (Dent, 1983), and similar to the La Téne culture found in continental Europe. They were defined by the peculiar burials of local elite in disassembled chariots and square enclosures, accompanied by high quality metalwork, which is often unique in Iron Age Britain, as well as the use of imported materials in the grave goods. The skeletons were placed in a north-south orientation, with the head facing north, and pig and horse bones were frequently buried alongside. Genetic analysis of Late Iron Age individuals from southern England has shown that these individuals shared ancestors with multiple northern European populations, such as England, Denmark, Finland and the Netherlands (Schiffels et al., 2016).

1.6.6 Roman conquest

The main landscape change during the Roman period in Britain was the building of roads and towns, which still influence the landscape today, and timber for building was replaced by stone during the Iron Age/Roman overlap (Pryor, 2004). Southern Britain was already in regular contact with the Roman Empire during the later Iron Age (Mallory, 2013), although Roman contact with Scotland, which had been largely unconquered, and Ireland, was less established. Even more so than during the Bronze and Iron Ages, people moved around during this time , as being incorporated into the Roman Empire, ranging from Europe, over North Africa to the Near East, resulted in movement, both voluntary and forced (Leach et al., 2010). Trade was also a significant part of daily life, with Britain providing slaves and raw-materials to the Empire in exchange for high-status commodities, such as wine (Pryor, 2004). The population structure was quite diverse, with a strong Roman impact in south-east Britain compared to the rest of the British

Isles (Pryor, 2004). The large proportion of Roman soldiers of Germanic origin by the late 4th century AD, and the hire of free mercenaries to support the Roman army during the 5th century AD, were reflected in the diversity of the British population during this period (Pryor, 2004). Not only a range of European people, but also individuals of varying status from further abroad, made it to Britain (Leach et al., 2010), leading to the possiblity of further admixture.

The Roman influence on Scotland was limited to the south, which was later manifested through the building of Hadrian's Wall and the Antonine Wall (Breeze et al., 1976). These restricted movements between the north and south of Britain has influenced the diversity and admixture of modern-day populations. Romano-British genomes from a cemetery in York showed a significant divergence from modern Yorkshire and eastern English individuals, whilst they showed particular affinity to the modern Welsh population (Martiniano et al., 2016). When compared to earlier Iron Age samples from East Yorkshire, there appeared to be continuity in the area. However, the Roman genomes differed from Anglo-Saxon genomes from the same region (Martiniano et al., 2016), suggesting a population continuity before the Anglo-Saxon migrations into Britain.

1.6.7 Anglo-Saxon Britain

The Romano-British and Anglo-Saxon periods in Britain were very different, separated by ethnicity, culture and religion. Despite this, in some places there was little or no break in the field system and settlement patterns, and Romano-British fields were subsequently used by the incoming Anglo-Saxons (Pryor, 2004). The withdrawal of the Roman legions from the British Isles led to a decrease in population (Pryor, 2004), and the main question concerning the Anglo-Saxon arrivals in Britain is: what happened to the indigenous people?

In order to try and answer this question of population replacement or admixture, several studies used modern Y-chromosome DNA data, and between 25 and 100% contribution to the modern male English gene pool was estimated (Weale et al., 2002; Capelli et al., 2003; Thomas et al., 2006). The overlay of subsequent migrations on any original Anglo-Saxon introgression renders this approach problematic, especially when incursions from Danish populations are considered, as their male gene pool is basically identical to the Anglo-Saxon one (Capelli et al., 2003). This is due to both low resolution of the genetic markers and poorly defined and overlapping source regions.

Future high resolution studies will likely be able to show differences, and a whole genome approach would greatly benefit any conclusions.
In 2002, Weale et al. suggested, based on an east-west transect of modern-day Britain, that an Anglo-Saxon mass migration, with an associated Y-chromosome gene flow, occurred into central England, but not into Wales.The first detailed dataset of modern British Y-chromosome variation, using low-resolution microsatellites and a small number of haplogroup-defining SNPs, was built up by Capelli et al. in 2003. They saw a trend for Wales, Scotland and Ireland to resemble more closely the Basques, and for English regions to be more similar to the European continent, especially to the likely Anglo-Saxon source populations. This finding was more recently confirmed at the genome-wide level by Leslie et al. (2015), and has since further been supported by the analysis of ancient genomes from Yorkshire, which showed that Romano-British and Anglo-Saxon genomes from the same region differ significantly (Martiniano et al., 2016).
In order to explain the higher reproductive success of Anglo-Saxon immigrants, differential status and reproductive isolation among the different ethnic groups have been suggested. A 7th century AD Wessex law code clearly distinguished between Saxons and native Britons, giving the former a higher legal status (Thomas et al., 2006). Limited intermarriage and differences in reproductive success could also be an explanation for the higher contribution to the modern English gene pool of an initially small Anglo-Saxon immigrant population (Thomas et al., 2006). However, analysis of Anglo-Saxon individuals from the early (fifth to sixth century AD) and middle (seventh to ninth century AD) Anglo-Saxon periods at Oakington in Cambridgeshire, England, produced no evidence for segregation of indigenous people and Anglo-Saxon incomers, but rather genetic admixture and identical cultural community (Schiffels et al., 2016). This was reflected in the archaeological record, as genetically indigenous, admixed and foreign individuals were all buried in similar graves, alongside similar grave goods. One of the genetically foreign Anglo-Saxon individuals was the only burial without any grave goods at this site, suggesting that new immigrants may have been less wealthy than locals (Schiffels et al., 2016).

1.6.8 Vikings in Britain

In $\sim$800 AD, Scandinavian raiding parties started to hit vulnerable coastal areas of the British Isles, mainly monasteries (Richards, 1999). These raids started as hit-and-run affairs, but eventually the Vikings settled. The north

of the British Isles (Scotland, Scottish Isles and the Isle of Man) was colonised by the Norse, whilst Anglo-Saxon England was invaded by Danes, with the exception of the extreme south of England. It is unknown when and how the Vikings started to occupy, conquer and colonise the islands (Ó Corráin, 1998), but it is clear that raiding and exploiting areas in the British Isles created destabilisation that preceded the permanent Viking settlements. The establishment of a slave trade among the settlements led to enhanced possibilities for further admixture (Pearson, 2006). For the Anglo-Saxons especially, invasions led to major economic and social change, and the formation of a network of fortified towns, which later developed into medieval towns (Richards, 1999). As these fortified towns were under royal protection and control, they became the centres of trade, industry, minting and taxation.
The settlements of the islands of Orkney, Shetland and the Isle of Man are special cases, since the cultural connections in these areas are nowadays strongest to Scandinavia. The analysis of Y-chromosomal DNA in modern Orkney populations has revealed closer relationships with Norway than with the rest of the British Isles (Wilson et al., 2001). Similar frequency patterns have been observed in other Norse colonies of Iceland and the Faroes (Helgason et al., 2001; Goodacre et al., 2005). Further analyses of both uniparental markers in Scotland have inferred a considerable Norse component (between 15% and 30%) of the maternal lineage, and suggested family-based settlements with roughly equal numbers of Scandinavian female and male migrants (Goodacre et al., 2005). The Earldom of Orkney was the centre of diverse Norse activities in the North Atlantic region and the intensity of these activities is consistent with the different estimates of Scandinavian ancestry for mtDNA lineages in Orkney, the Western Isles, the Isle of Skye and the north-west coast of Scotland (Helgason et al., 2001).

1.6.9 Ireland

Sea levels fell during the Ice Age, and, prior to 16 kya BP, Ireland became accessible from Britain (Figure 1.3) (Clark et al., 2012; Mallory, 2013), making it possible for hunter-gatherers to follow the ice sheets and reach Ireland on foot. It is argued that southern Ireland was even connected to northern France (Woodman, 1986), although this is not very likely. With the discovery of a human-modified brown bear patella in Alice and Gwendoline Cave in Ireland, colonisation of Ireland during the Palaeolithic is now argued to be evident (Dowd et al., 2016), although this patella might have originated elsewhere and been brought to Ireland at a later date.

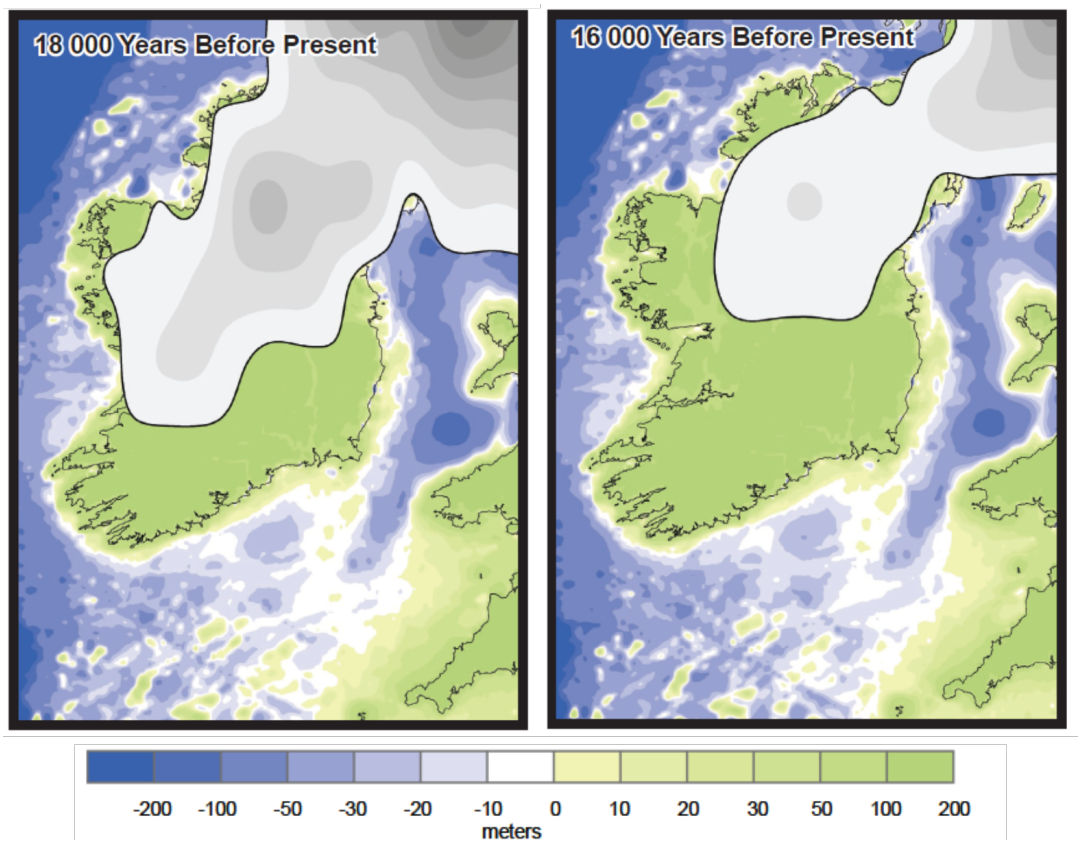

FIGURE 1.3: Map showing the post-LGM landbridges between Britain and Ireland with ice sheets retracting (adapted from Edwards et al. (2008)).

Around 20 kya, ice sheets started retreating and parts of north-eastern and south-western Ireland once more became ice-free (Clark et al., 2012). By 14,000 BP, Ireland was completely ice-free (Edwards et al., 2008), allowing people into Ireland once again. Sea levels rose again with the end of the last Ice Age (∼12,000 BP), making Ireland as the island it is today, and possibly leading to a loss of archaeological coastal sites and their artefacts (Mallory, 2013).

The oldest Mesolithic settlement in Ireland is Mount Sandel, where the remains of huts, stone tools, and assemblages of animal bones showing cut-marks have been discovered (Mallory, 2013). The local elements occurring in the Mount Sandel assemblage differ from any forms found in Britain. The time required for local forms of artefacts to develop suggests that Mount Sandel was occupied some time after an initial phase of settlement in Ireland (Bayliss et al., 2009), and the situation in Ireland does not resemble that of any other area in Europe. Late Mesolithic hunter-gatherer-fisher settlements, such as at Ferriter's Cove, where domesticated cattle remains were found, suggested the possibility of episodes of early contact between Late Mesolithic Irish populations and Mesolithic and/or Neolithic groups from

the continent (Sheridan, 2010). The ceramic and Mesolithic tomb types found in Ireland seem to have a common origin with those found in Britain, yet they are not quite the same (Sheridan, 2010). Passage tombs in Ireland follow a long sequence of development and persist much longer than in most regions of Britain, where they were possibly re-introduced from Ireland at a later stage (Sheridan, 2010).

Bayesian chronologies of plant macro-remains (Rowley-Conwy, 2004; Mallory, 2013; Whitehouse et al., 2014), suggest that the spread of the Neolithic in Ireland must have been extremely rapid, which has been argued to have been facilitated by the use of waterways (Davison et al., 2006). Since any cultural change seen in the archaeological record may have been due to migrations of people into Ireland, the genetic continuity of Irish populations at this time has been a focus for analysis. A Neolithic individual from Northern Ireland was found to have early farmer ancestry similar to other Neolithic individuals from the Atlantic fringe, whose populations originated in the Near East and spread through Europe during the Neolithic (Cassidy et al., 2016). This genome also showed hunter-gatherer introgression, suggesting admixture of ancestral farmers with Mesolithic hunter-gatherers, although it is unknown where this introgression took place. Bronze Age samples from Ireland, like those from Britain, suggested genetic discontinuity, with significant amounts of introgression of an ancestral component from the Steppe, implying that population movements in the Bronze Age, largely attributed in western Europe to the Bell Beaker culture, reached all the way to the north-western edge of Europe into Ireland (Cassidy et al., 2016). However, the sample number from Ireland to date is extremely low (one Neolithic and three Bronze Age individuals, all from the north-east close to Scotland) and so further studies are needed to support this.

While people in Britain increasingly started using iron around 800 AD, there is no extreme shift from bronze to iron in Ireland and, although small-scale movements between Britain and Ireland are suggested based on the comparability of the earliest iron metallurgy and material culture, no evidence for major migrations or population movements into the island currently exist (Mallory, 2013).

The Vikings started to raid Ireland in the late 8th century AD, followed by the establishment of settlements in the early 9th century AD, starting from Norway, but later coming from their established settlements in Scotland (Ó Corráin, 1998). Initially, raids were only seasonal but, over the years, these gave way to the establishment of winter-camps. While most camps were not

occupied for long, a few developed into permanent bases and larger settlements, where trading and other economic functions developed, such as at Dublin and Waterford (Sheehan, 2008). The Viking settlements in Ireland that engaged in international trade became areas of urbanisation, especially in the coastal areas (Doherty, 1980).

A modern genetic investigation of patrilineal kinship structures in Ireland showed no universal feature of Irish tribal units, such as under Ui Neill (McEvoy et al., 2008), while the study by Moore et al. (2006) suggested a strongly influenced social selection of the Y-chromosome. A much more recent divergence of population structure within the Irish population are the Irish travellers (North et al., 2000; Relethford et al., 2013; Gilbert et al., 2017a). After much debate, modern population analysis showed that the travellers are of Irish origin, rather than Roma descendants as originally believed, and that this divergence was not a sudden event, but a rather gradual population development due to separation over time (North et al., 2000), and the genetic differences between Irish travellers and settled Irish populations has been shown to be due to genetic drift rather than external gene flow (Relethford et al., 2013). This most likely occurred due to the potato famine (1845-1849 AD: Woodham-Smith 1962), which caused people to leave their homes, and many varied groups joined the original population of travellers of craftsmen and artisans (Crawford et al., 1974; North et al., 2000). A fine-structure analysis of the modern Irish population, including the Irish travellers, linked individuals with four generations of ancestry to specific regions in Ireland (Gilbert et al., 2017a). Using a fine-structure approach, the Irish population could be divided in 10 distinct geographically stratified genetic clusters, seven of "Gaelic" Irish ancestry and three of shared Irish-British ancestry. The study revealed a major genetic barrier to the north of Ireland in Ulster and a surprising amount of Norwegian-like ancestry (similar to the north and western coasts of Norway) accounting for $\sim$20% of the gene pool. Furthermore, the study suggested a low, constant level of gene flow across Ireland and Scotland prior to Viking admixture. Genetic clusters specific to the Irish traveller population, which reflect sociolinguistic differences, could also be identified, and a high degree of homozygosity and high frequencies of otherwise rare variants were observed in the fine-scale genetic analysis of the Irish traveller population (Gilbert et al., 2017a). Overall, the Irish population is genetically closest to south-western Scottish populations, which have deep historical links through time (Gilbert et al., 2017a).

1.7 Scope of this thesis

After a review of the current literature, several questions about the processes of population movements into Britain remain. Although population replacements during the Neolithic and Bronze Age have already been suggested (Brace et al., 2018; Olalde et al., 2018), it remains unknown what happened to the replaced indigenous populations. Possible introgression events during the Bronze Age to Iron Age transition are yet to be discovered. Similarly, ancient DNA studies have given glimpses of the genetic composition of the Roman period in Britain and their similarities and differences to Iron Age and Anglo-Saxon populations, but sample numbers are still very small and from few archaeological sites. Analyses of the genetic composition of individuals from the Viking period are also missing from the genetic history of the British Isles. The main objective of this thesis was to gain further insights into the genetic diversity and human migration events that can be learned from ancient British samples across these time periods, ranging from the Neolithic to the present, and from the south to the north of Britain, using whole genome sequencing techniques.

Chapter 2 describes the analysis of eight Neolithic individuals from Strathglebe on the Isle of Skye. One tooth sample turned out to be broken at the root and was thus excluded from sample preparation and analysis. Overall, genome coverage of the samples was very low, with only one sample producing over 10,000 SNPs and thus being suitable for genome-wide analysis. This individual showed most similar ancestry proportions to other Neolithic individuals from Europe. Uniparental markers for the seven Neolithic individuals were analysed and identified as belonging to ancient European mitochondrial lineages.

In Chapter 3, 25 individuals from the Bronze Age cemetery at the Links of Noltland, on Westray, Orkney, and three individuals from an Iron Age cemetery at Knowe of Skea, also on Westray, were analysed for both uniparental markers and autosomal data. Familial relationships among the individuals within the Bronze Age cemetery were investigated, and showed that most individuals were likely first-degree related, especially those buried within a large boat-shaped grave. No connection based on the female line of descent could be observed, suggesting patrilocality. Familial relationships were also identified among the three Iron Age individuals from the Knowe of Skea.

Genome-wide analysis of all samples with high enough coverage showed that both Bronze and Iron Age individuals cluster closest with other European Late Neolithic/Bronze Age individuals and show significantly similar admixture patterns, suggesting a continuous ancestry throughout the Bronze and Iron Ages on Westray.

Chapter 4 describes the analysis of a Bronze Age Bell Beaker individual from Low Hauxley, Northumberland, England. Interestingly, the individual showed a small proportion of ancestry most similar to ancient Levantine Bronze Age and modern North African populations, suggesting possible long-distance migrations during the Bell Beaker period.

Chapter 5 describes the analysis of maternal relationships between an adult female and a foetus and a perinate buried within the blockage of a stairwell to a cave entrance, at the Iron Age site of High Pasture Cave, Skye, Scotland. Due to the very low-coverage genomes retrieved from the bone fragments of the foetus and the perinate, only their mitochondrial genomes could be analysed for a possible maternal relationship between the individuals. The mitochondrial genome coverage was high enough to allow detailed comparisons of the mitochondrial haplotypes, and analysis suggested that the female was most likely to be the mother of the foetus, but not the perinate.

Chapter 6 describes the analysis of a rather unusual burial of an adult male in the Rosemarkie Cave, Black Isle, Scotland in the Pictish period. The murdered individual was buried with several big stones holding him down within the burial, as if to prevent him from returning from the dead. Genetic analysis showed that, despite this unusual treatment, he was likely a local, as he had typical northern European mitochondrial and Y-chromosome haplotypes. A comparison with the adult female from High Pasture Cave (discussed in Chapter 5), and published ancient and modern autosomal data, showed that both Scottish individuals displayed similar ancestry to Late Neolithic/Bronze Age European people. Although this only includes two adults, the results are suggestive of a possible continuous population from the Late Neolithic to the Pictish era in Scotland.

In Chapter 7, 16 individuals from a post-Roman cemetery at Worth Matravers were analysed for their genetic kinship, using both uniparental markers and

autosomal data. Several first- and second-degree kin relationships were identified across the cemetery. The majority clustered within modern-day European variation, whereas one individual also exhibited Near Eastern and/or North African ancestry. This was reflected in his Y-chromosome haplogroup - although he belonged to a typical European mitochondrial haplogroup, he had a Near Eastern/North African Y-chromosome. This individual highlights the diversity during the post-Roman period, and the possible long-distance migration of individuals.

Chapter 8 describes the analysis of an Early Anglo-Saxon cemetery at West Heslerton, North Yorkshire, England, which is located at a site that was already in use during the Bronze and Iron Age. Individuals from the Bronze Age (n = 3), Iron Age (n = 2) and the Anglo-Saxon cemetery (n = 9) were analysed in order to compare their uniparental markers and autosomal data. Although the Iron Age individuals were excluded from autosomal analysis due to low coverage, a population shift could be observed between the Bronze Age and the Anglo-Saxon period.

In Chapter 9, genetic analysis of a mutilated skull was used to accompany archaeological investigations on an Anglo-Saxon individual excavated at Oakridge, Basingstoke, England. As mutilation was typically a form of extreme punishment during the Late Anglo-Saxon era, it was surprising to find that the individual, whose nose had deliberately been cut off, was a young woman. From stable isotopic analysis, she was probably not local to the Hampshire area, but possibly from either northern England or southern Scandinavia. In the overall PCA she clustered closest with Bronze Age individuals from Links of Noltland in Orkney.

Two additional chapters with preliminary results can be found in the appendix (Appendix B and C), which focus on the modern phylogeography of the British Isles and on the dissemination of major European mitochondrial haplogroup H, respectively. Mitochondrial haplogroup frequencies were compared and founder analysis was run. The haplogroup composition among the countries within the British Isles varied, with Wales being rather different to other areas in the UK. Founder analysis of H3 showed that the oldest arrivals date to $\sim$8 and $\sim$7 kya in Britain, for central Europe and Iberia as source populations, respectively. Lineages with the biggest impact during

these migrations were very similar between the two source populations (Appendix B).
Phylogenetic trees of subhaplogroups of H were built and branches dated using ρ. The phylogenetic trees showed the likely origins of the haplogroups analysed and their distribution (Appendix C).

Chapter 2

Neolithic human remains from Strathglebe, Isle of Skye, Scotland

2.1 Introduction

The Neolithic first arrived in Britain *c.* 6,000 years ago, a thousand years after its arrival in adjacent areas of north-western Europe (Gkiasta et al., 2003). The migration patterns and cultural changes associated with the Neolithic transition in Britain have been discussed extensively (e.g. Thomas 1988; Sheridan 2010), but large scale genomic analyses of ancient British populations have added a new layer to the previous archaeological picture of the prehistory of the British Isles. The recent analysis of over 70 Neolithic individuals from across Britain found a significant population change from Mesolithic hunter-gatherers to Neolithic farmers (Brace et al., 2018). Of these samples, 37 were from Scotland and these showed the highest admixture proportions of western hunter-gatherers (WHG) compared to the other regions of the British Isles. Generally, little introgression from British hunter-gatherers could be inferred in the analysed British Neolithic samples, compared to inferred admixture with hunter-gatherer lineages brought in from the continent (Brace et al., 2018). However, this excludes the possibility of local adoption, especially, when comparing the different levels of introgression between Scotland, England and Wales. Large-scale studies, such as the above, add important information to the overall picture of the population structure of the British Isles, but they tend to ignore local, familial relationships and the stories behind the analysed individuals.

During a wider landscape survey, which was centred around the excavation of High Pasture Cave on the Isle of Skye (Scotland) (Figure 2.1), the remains of a degraded cairn were recorded. The assemblage of human remains found in the cairn were believed to be associated with the Neolithic as two leaf-shaped arrowheads were found amongst the remains. As the interred

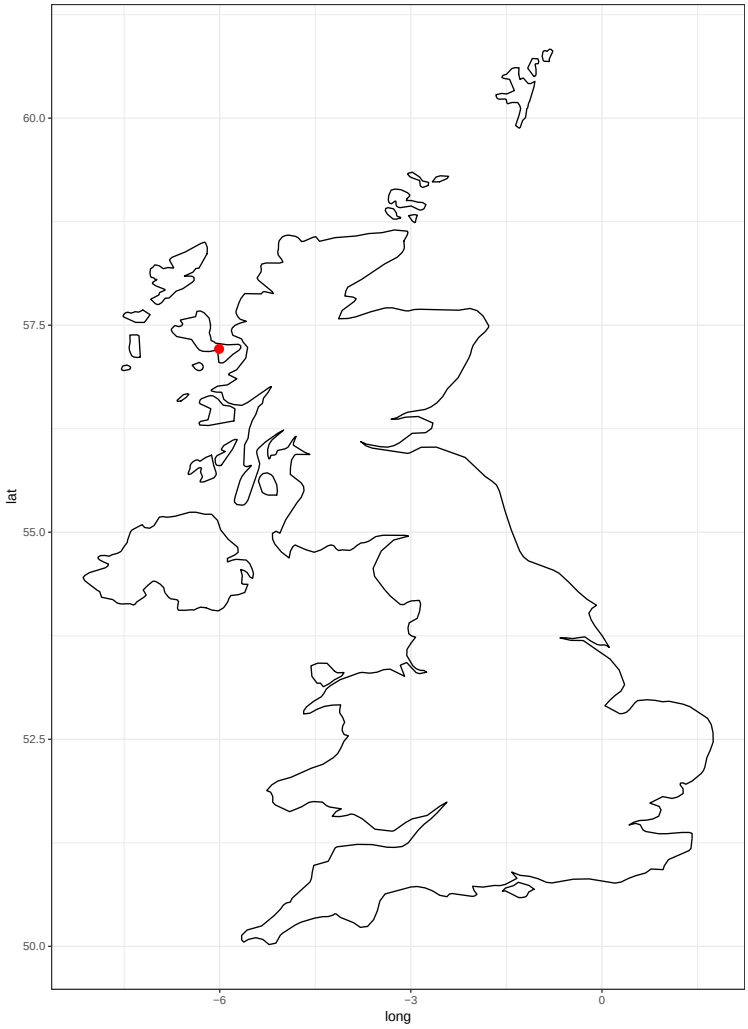

FIGURE 2.1: The location of Strathglebe within the British Isles.

remains were from disarticulated skeletons, it was difficult to determine the number of individuals buried within the cairn. However, based on the number of cranial elements, a minimum of six individuals was estimated. After further analysis of the remains, eight teeth were selected for ancient DNA (aDNA) analysis, stable isotope analysis and radiocarbon dating to confirm the age of the assemblage. The analysis of these individuals should allow for a more localised story of the Neolithic in the Isle of Skye, giving an insight into familial structures and maternal lineages present during this period in Scotland.

2.2 Materials and Methods

The human remains found in the degraded cairn were found in disarticulated bone and teeth mixtures, and a total of eight molars were selected to exclude the possibility of sampling the same individual twice (Table 2.1). The teeth were sampled for genetic and dietary stable isotopic analysis. One tooth (KD030) turned out to be broken at the root and was excluded from DNA analysis, as the amount of retrieved tooth powder was too low and, therefore, this sample was only used for dietary stable isotopic analysis. Tooth powder from KD027 was used for radiocarbon dating.

TABLE 2.1: Processed samples from Strathglebe. Analysis type: M – Macrogen; N/A – excluded genetic from analysis.

DNA code	DNA sample	Analysis type	Era	Sample code
KD023	LRM1	M	Neolithic	SG2 / N2 / Find 03 (sample 1)
KD024	LLM2	M	Neolithic	SG2 / S2 / Find 14 (sample 2)
KD025	LR?M1	M	Neolithic	SG2 / S1 / Find 06 / Context 1 (sample 3)
KD026	UL?M1?	M	Neolithic	SG2 / N3 / Find 50 / Context 3 (sample 4)
KD027	L?L?M1?	M	Neolithic	SG2 / N4 / Find 51 / Feature B (sample 5)
KD028	LR?M1	M	Neolithic	SG2 / S03 / Find 49 (sample 6)
KD029	LLM1	M	Neolithic	SG2 / E02 / Find 43 (sample 7)
KD030	URM3	N/A	Neolithic	SG2 / S2 / Find 48 (sample 8)

2.2.1 Sampling and DNA extraction

The sampling took place at the Ancient DNA Facility at the University of Huddersfield under dedicated clean-room conditions. Full body suits, hairnets, gloves and face masks were worn throughout the drilling, extraction and library preparation processes. All tools and surfaces were constantly cleaned with LookOut® DNA Erase (SIGMA Life Sciences), as well as with bleach, ethanol and long exposures to UV light. The tooth surfaces were decontaminated by UV radiation for 10 minutes on each side, followed by cleaning with 5µm aluminium oxide powder using a compressed air abrasive system. Each root was removed using a cutting saw attached to a hobby drill, and this was then shaken with a steel ball inside a metal shaker in a Mixer Mill (Retsch MM400) for 30 seconds at 30 Hz/second. DNA was extracted from approximately 150mg of the root powder produced, following the protocol by Yang et al. (1998) with modifications by MacHugh et al. (2000). Blank controls were included throughout the sampling procedure and extraction to allow quantification of possible contamination.

2.2.2 Library preparation and sequencing

Next-generation sequencing libraries were constructed from DNA extracts using the method by Meyer et al. (2010) with modifications outlined in Gamba et al. (2014) and Martiniano et al. (2014). The DNA extracts were treated with the USER enzyme (NEBNext) prior to library preparation, by adding 5 µl of USER enzyme (1U/µl), mixing and incubating for 3 hours at 37 °C. All DNA purification steps were performed using the QIAQuick MinElute purification kit (Qiagen) following the manufacturers protocol, with the modification of adding 0.05% Tween 20 (Fisher BioReagents) to the Elution Buffer. Libraries were single-indexed, and blanks were included.
In brief the method was as follows, T4 DNA polymerase buffer (Thermo Scientific) was used instead of Tango buffer in the Blunt-end repair step. Bst polymerase activity was heat inactivated by incubating the libraries for 20 minutes at 80°C. Indexing amplifications were set up using Accuprime *Pfx* Supermix (Life Technology), primer IS4 (0.2µM), a specific indexing primer (0.2µM) and 3 µl of sample library, with a total reaction volume of 25 µl. Amplification was run under the following thermal cycling conditions: 5 minutes at 95 °C; 14 cycles of 15 seconds at 95 °C, 30 seconds at 60 °C and 30 seconds at 68 °C; and a final extension step of 5 minutes at 68 °C. The resulting amplification product was purified using the QIAQuick MinElute purification kit (Qiagen) described above.
For the second amplification round Accuprime *Pfx* Supermix (Life Technology) was used, together with primers IS5 (10µM), IS6 (10µM) and 2.5 µl of sample library. Amplification took place under the following thermal cycling conditions: 30 seconds at 98 °C; 10 cycles of 20 seconds at 98 °C, 30 seconds at 60 °C, 40 seconds at 72 °C, and a final extension of 5 minutes at 72 °C. The PCR product was purified as described above. Amplifications were quantified using a Qubit 3.0 Fluorometer, and then analysed using an Agilent 2100 Bioanalyzer High Sensitivity DNA kit. Libraries were pooled equimolarly, as twelve libraries were sequenced per lane. NGS libraries were sequenced on an Illumina HiSeq4000 (100 base pair, paired-end sequencing; Macrogen, South Korea).

2.2.3 Radiocarbon dating

Radiocarbon dating of sample KD027 was carried out by Dr. Peter Ditchfield at the Oxford Radiocarbon Accelerator Unit. This was processed using the standard Oxford Laboratory protocol (Ramsey et al., 2002; Ramsey

et al., 2004a; Ramsey et al., 2004b). The resultant data was calibrated using OxCal v.4.3.2 (Ramsey et al., 2010), using the IntCal13 calibration curve. IntCal13 represents the mid-latitude northern hemisphere atmospheric reservoir, which consists of compiled measurements of tree rings, plant macrofossils, speleothems, corals, and foraminifera from independent dating (Reimer et al., 2013).

2.2.4 Dietary stable isotope analysis

Stable isotope analysis was undertaken by Dr. Peter Ditchfield at the University of Oxford, according to the following protocol.
Collagen for carbon and nitrogen stable isotopic analysis was prepared from the dentine of the tooth crown. The sample was demineralised using 10ml aliquots of 0.5M HCl solution at 4°C. The acid was changed at 48 hour intervals until no further reaction was seen. The sample was then rinsed three times with milli-µ ultra-pure water and placed in 10ml of pH3 water at 75°C for 48 hours. The sample was filtered using an Eezee™ filter, and the supernatant liquid decanted into a Nalgene™ tube with a temporary Parafilm™ cover. The sample was pre-frozen at -40°C prior to freeze drying in a Zirbus VaCo5 freeze drier fitted with an oil free vacuum system for 72 hours.
The resulting purified collagen was weighed out for analysis using *c*. 1mg aliquots weighed into pre-cleaned tin capsules. Samples were combusted on a Sercon GSL elemental analyser system using a helium carrier gas with a flow of approximately 80ml per minute. A 2% split of the gases evolved was analysed for nitrogen and carbon stable isotopic composition using a Sercon $-20/22$ gas source mass spectrometer. Isotopic values as well as elemental abundances and carbon-to-nitrogen ratios were calibrated against an in-house alanine standard, which itself is routinely measured against international standards (USGS 40 and USGS 41 glutamic acid, whose values are traceable back to the V-PDB and AIR international standards for carbon and nitrogen, respectively). Further aliquots of the alanine standard were used to monitor and correct for instrumental drift. The sample was run in triplicate. Stable isotopic results are reported in delta notation relative to V-PDB for carbon and AIR for nitrogen. Replicate analysis of the alanine in-house standard gave the following results (n= 10 measurements): $\delta^{13}C = -27.15 \pm 0.13‰$, expected value $-27.11‰$; $\delta^{15}N = -1.55 \pm 0.15‰$, expected value $-1.56‰$. This suggests that the individual values obtained during the analyses reported here are typically accurate to be better than $\pm 0.1‰$ for both $\delta^{13}C$ and $\delta^{15}N$.

2.2.5 Data processing and read mapping

The resulting NGS reads were trimmed using AdapterRemoval version 2.1.7 (Schubert et al., 2016), to remove the adapters from the read pairs, trim Ns, remove low quality bases and merge the reads (`-trimns -trimqualities -collapse -adapter1 AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -adapter2 AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT`). BWA aln, version (0.7.12-r1039) (Li et al., 2009a), was used to map reads to the human mitochondrial genome (rCRS, NC_012920.1; Andrews et al. 1999) and the human reference genome (UCSC hg19), filtering by base quality 15 and disabling seed length as recommended for aDNA data (Schubert et al., 2012). Samtools version 1.3 was used to sort, and filter reads, and to remove PCR duplicates. DNA damage patterns were assessed using mapDamage 2.0 (Jónsson et al., 2013) and bamdamage (Malaspinas et al., 2014). Picard tools version 2.18 (`http://broadinstitute.github.io/picard`) was used to add read groups to the sequencing reads aligned to hg19. Variants were called using samtools mpileup with -B -q20 and -Q20 and pileupCaller with the published Lazaridis et al. (2016) and Olalde et al. (2018) datasets as reference panel. The sequencing quality was assessed using FastQC v0.11.5 (Andrews, 2016) and qualimap version 2.2.1 (Okonechnikov et al., 2016). Samtools mpileup and bcftools v.1.3 were used to create a consensus fastq file of the alignment to the rCRS. SEQTK (v.1.2-r102-dirty) was used to convert the fastq file into fasta format. The mitochondrial contamination was estimated using schmutzi (Renaud et al., 2015).

2.2.6 Bioinformatic analysis

The genetic sex was determined using a script published by Skoglund et al. (2013), and in cases where the sex could not be assigned using this method karyotype plots and the Skoglund score (Skoglund et al., 2015) were used. For the mitochondrial haplotype assignment, variant lists of each sample were extracted from the consensus sequences and uploaded to HaploGrep 2 (Weissensteiner et al., 2016). For the paternal lineage prediction Yleaf was used (Ralf et al., 2018).

In order to compare the newly generated ancient data to published modern and ancient human populations, known informative SNP positions of the publicly available Lazaridis et al. (2016) and Olalde et al. (2018) datasets were used and SNPs were filtered as described previously (Lazaridis et al.,

2016; Mathieson et al., 2018; Olalde et al., 2018). Variants in linkage disequilibrium were filtered using the -indep-pairwise option in PLINK v0.67 with the parameters 200, 25 and 0.4. Principal component analysis was run using smartpca of the EIGENSOFT package (Patterson et al., 2006; Price et al., 2006) and plotted with R version 3.4.4 (R Core Team, 2018). A combined dataset of the Neolithic genotypes, together with the reference panel, was run using ADMIXTURE (Alexander et al., 2009) for ancestral population numbers from K=2 to K=14, with cross-validation enabled (-cv flag).

2.3 Results

2.3.1 Sequencing results

All seven libraries of the sampled individuals produced genetic data. The endogenous content of the samples ranged from 5.83 to 26.73%. The duplication rate was generally very low, with <0.1%. The overall genome coverage was very low, as all libraries produced less than 0.1x genome coverage. All results are given in Table 2.2. Contamination estimates of the negative controls, as well as the overall contamination fraction of the sequencing lane were very low (for details see Appendix D).

TABLE 2.2: Sequencing results of the Neolithic individuals from Strathglebe. 'SNPs' denotes the number of SNPs called with the published datasets as reference panel.

Sample ID	Read count	Endogenous content in %	Duplication rate in %	Genome coverage	SNPs
KD023	58,063,490	5.83	0.0733	0.0113	3,598
KD024	79,832,908	6.95	0.0848	0.0157	5,802
KD025	43,700,926	8.15	0.0754	0.0087	1,568
KD026	90,603,286	26.73	0.0910	0.0978	41,748
KD027	80,631,004	7.21	0.0846	0.0193	5,978
KD028	55,547,446	7.65	0.0691	0.0100	2,680
KD029	62,971,140	11.76	0.0728	0.0111	4,515

2.3.2 Radiocarbon dating

The uncalibrated radiocarbon date was 4569 ± 35 bp (OxA-37513), which corresponds to 3495-3104 calibrated BC (95.4%) shown in Figure 2.2.

2.3.3 Dietary stable isotope analysis

In order to identify the type of diet the Strathglebe individuals were eating, dietary stable isotopes were analysed and compared to four other Neolithic sites from western Scotland: Carding Mill Bay, Oban (n = 11); Clachaig A,

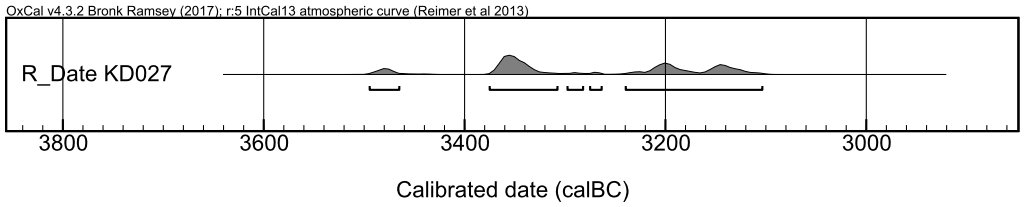

FIGURE 2.2: Calibrated radiocarbon date of KD027.

West coast of Scotland (n = 6); Crarae, Inveraray (n = 3); and Raschoille Cave, Oban (n = 9) (Schulting et al., 2002b; Bownes, 2018). The resulting plot is shown in Figure 2.3.

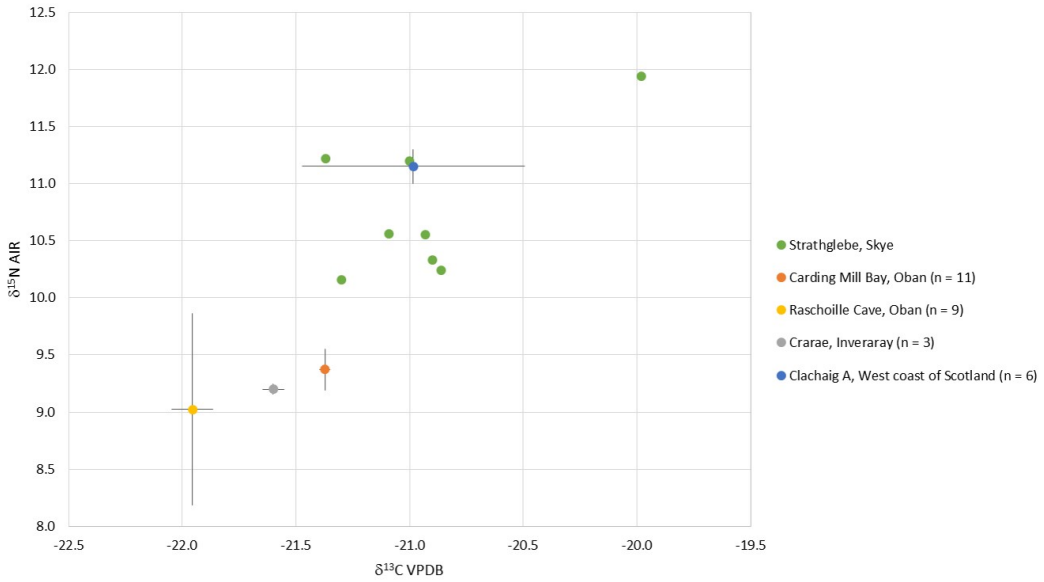

FIGURE 2.3: Dietary stable isotope data of the Neolithic Strathglebe individuals, compared to four other Neolithic human assemblages from western Scotland (Schulting et al., 2002b; Bownes, 2018).

2.3.4 Genetic sex determination

Using the Skoglund script (Skoglund et al., 2013), the genetic sex could only be determined for one individual (KD026), which was a male. For individual KD029, the sex determination was consistent with XX, but not XY. No sex could be assigned to any other individual using this method. For the undetermined individuals, karyotype plotting was used and the percentage of chromosome X reads was calculated following Skoglund et al. (2015) (for details see Appendix E). Following this approach, the genetic sex of all previously undetermined individuals could be estimated (Table 2.3).

TABLE 2.3: Karyotype-based genetic sex determination using the Skoglund score (Skoglund et al., 2015).

Sample ID	X-chromosome in %	Sex determination
KD023	78.6	XX
KD024	82.5	XX
KD025	60.0	XY
KD027	53.2	XY
KD028	73.3	XX
KD029	87.3	XX

2.3.5 Mitochondrial genome analysis

The mitochondrial haplotype analysis revealed that two haplogroups are shared among four individuals, T2b and U5b2c. Haplogroups represented by only one individual within this assemblage are J1c1, H1c and H5b (Table 2.4). Individual haplotypes for all samples analysed can be found in Appendix F.

TABLE 2.4: Mitochondrial genome analysis of the Strathglebe individuals.

Sample ID	Mitogenome coverage	Std deviation	mtDNA Haplogroup	Overall rank
KD023	13.0923	3.6492	J1c1	0.9975
KD024	23.0202	4.7622	T2b	0.9902
KD025	0.8991	1.0684	H1c	0.6067
KD026	22.1220	4.5654	U5b2c	0.9222
KD027	11.9543	3.6918	U5b2c	0.9277
KD028	15.4872	3.8070	T2b	0.9902
KD029	16.6424	4.1540	H5b	1.0000

2.3.6 Y-chromosome analysis

Only one sample (KD026) had enough coverage to be securely identified as male. This sample had four 'most downstream' predictive markers for Y-chromosome haplogroup I, namely M258, CTS4088, CTS4982 and L751, with a total of 1,672 detected SNPs (for details see Appendix G Table G.1). The number of retrieved predictive markers for chromosome Y were too low in the other two putative male samples (KD025 and KD027).

2.3.7 Principal component analysis

Principal component analysis (PCA) was run for KD026 only, due to the number SNPs overlapping with the Lazaridis et al. (2016) dataset. A PCA

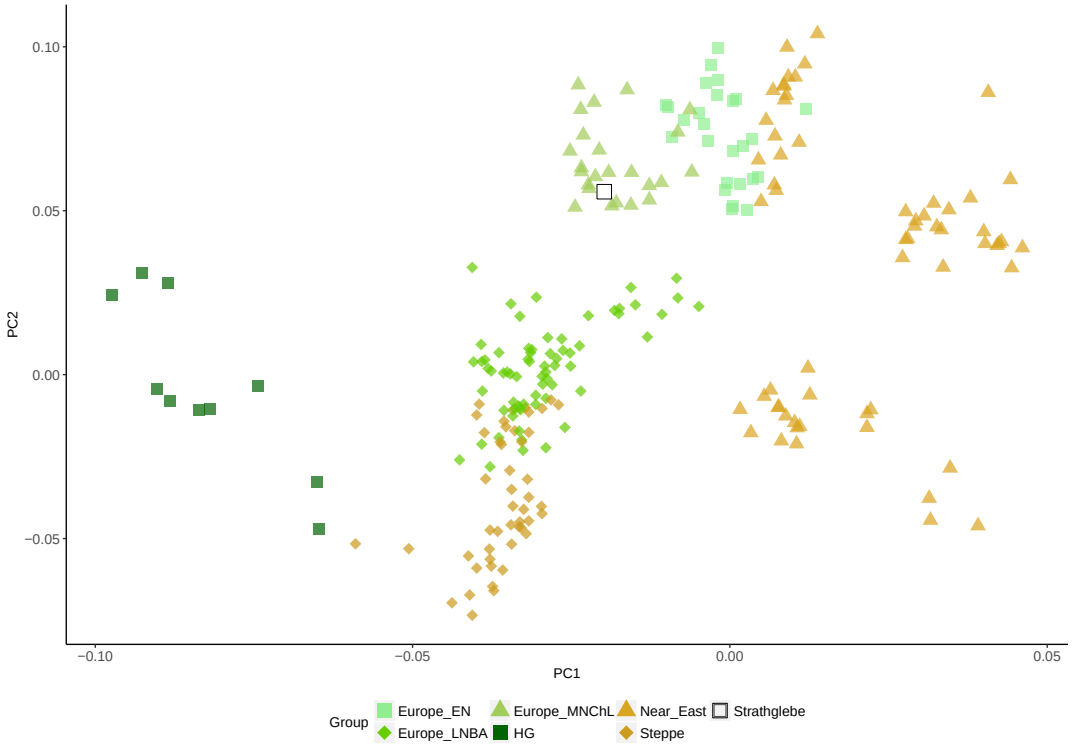

FIGURE 2.4: Principal component analysis of the Strathglebe individual, with previously published ancient samples of Lazaridis et al. (2016) projected onto the first two principal components.

was run on the sample merged with the Lazaridis dataset. The sample is projected on the first and second principal components of the published dataset (Figure 2.4).

2.3.8 ADMIXTURE analysis

ADMIXTURE was run for KD026 merged with the complete Lazaridis et al. (2016) and Olalde et al. (2018) datasets, with cross-validation between *K* values 2 to 14, with *K*=12 giving the lowest standard error (Appendix H) and therefore being the sensible modelling choice (Figure 2.5).

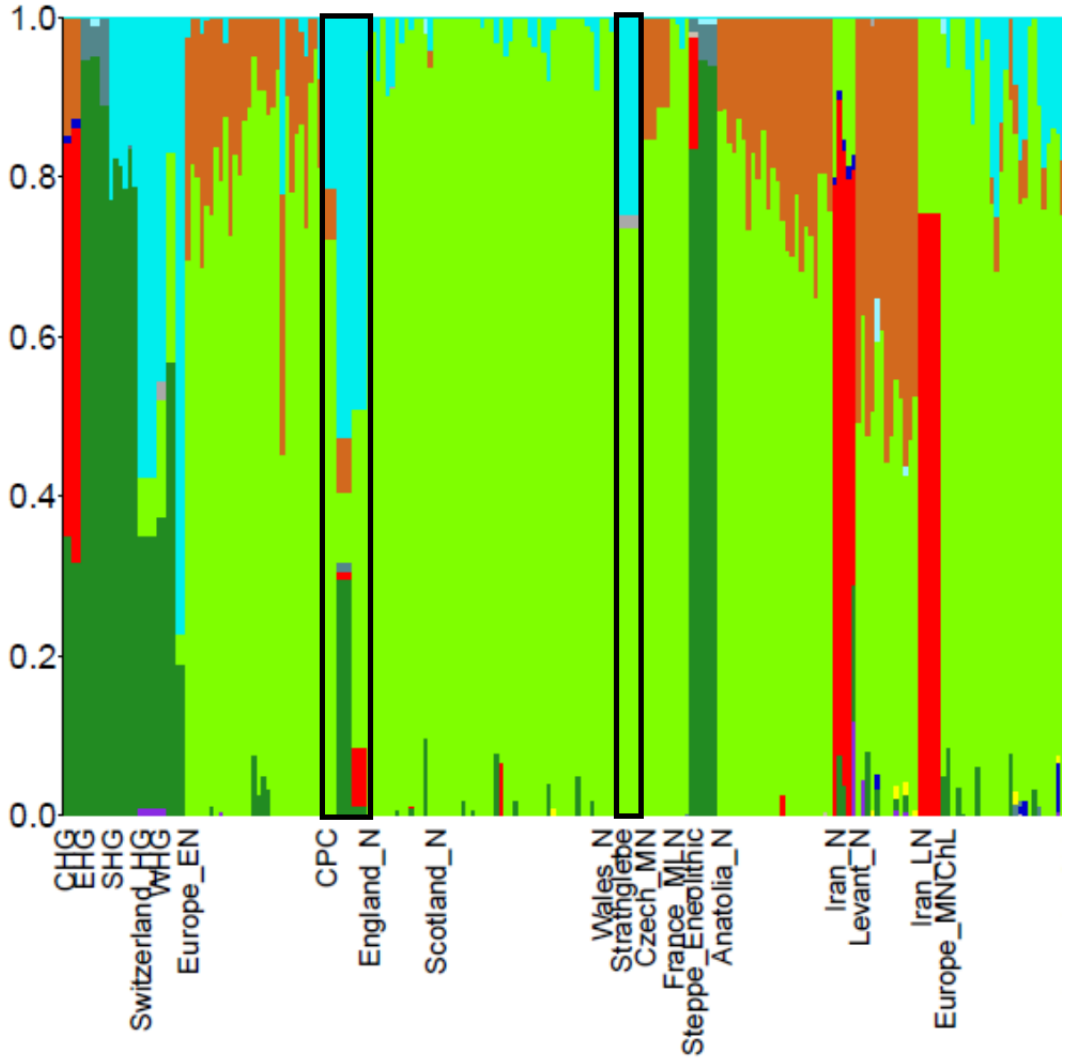

FIGURE 2.5: ADMIXTURE analysis of the Strathglebe individual (KD026), with previously published ancient samples of Lazaridis et al. (2016) and Olalde et al. (2018) for K=12. Figure displays relevant subsection of Figure 10.2. Abbreviations used: E: Early, M: Middle, L: Late, HG: Hunter-Gatherer (C - Caucasus, E - Eastern, S - Scandinavian, W - Western), N: Neolithic, ChL: Chalcolithic, CPC: Carsington Pasture Cave.

2.4 Discussion

2.4.1 Distinguishing between individuals

Since the samples originated from loose teeth, and because it was not possible to sample the same tooth type from each (potential) individual, the genetic data was compared in order to possibly distinguish between individuals. Furthermore, one sample (KD027) was dated in order to verify the archaeological determination of the assemblage belonging to the Neolithic. The radiocarbon date of KD027 of 4569 $\pm$ 35 bp (OxA-37513) confirmed the Neolithic origin of these samples.
The two samples belonging to mtDNA haplogroup U5b2c were both determined as male, and unfortunately, the Y-chromosome coverage of KD026 and KD027 was too low to compare the Y-chromosome haplogroup and check for differences to exclude the possibility of them having originated from the same individual. However, KD027 was from Feature B in a group with several other teeth, whereas KD026 was from a different context. Therefore, they are most likely different individuals. Similarly, KD024 and KD028 were both determined as female and both belonged to mtDNA haplogroup T2b. As they were sampled from a lower left second molar and a lower right(?) first molar, respectively, it could not be excluded in this case that these two samples originated from the same individual.

2.4.2 Neolithic diet in Strathglebe

The dietary stable isotope data showed that the people at Strathglebe were eating a high protein diet, as shown by their high nitrogen values. They are eating a similar diet to the people from Clachaig A on the west coast of Scotland, but both sites are different to Carding Mill Bay and Raschoille Cave in Oban, and Crarae in Inveraray. These latter people have little to no marine component in their diet (Schulting et al., 2002b), whereas Strathglebe and Clachiag A show more evidence of a fish-rich diet. KD025 has the highest nitrogen value, which, along with the increased carbon result, suggests that this individual was eating a marine diet.

2.4.3 Maternal lineages

The genome coverage was too low to run any admixture analysis, so it was not possible to compare the genomes of the studied individuals with previously reported Neolithic individuals from Britain. Nevertheless, the low coverage data did allow for analysis of maternal familial relationships amongst the assemblage. The seven individuals analysed in this study were identified as belonging to five different mtDNA haplogroups; i.e. H1c, H5b, J1c1, T2b and U5b2c, with both T2b and U5b2c seen in two samples each.

2.4.3.1 Haplogroup H

Two individuals belong to H1c and H5b, sublineages of the major European mtDNA haplogroup H (more details on this haplogroup can be found in Appendix C). The sample (KD025) that was determined to belong to mitochondrial haplogroup H1c only had a rank of 0.60, which indicated inaccuracy. After manual checking of the sequence only the defining mutation for H1c could be found. The sample did not show the defining mutation for H1 G3010A, although the postion was covered. Instead, the positions for both defining mutations of H were not covered, therefore, any assumptions made on the presence of H1c in the Neolithic on the Isle of Skye need to be made with caution. H1 is the most frequent subhaplogroup of H and has a similar star-like structure to H. Its frequency peaks are in Iberia and it is argued that H1 spread with the Neolithic migrations from the ice age refuge in Iberia into Europe. In modern-day Britain haplogroup H1 has a frequency of $\sim$30%.
Haplogroup H5, on the other hand, is less frequent than H1, although it is still one of the more frequent subclades of haplogroup H, accounting for $\sim$8% in modern-day Britain (see Appendix B). Its highest frequencies are in south-eastern Europe (mainly Italy), Latvia and Wales. Haplogroup H1c has previously been reported in Mesolithic/Neolithic remains from Sweden (Skoglund et al., 2014), from the Neolithic in England (Cheddar, Somerset) and Scotland (Oban, Argyll and Bute) (Olalde et al., 2018) and in the Bronze Age from England (Stockton-On-Tees, North Yorkshire) (Olalde et al., 2018), Germany (Knipper et al., 2017) and Latvia (Mittnik et al., 2018).
While H1c has previously been found in the British Isles dating as far back as the Neolithic, H5b has, to date, not been reported from ancient human remains from the British Isles. Neolithic remains belonging to H5b were found in Bulgaria (Mathieson et al., 2018) and Germany (Brotherton et al.,

2013). Human remains from the Bronze Age were found in the Czech Republic (Olalde et al., 2018) and Russia (Mathieson et al., 2015), whereas two Iron Age individuals belonging to H5b were found in Moldova (Juras et al., 2017). The newly reported H5b mitochondrial genome from the Isle of Skye is important for the analysis of H5 and its arrival in the British Isles, as its current distribution in the Neolithic suggests a contribution of mitochondrial lineages from central Europe into the British Isles.

2.4.3.2 Haplogroups J and T

Mitochondrial haplogroups J and T both originated in the Near East and spread into Europe after the peak of the LGM (Pala et al., 2012), however, multiple dispersals during the Late LGM and the Neolithic have been suggested as well (Pereira et al., 2017). The main subclades of both haplogroups that are seen as signals for the major expansions in the Late Glacial period are J1c, J1b1a, J2a1, T1a1, T2a1b and T2f1, as well as T2b and T2e, which signal the expansions in the immediate postglacial period (Pala et al., 2012). Haplogroup J1 is in general mainly Near Eastern, whereas its subclade J1c is primarily European. Haplogroup J1c1, which has been identified in an individual from Strathglebe, has an age estimate of $\sim$12 kya (Pala et al., 2012). T2b is the major European clade of haplogroup T and comprises of about half of all T2 seen within Europe. T2b likely dispersed in Europe during the Early Neolithic and has an age estimate of $\sim$10 kya. Similar to the major European haplogroup H (see Appendix C), T2b is quite star-like and very widespread within Europe (Pala et al., 2012). Both J1c and T2b have been reported from Neolithic and Bronze Age remains throughout Europe.
J1c has been identified in Neolithic remains from England (Brace et al., 2018), Hungary (Gamba et al., 2014), Germany, (Haak et al., 2015; Lipson et al., 2017) and Bulgaria (Mathieson et al., 2018). Other remains from Germany (Knipper et al., 2017; Olalde et al., 2018), Hungary (Lipson et al., 2017), Romania and Ukraine (Mathieson et al., 2018), and the British Isles (Olalde et al., 2018) date to the Bronze Age. The subclade J1c1 has been mainly reported in Neolithic and Bronze Age remains from the British Isles and Spain, but was also found in Neolithic remains from Hungary, Greece and Macedonia (Gamba et al., 2014; Mathieson et al., 2015; Hofmanová et al., 2016; Lipson et al., 2017; Brace et al., 2018; Mathieson et al., 2018; Olalde et al., 2018). Haplogroup T2b has, to date, been reported in 41 ancient samples from the Neolithic, Chalcolithic, Bronze Age and Iron Age in Europe. The remains dating

to the Neolithic are from Austria (Mathieson et al., 2018), Bulgaria (Mathieson et al., 2018), England (Winterbourne Monkton, Wiltshire) (Olalde et al., 2018), Germany (Haak et al., 2015; Lipson et al., 2017), Hungary (Lipson et al., 2017), Turkey (Mathieson et al., 2015), and Ukraine (Nikitin et al., 2017; Mathieson et al., 2018). Bronze Age remains were found in Denmark (Allentoft et al., 2015), Hungary (Allentoft et al., 2015; Lipson et al., 2017; Olalde et al., 2018), Czech Republic (Allentoft et al., 2015; Olalde et al., 2018), Turkey (Lazaridis et al., 2017), Ukraine (Mathieson et al., 2018), Lithuania and Sweden (Mittnik et al., 2018), the British Isles (Biddenham Loop, Bedfordshire; Mortlake, London; Trumpington Meadows, Cambridge; Stockton-On-Tees, North Yorkshire), Germany and Spain (Olalde et al., 2018). Furthermore, one Iron Age individual from Moldova (Juras et al., 2017) was found belonging to T2b.

2.4.3.3 Haplogroup U5

The mtDNA haplogroup U5 is the most ancient European maternal haplogroup and is argued to have arisen in Europe (Malyarchuk et al., 2010b). Its most ancient identified subhaplogroup is U5b2, which dates to $\sim$24 kya, suggesting a pre-LGM divergence within Europe (Malyarchuk et al., 2010b). Haplogroup U5 has been found in ancient individuals from the Czech Republic and Belgium (Fu et al., 2016; Posth et al., 2016), all dating to the Upper Palaeolithic. U5b* is mainly reported from the Late Mesolithic and Early Neolithic of the British Isles (Brace et al., 2018; Olalde et al., 2018), but also from the Late Chalcolithic in Hungary (Lipson et al., 2017). To date four ancient individuals belonging to haplogroup U5b2* have been reported. One individual from the Early Neolithic in Hampnett, Cheltenham, England, (Brace et al., 2018) and three individuals from the Ukraine dating to the Mesolithic (Jones et al., 2017; Mathieson et al., 2018). Within the subclade of U5b2c*, nine ancient individuals have been reported to date. These samples were found in the Mesolithic in Serbia, the Late and Middle Neolithic in Hungary, the Chalcolithic/Early Bronze Age in the Czech Republic, the Bell Beaker period in Germany and the Corded Ware in Estonia (Gamba et al., 2014; Allentoft et al., 2015; Lipson et al., 2017; Saag et al., 2017; Mathieson et al., 2018; Olalde et al., 2018).

2.4.4 Paternal lineages

One of the individuals from the Late Neolithic in Hungary that belongs to haplogroup U5b2c has the Y-chromosome haplogroup I2a1, and appears to share a similar ancestry as the male from this study (KD026). Although the coverage of the Y-chromosome was too low to further specify the predicted Y-chromosome haplogroup, the reported individual likely belongs to Y haplogroup I. Virtually absent anywhere else, Y-chromosome haplogroup I is the only major clade that is widespread in Europe (Rootsi et al., 2004). The phylogeography of haplogroup I suggests a European origin, most likely before the Last Glacial Maximum (LGM) (Semino et al., 2000), and thus a pivotal role of this lineage during the recolonisation of Europe after the LGM (Rootsi et al., 2004). The haplogroup is most common among modern day Scandinavian populations and those in the north-western Balkans (Rootsi et al., 2004).

2.4.5 Conclusions

Unfortunately, the coverage of the sequenced genomes was too low to securely determine the genetic sex of the studied individuals. A full Y-chromosome analysis of the single possible male individual was not possible, as a minimum coverage of the Y-chromosome is needed to predict the haplogroup. Thus, only one male individual could be analysed for both its maternal and paternal lineages. A comparative analysis of the presented Neolithic genomes with published genomes from the Neolithic in Britain and the European continent was only possible for one sample (KD026), due to the very low coverage of the genomes. In the PCA KD026 clustered closest with European Neolithic individuals, which was supported by the admixture analysis, that showed similar patterns of ancestry for the Neolithic individual from Strathglebe and European Middle Neolithic individuals from the Lazaridis et al. (2016) dataset, as well as to the Neolithic individual from the Carsington Pasture Cave (supplementary information in Appendix I).

Chapter 3

A Bell Beaker individual from Low Hauxley, Northumberland

3.1 Introduction

Over the past centuries, archaeological and palaeoenvironmental deposits have been lost to the sea due to the impact of coastal erosion (Waddington et al., 2016). The cliff face to the south of Low Hauxley, on the Northumberland coast (Figure 3.1), has been eroding for the past 30 years (Waddington et al., 2016). The main archaeological site occupies a slight natural hillock and is flanked to the north and south by separate organic sediment units, described as "peats", that would have originated as wetlands during the Neolithic period, around the 4th millennium cal. BC (Waddington et al., 2016). A team, led by Clive Bonsall, conducted a long-term investigation of the site between 1982 and 1988, which involved systematic excavation and extensive palaeoenvironmental work (Bonsall, 1984). Evidence for human activity dating from the Mesolithic through to the medieval period was found, including the remains of a Neolithic stone-built burial mound, or cairn, from which pottery sherds and cist burials with cremated human remains were eroding (Waddington et al., 2016). In addition, Mesolithic flint artefacts were found in a buried soil horizon underneath the cairn, and pottery was recovered from a later soil layer within the sand dune above the cairn.

Human remains, including a complete skeleton and a Beaker, were recovered from within the cairn, all from cist burials, made by digging pits through the ground surface and constructing well-made stone cist boxes with capstones that would have stood proud of the ground surface (Waddington et al., 2016). The inhumation (Burial 1) was removed from its cist through a small "side" slab on its north side (Waddington et al., 2016). The individual was a young, possibly male, young adult, aged between 12 and 16, crouched and facing towards the north. He was radiocarbon dated to the early part of

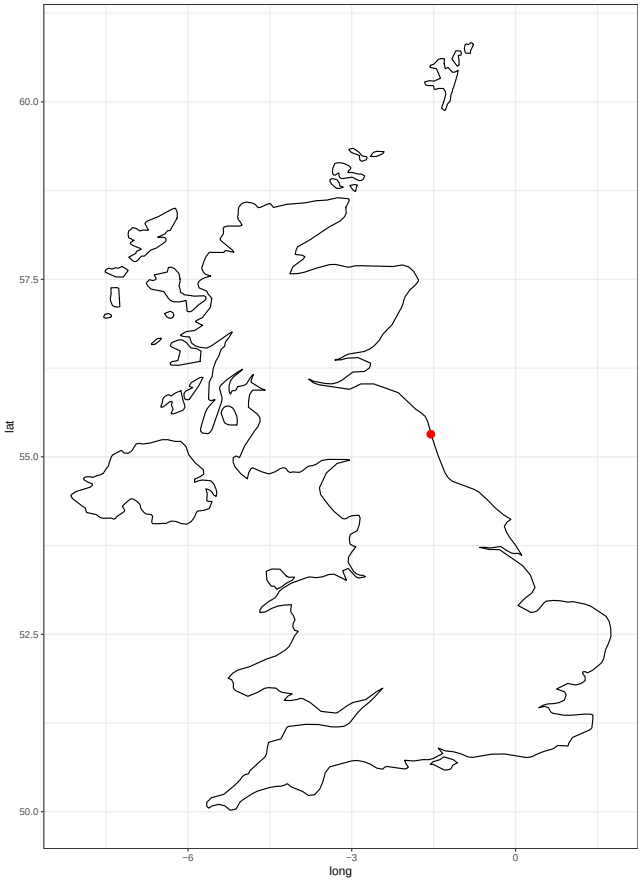

FIGURE 3.1: The location of Low Hauxley within the British Isles.

the Beaker period (3874 ± 32 bp; SUERC-49872; approximately 2466-2279 cal. BC) (Waddington et al., 2016), and appeared to have been accorded particular significance as a primary or founding burial. Two other cists contained cremated remains from two other people (Burials 2 and 3), who were not accompanied by a Beaker vessel. To date, there are no other instances where both cremation and inhumation remains accompanied by a Beaker have been reported at the same site (Waddington et al., 2016). The tops of the three cists would have remained visible after burial and they formed the first phase of the Low Hauxley cemetery, which can be directly associated with the Beaker people (Waddington et al., 2016). The genetic analysis of the Beaker inhumation can contribute to the understanding of the arrival of the Bell Beakers into Britain and their genetic heritage.

3.2 Materials and Methods

The left petrous portion of Burial 1 was sampled for DNA analysis at the Ancient DNA Facility of the University of Huddersfield, under dedicated

clean-room conditions. Two single-indexed libraries were prepared, and sequenced on an Illumina HiSeq 4000 (100 base-pair, paired-end sequencing; Macrogen, South Korea). The analytic steps, sample preparation and bioinformatic analyses followed as outlined in Chapter 2.

3.3 Results

3.3.1 Sequencing results

Sequencing results are displayed in Table 3.1.

TABLE 3.1: Sequencing results of the Low Hauxley sample. 'SNPs' denotes the number of SNPs called with the published datasets as reference panel.

Sample ID	Read count	Endogenous content in %	Duplication rate in %	Genome coverage	SNPs
Low Hauxley	69,129,236 & 53,129,192	16.00 & 19.05	1.39	0.0553x	21,101

3.3.2 Genetic sex determination

The genetic sex was determined as male. The score by Skoglund et al. (2015) gave a result of 'only consistent with XY but not XX' (Skoglund et al., 2013), and, therefore, karyotype-plotting (Figure 3.2), was undertaken, together with the calculation of the Skoglund score (Skoglund et al., 2015), which was 57.4% and, therefore, indicates a male.

3.3.3 Mitochondrial genome analysis

The haplotype of the presented ancient sample was determined as described in Chapter 2, using the rCRS as reference genome. The individual was found to have all the defining mutations of haplogroup T2e1a (Table 3.2). The full mutation list can be found in Appendix F.

3.3.4 Y-chromosome analysis

The Y-chromosome analysis gave R1b1a1a2 as the predicted haplogroup, with CTS7880, CTS5676, L820, L389, and CTS12972, as most downstream predictive markers (see Appendix G for more details).

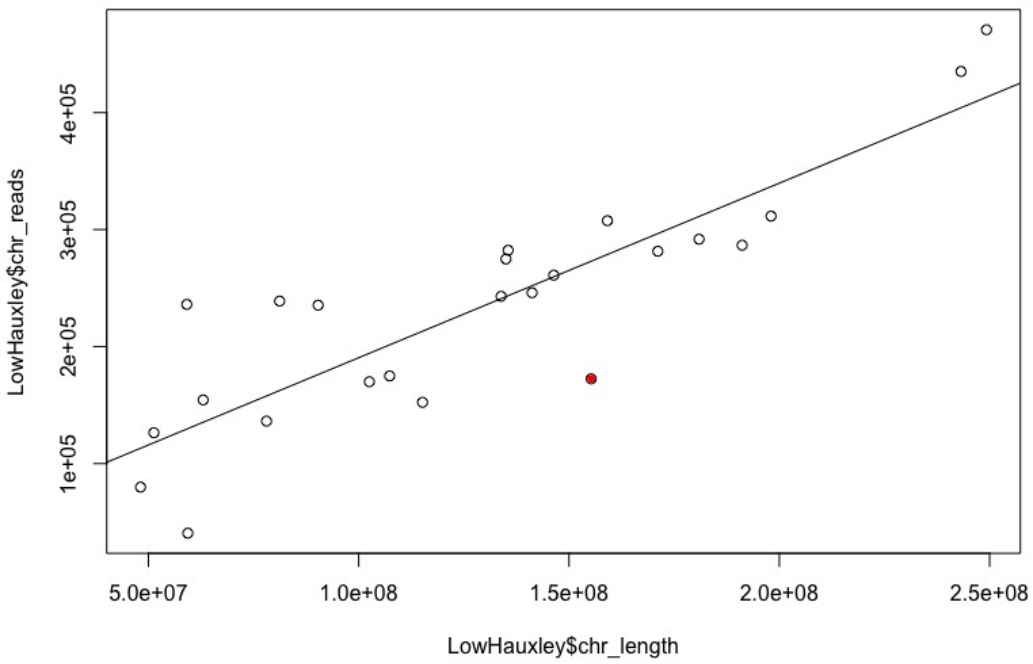

FIGURE 3.2: Karyotype plot of the Low Hauxley individual.

3.3.5 Principal component analysis

In the PCA, the individual of Burial 1 plots alongside published European Late Neolithic/Bronze Age samples used in Lazaridis et al. (2016) (Figure 3.3).

3.3.6 ADMIXTURE analysis

The ADMIXTURE cross-validation error for K=12 was the lowest (see Appendix H). The admixture plot is shown in Figure 3.4.

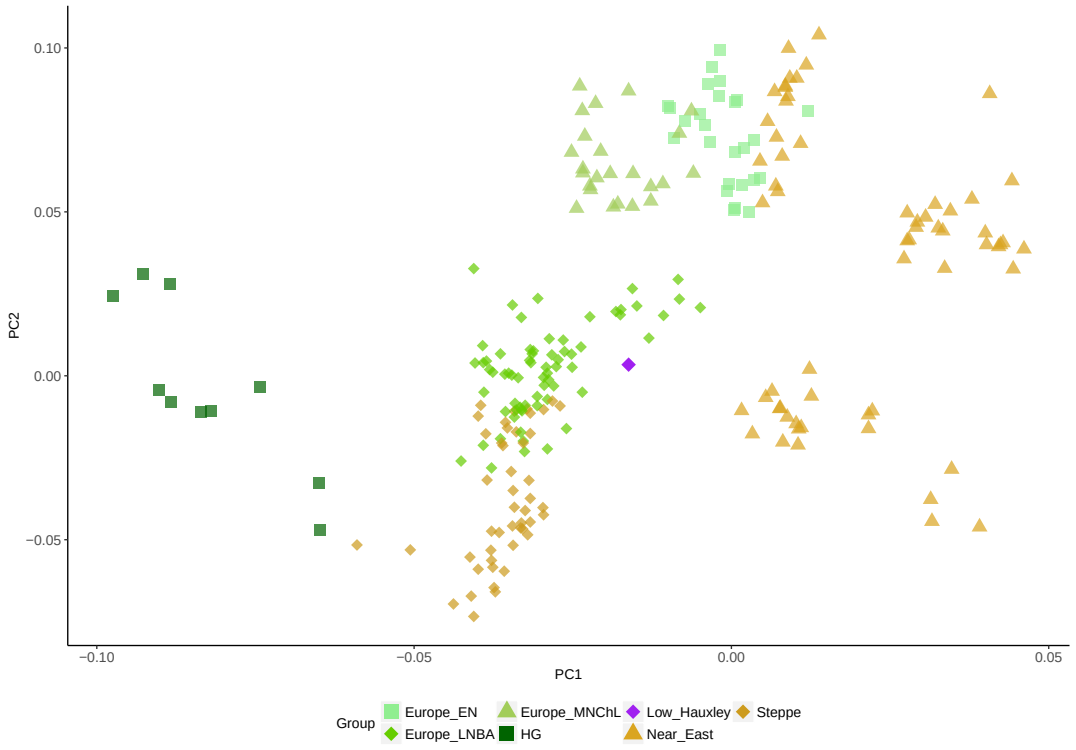

FIGURE 3.3: Principal component analysis of the Low Hauxley individual, with previously published ancient samples of Lazaridis et al. (2016) projected onto the first two principal components.

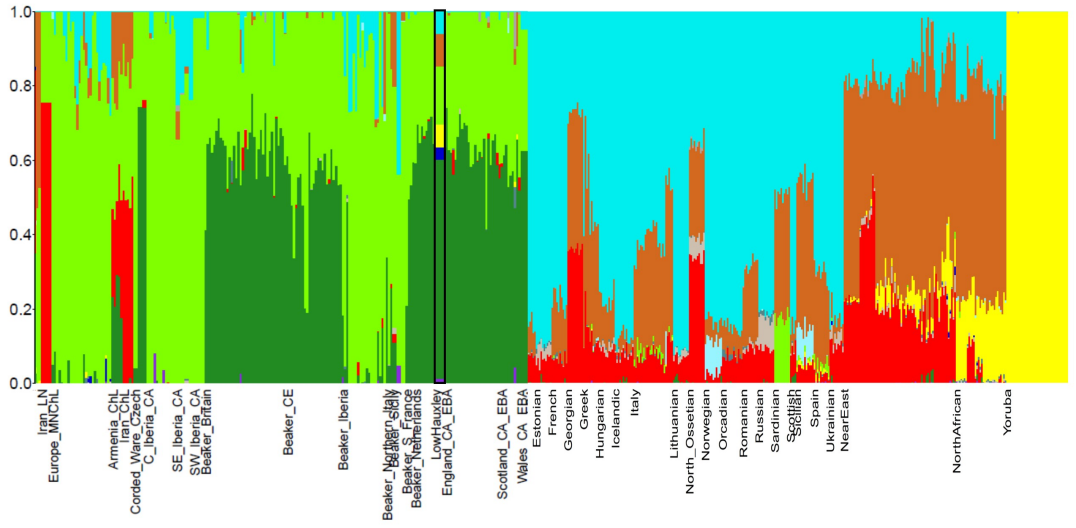

FIGURE 3.4: ADMIXTURE analysis of the Beaker individual from Low Hauxley, with previously published ancient Beaker and Copper Age samples of Lazaridis et al. (2016) and Olalde et al. (2018) for K=12. Abbreviations used: E: Early, M: Middle, L: Late, N: Neolithic, ChL: Chalcolithic, CA: Copper Age, BA: Bronze Age, CE: Central Europe, LON: Links of Noltland.

TABLE 3.2: Mitochondrial DNA haplogroup and mutations identified in the Low Hauxley individual.

Sample	Haplogroup	Average coverage	Polymorphisms
Low Hauxley	T2e1a	8.8633x	41T 73G 150T 263G 750G 1438G 1888A 2706G 4216C 4769G 4917G 7028T 8697A 8860G 10463C 11251G 11719A 11812G 13368A 14233G 14766T 14905A 15326G 15452A 15607G 15928A 16126C 16294T

3.4 Discussion

3.4.1 Genome coverage

The genome coverage of the Beaker individual from Low Hauxley was 0.0553x and allowed genome-wide analysis with published modern and ancient genomes, as well as detailed analysis of the uniparental markers. Although it is difficult to sex sub-adult remains using osteological methods, the individual was classified as male (Waddington et al., 2016). The genetic sex confirmed this identification.

3.4.2 Maternal and paternal lineage

The mitochondrial coverage was close to 10x and the complete genome was recovered. The individual was securely determined as belonging to haplogroup T2e1a. The subclade of T2, T2e, is a predominantly southern European and Mediterranean haplogroup (Pala et al., 2012), which has been found in Bell Beakers in central Europe and Bronze Age England (Olalde et al., 2018). Subhaplogroup T2e1a has been reported from only one ancient individual, from Amesbury Down in Wiltshire, England (Olalde et al., 2018). The predicted Y-chromosome haplogroup of the Low Hauxley Beaker individual, R1b1a1a2, has previously been found in Bell Beakers from Poland (Olalde et al., 2018), making both mitochondrial and Y-chromosome lineages of the Low Hauxley individual well established lineages within the Bell Beakers in Europe.

3.4.3 Bell Beaker ancestry

Similar to the Bronze Age individuals from Orkney (Chapter 4), the Bronze Age individual from Low Hauxley clustered closest with European Late Neolithic/Bronze Age individuals (Figure 3.3). However, admixture analysis showed that, although the highest proportions of ancestry were comparable to Late Neolithic/Bronze Age Europeans, he also had a small portion of shared ancestry with modern Near Eastern/North Africans. This 'African' component has also been found in modern Near Eastern and southern European populations, such as, for example, Spain and Sicily, but also to a certain extent in Neolithic and Bronze Age individuals from the Levant (Figure 3.4). North African ancestry has been shown to be highest in southwestern Europe with decreasing proportion towards the northern latitudes, and the sharpest difference appearing between Iberia and France (Botigué et al., 2013). However, this admixture has been estimated to be much more recent, dating to the Roman Empire and subsequent Arab migrations (Moorjani et al., 2011), and has not been observed in any European Bronze Age individuals analysed to date.

Based on suggested Bell Beaker migrations to Britain (Olalde et al., 2018), it is unlikely that the North African component in the Low Hauxley individual derived from Levantine or Near Eastern ancestry. Genetic analysis of Iberian Bronze Age cattle has shown that African and Iberian cultures must have been linked, as mitochondrial genomes of African cattle have been found in Iberian Bronze Age cattle (Anderung et al., 2005), suggesting interbreeding and cultural exchange. An extensive study on the genetic history of humans in Iberia found that a Copper Age individual, excavated in a Bell Beaker context at Camino de las Yeseras (San Fernando de Henares, Madrid, Spain), clustered with modern and ancient North Africans (Olalde et al. in review), and the authors suggested that the individual was a North African migrant. The association of Iberian and North African cultures during the Copper and Bronze Age has further been confirmed by analysis of a Bronze Age individual from Loma del Puerco (Chiclana de la Frontera, Cádiz), who carried 25% North African ancestry. The genetic similarity and shared association to the Bell Beaker culture between the reported individuals from Iberia (Olalde et al. in review) and the Bronze Age individual from Low Hauxley, England, suggests long-distance migrations of Bronze Age individuals. The North African component of the Low Hauxley individual suggests an ancestry-link to North Africa through Iberia, and supports the cultural connection between

Iberia and Britain during the Bronze Age, if low resolution artefacts can be ruled out by further sequencing.

Chapter 4

The Bronze Age and Iron Age in Orkney

4.1 Introduction

A Bronze Age cemetery was uncovered during excavations on the Links of Noltland (LoN), Westray, Orkney (Figure 4.1). The cemetery contained both cremation and inhumation burials, with the inhumations arranged to the north and west of a central cremation deposit. Two paired Bronze Age structures appeared to be the only boundaries, with formal cemetery boundaries missing. Two burial types are typically found in Orkney: flat cist cemeteries associated with mounds, and artificial mounds with cists inside (Moore and Wilson, 1995). However, the Bronze Age cemetery on the Links of Noltland does not fit into either of these categories. No mound could be found in relation to the cemetery, but it is possible that this might be due to severe erosion in the area. Considering the position of the Noltland cemetery to the paired Bronze Age structures, and the possible field systems to the north of the site, the funerary area appears to have had close associations with agricultural complexes (as discussed by Ovrevik 1985).

Twenty five articulated and twenty six disarticulated skeletons, as well as thirty six cremation deposits, were discovered. The articulated human remains could be aged accurately. Most women and all children were buried with grave goods, as was common practice during the Bronze Age, particularly in Ireland (Woddell 1990). Children were buried in stone-lined cists, while the majority of the adult inhumations were buried in simple sand-cut graves. Multiple burials appeared to be a common feature of this cemetery, with at least three identified multiple burials. The cremations were buried in simple sand-dug pits. At other sites in Bronze Age Orkney, multiple burials began to be replaced by single cist burials during the later Bronze Age, which

were then replaced by cremations. At Noltland, however, it appears from radiocarbon dates that inhumations and cremations occurred at the same time, suggesting that both rituals were of equal importance here. 25 of the excavated individuals were sampled for genome-wide analysis.

FIGURE 4.1: Map of the island of Westray, Orkney. The green circle indicates the position of the Bronze Age excavation site at the Links of Noltland, while the red circle indicates the position of the Iron Age excavation site at Knowe of Skea.

Additional to the Bronze Age cemetery on Westray, an Iron Age burial site was excavated at Knowe of Skea (KoS). The site of Knowe of Skea encompassed burials and material from the Neolithic to the Early Medieval era, such as a monumental stone building, and several buildings dating to the Bronze and Iron Age. The incredible number of individuals (>100) dating to the Iron Age within this site indicates an exceptional and unlikely domestic function of the cemetery, especially as over 60 individuals were neonates (Armit et al., 2007), and the cemetery is the largest yet discovered in Scotland (Moore and Wilson, unpublished). Three disarticulated inhumations of an adolescent, a middle aged adult and an older adult from Knowe of Skea were sampled for genetic analysis.

Uniparental markers have been used to analyse kinship among ancient multiple burials (Haak et al., 2008). Other studies have used mitochondrial HVS analysis and nuclear short tandem repeats (STRs) on ancient human remains to analyse kin relationships; for example, between individuals from a villa in Pompeii (Di Bernardo et al., 2009), bodies from sarcophaghi (Deguilloux

et al., 2014), Mongolian noble burials (Cui et al., 2015), or among Bronze Age pit burials (Palomo-Díez et al., 2018). However, STRs are generally difficult to type in ancient remains, and uniparental markers can only exclude certain direct relationships, as, even among unrelated individuals, most of the mitochondrial and Y-chromosome haplogroups found are relatively common (Kuhn et al., 2018), especially within Europe. As uniparental markers can not be used to infer any direct degree of kinship, genome-wide data is nowadays beginning to be used to identify individuals and their kin relationships (Kuhn et al., 2018). Tools for kinship analysis of ancient remains, such as the recently published READ (Kuhn et al., 2018), now allow for first- and second-degree kin analysis on low coverage whole-genome sequence data. READ assumes pseudohaploid data and performs a normalisation using an estimate for the pairwise distance of unrelated individuals to help overcome biases potentially arising from e.g. SNP ascertainment of general population diversity. Pair-wise distances are calculated between individuals and the given normalised P0 score, and standard errors denoted as Z_upper and Z_lower are used to assess the kinship classification. However, the results need to be interpreted with care, as the chance of false positive classifications is increased as the number of SNP loci decreases. Uniparental markers and archaeological context information should be used to facilitate the interpretation of the results.
The genetic analysis of 25 articulated inhumations of the Links of Noltland cemetery and three adult individuals from the Iron Age mass burial site at the Knowe of Skea cemetery aimed to identify possible family structures in multiple burials and across the two cemeteries using uniparental markers and kinship analysis. It was further envisaged that nuclear data might give insight into possible migration patterns to Orkney during the Bronze Age, and how/whether this persisted into the Iron Age.

4.2 Materials and Methods

Overall, 25 samples from the Bronze Age cemetery at Links of Noltland, consisting of 16 petrous bones and 9 teeth, were processed. All of the samples were part of the main cemetery and buried close to one another. Five of the individuals derived from a single grave that included multiple inhumations: inhumation numbers 9275 (1620-1450 cal. BC), 9302 (1499-1386 cal. BC), 9295 (1623-1458 cal. BC), 9291G (1610-1440 cal. BC) and 9304. Two other inhumations, 9414P and 9414S, also derived from the same grave. Of the remaining

TABLE 4.1: Processed samples of the cemeteries from Links of Noltland and Knowe of Skea.

DNA code	DNA sample	Analysis type	Era	Site information	Sample code
KD004	R petrous	L	Iron Age	Knowe of Skea	BNKS 4038
KD006	R petrous	L	Bronze Age	Links of Noltland	LON 9293
KD042	R petrous	M	Iron Age	Knowe of Skea	BNKS 2115
KD043	L petrous	M	Iron Age	Knowe of Skea	BNKS 4045
KD044	LM3	M	Bronze Age	Links of Noltland	LON 9488
KD045	U incisor	M	Bronze Age	Links of Noltland	LON 9395
KD046	molar	M	Bronze Age	Links of Noltland	LON 9275 (Group P)
KD047	R petrous	M	Bronze Age	Links of Noltland	LON 9302
KD048	L petrous	M	Bronze Age	Links of Noltland	LON 9244
KD049	R petrous	M	Bronze Age	Links of Noltland	LON 9275 (Group I)
KD050	R petrous	M	Bronze Age	Links of Noltland	LON 9275 (Group C)
KD051	R petrous	M	Bronze Age	Links of Noltland	LON 9280
KD052	R petrous	M	Bronze Age	Links of Noltland	LON 9414
KD053	R petrous	M	Bronze Age	Links of Noltland	LON 9414
KD054	decidious molar	n/a	Bronze Age	Links of Noltland	LON 9303
KD055	URM3	M	Bronze Age	Links of Noltland	LON 9373
KD056	UL incisor	no library	Bronze Age	Links of Noltland	LON 9404
KD057	lower incisor	M	Bronze Age	Links of Noltland	LON 9374
KD058	R petrous	M	Bronze Age	Links of Noltland	LON 9295
KD059	lower incisor	M	Bronze Age	Links of Noltland	LON 9053
KD060	R petrous	M	Bronze Age	Links of Noltland	LON 9412
KD061	L petrous	M	Bronze Age	Links of Noltland	LON 9326
KD062	R petrous	M	Bronze Age	Links of Noltland	LON 9290
KD063	L petrous	M	Bronze Age	Links of Noltland	LON 9335
KD064	R petrous	M	Bronze Age	Links of Noltland	LON 9370
KD065	L petrous	M	Bronze Age	Links of Noltland	LON 9291G
KD066	LLM1	M	Bronze Age	Links of Noltland	LON 9304
KD067	L petrous	M	Bronze Age	Links of Noltland	LON 9442

individuals, four had been dated: 9244 (1520-1410 cal. BC), 9280 (1750-1530 cal. BC), 9053 (1630-1460 cal. BC) and 9290 (1610-1420 cal. BC).

Three individuals from the Iron Age cemetery at Knowe of Skea were sampled for their petrous bones. Of these, individual 4045 was dated to 1AD to 220 cal. AD. The sampling procedure and NGS library preparation was as described in Chapter 2. Samples LON 9293 and BNKS 4038 were dual-indexed for 100bp single-end sequencing on a HiSeq2500 (NBAF Liverpool, UK), while all other samples were single-indexed for 100bp paired-end sequencing on a HiSeq4000 (Macrogen, South Korea). Bioinformatic processing followed all steps as described in Chapter 2. Variants in linkage disequilibrium were filtered using the –indep-pairwise option in PLINK v0.67 with the parameters 200, 25 and 0.4. ADMIXTURE was run for ancestral population numbers from K=2 to K=14, with cross-validation enabled (–cv flag), on a merged dataset including the here presented samples together with the ancient and modern data of Lazaridis et al. (2016) and Olalde et al. (2018).

4.3 Results

Out of the 25 samples from the Links of Noltland, the library preparation of one sample (9303, decidious molar) failed and, therefore, the sample was excluded from sequencing. The library preparation of sample 9404 failed and it was thus excluded from any further analysis. All other 23 samples were analysed as described in Chapter 2.

4.3.1 Sequencing results

The endogenous DNA content of the samples ranged from 3.08 to 66.46% (Table 4.2). Whole genome coverage ranged from 0.0014x to 0.1293x, while the coverage of the mitochondrial genome ranged from 0.3328x to 51.2858x. The four libraries sequenced for KD006 (KD006.1, KD006.2, KD006.3 and KD006.4) had read counts of 5,968,323, 12,920,228, 6,786,040, and 1,078,610, the range of read counts is given in Table 4.2. The endogenous content of the four libraries sequenced for KD006 was 18.67%, 17.05%, 18.27%, and 16.82% respectively, and so its overall endogenous content is shown as a range in Table 4.2.

All three samples from Knowe of Skea passed the library preparation stage and were sequenced successfully. Their endogenous DNA contents ranged from 16.32 to 46.28%. The read count and endogenous content for KD004 are given as a range (Table 4.3, as a total of four single-end libraries were produced (KD004.1, KD004.2, KD004.3 and KD004.4), resulting in 28.34%, 29.43%, 28.58% and 28.18% endogenous content, and 7,670,025, 10,642,858, 26,376,756, and 22,201,481 reads, respectively. Whole genome coverage ranged from 0.0799x to 0.1100x, whilst mitochondrial genome coverage ranged from 4.8221x to 46.3234x. The duplication rate was <0.8% for all samples from both sites (see Tables 4.2 and 4.3).

TABLE 4.2: Sequencing results from 23 Bronze Age Links of Noltland samples and three Iron Age Knowe of Skea samples, from Westray, Orkney. Endogenous content for KD004 and KD006 is given as a range as explained in the text. 'SNPs' denotes the number of SNPs called with the published datasets as reference panel.

Lab ID	Sample ID	Read count	endogenous content in %	duplication rate in %	genome coverage	SNPs
KD006	LON 9293	5,968,323-12,920,228	16.8200-18.6700	0.7800	0.1073	50,199
KD044	LON 9488	92,982,52	66.3500	0.1123	0.1293	75,076
KD045	LON 9395	59,183,206	33.2700	0.1023	0.0196	11,720
KD046	LON 9275 Group P	85,415,490	9.0500	0.0937	0.0054	3,214
KD047	LON 9302	63,542,862	19.6700	0.0853	0.0235	16,045
KD048	LON 9244	62,890,464	5.9500	0.0702	0.0027	1,931
KD049	LON 9275 Group I	55,679,898	23.0600	0.0786	0.0368	25,407
KD050	LON 9275 Group C	89,513,406	17.9500	0.9870	0.0493	32,503
KD051	LON 9280	46,069,394	3.3000	0.0721	0.0018	1,278
KD052	LON 9414 primary burial	53,124,892	4.6200	0.0595	0.0025	1,686
KD053	LON 9414 secondary burial	52,647,474	3.0800	0.6780	0.0014	945
KD055	LON 9373	59,521,362	59.7400	0.0916	0.0400	25,866
KD057	LON 9374	67,743,896	28.3700	0.0897	0.0180	11,293
KD058	LON 9295	90,002,570	14.6100	0.1049	0.0181	12,197
KD059	LON 9053	82,213,502	29.4900	0.0857	0.0120	8,017
KD060	LON 9412	93,204,030	22.2300	0.1056	0.0662	44,509
KD061	LON 9326	56,421,470	52.0300	0.0967	0.0830	53,263
KD062	LON 9290	65,335,370	22.9300	0.0883	0.0308	20,190
KD063	LON 9335	75,345,786	60.5600	0.0936	0.1261	78,392
KD064	LON 9370	69,254,088	66.4600	0.1037	0.1277	82,849
KD065	LON 9291G	71,422,136	29.1300	0.1075	0.0451	30,564
KD066	LON 9304	72,430,882	10.8500	0.0779	0.0065	4,495
KD067	LON 9442	74,888,922	60.4700	0.1071	0.1101	68,666

TABLE 4.3: Sequencing results from three Iron Age Knowe of Skea samples, from Westray, Orkney. Endogenous content for KD004 is given as a range as explained in the text. 'SNPs' denotes the number of SNPs called with the published datasets as reference panel.

Lab ID	Sample ID	Read count	endogenous content in %	duplication rate in %	genome coverage	SNPs
KD004	BNKS 4038	7,670025-26,376,756	28.1800-29.4300	0.0799	0.3106	150,633
KD042	BNKS 2115	47,338,034	16.3200	0.1041	0.0230	13,564
KD043	BNKS 4045	54,345,210	46.2800	0.1100	0.1539	86,019

4.3.2 Genetic sex determination

Of the individuals whose sex could be determined using the Skoglund script, seven individuals were assigned as female, six as male and three as consistent with XY but not with XX (Table 4.4). For the remaining seven individuals whose genetic sex could not be determined using this method, karyotype plots were prepared and the X-chromosome percentage calculated (see Appendix E for details). This resulted in the determination of two males, three females and the sex of two individuals could not be assigned, due to low coverage.

4.3.3 Mitochondrial genome analysis

In the overall dataset, the main haplogroups observed were H, U5, T2 and K1. Two individuals had such a low mitochondrial genome coverage (KD051 and KD053) that no haplotype could be assigned. Five individuals were identified as belonging to haplogroup H39 (KD047-50 and KD065), while four other individuals were assigned to other subclades of the major European haplogroup H, namely H1n1 (KD057 and KD060), H58a (KD063), and H+195 (KD067). KD006, KD046 and KD066 belonged to haplogroup T2a1b1a, KD059 and KD064 to T2b21, KD045 and KD55 to J1c2a, and KD044 and KD062 to U5b2a3. One individual each was identified to belong to haplogroups K1a29a (KD052), K1a3a (KD058) and K1c2 (KD061).
The two males from Knowe of Skea belonged to the same mitochondrial haplogroup, H1b (KD004 and KD043), and had Y-chromosome haplogroups R1b1a1a2 and R1b1a1a2a1a2c1a2, respectively. The third individual (KD042), a woman, had mitochondrial haplogroup U5a1b1a. The complete mutation lists of each individual can be found in Appendix F.

4.3.4 Y-chromosome analysis

The results of the Y-chromosome analysis are summarised in Table 4.4. The complete lists of derived predictive markers per individual can be found in Appendix G.

TABLE 4.4: Osteological age estimate, genetic sex and identified uniparental markers are given. Bolded genetic sex indicates assignment based on Skoglund score. Ages with "?" are based on size of petrous bone only.

Lab ID	Age	Genetic sex identification	Mitogenome coverage	mtDNA haplogroup	Y-chromosome haplogroup	chr Y markers
KD006	Mature Adult	XY	12.7250	T2a1b1a	I2a1b	1180
KD044	Old Middle Adult	XX	24.9947	U5b2a3	–	–
KD045	Young Adult	consistent with XY but not XX	15.6955	J1c2a	I2a1b	408
KD046	Undetermined	**XY**	6.3507	T2a1b1a	–	–
KD047	Mature Adult	XY	2.7727	H39	I2a1a1a1∼	572
KD048	Infant/child	**Not assigned**	1.1570	H39	–	–
KD049	Juvenile?	XY	7.8708	H39	I	936
KD050	Juvenile?	XX	12.3676	H39	–	–
KD051	Juvenile	**Not assigned**	0.3328	–	–	–
KD052	Neonate	**XX**	0.8791	K1a29a	–	–
KD053	Foetus	**XY**	0.5172	–	–	–
KD055	Adult	XX	15.1104	J1c2a	-	–
KD057	Mature Adult	consistent with XY but not XX	19.8839	H1n1	I	346
KD058	Young Middle Adult	**XX**	4.1759	K1a3a	–	–
KD059	Undetermined	consistent with XY but not XX	15.4143	T2b21	I	285
KD060	Young Middle Adult	XY	12.8498	H1n1	I	1624
KD061	Neonate	XY	21.0532	K1c2	R1b1a1a2	2051
KD062	Foetus	XX	10.7808	U5b2a3	–	–
KD063	Foetus	XX	51.2858	H58a	–	–
KD064	Neonate	XY	47.2928	T2b21	I2a1b	3179
KD065	Foetus	XX	12.5107	H39	–	–
KD066	Undetermined	**XX**	4.4250	T2a1b1a	–	–
KD067	Foetus	XX	48.8611	H+195	–	–
KD004	Adolescent	XY	46.3234	H1b	R1b1a1a2a1a2c1a2	1971
KD042	Middle Adult	XX	4.8221	U5a1b1a	–	–
KD043	Older Adult	XY	17.0452	H1b	R1b1a1a2	106

4.3.5 Kin relationship estimation

4.3.5.1 Links of Noltland

The kinship estimation with READ indicated that most individuals were first or second-degree related. The majority of the five individuals buried together in one grave (sample numbers KD047, KD049, KD050, KD058, KD065, see Table 4.1) were estimated to be first-degree related. The estimated degrees of kinship are shown in Table 4.5. The likely relationship between these individuals is shown schematically in Figure 4.2, as determined from the information shown in the comparison square (Table 4.6).

TABLE 4.5: Estimated genetic kin relationships of the Links of Noltland individuals. Higher and lower class of genetic difference (Z_upper and Z_lower) are given as standard errors from the normalised mean *P*0 score.

PairIndividuals	Relationship	Z_upper	Z_lower
KD047+KD049	Second Degree	3.1727	-0.0328
KD047+KD050	Second Degree	3.4483	-0.1341
KD047+KD058	First Degree	0.5374	-4.1246
KD047+KD059	Second Degree	0.8253	-0.8373
KD047+KD065	Second Degree	1.3654	-2.0008
KD049+KD050	First Degree	3.1947	-5.1637
KD049+KD058	First Degree	1.04524	-4.5297
KD049+KD062	Second Degree	0.6992	-2.6773
KD049+KD065	First Degree	0.8454	-7.0129
KD050+KD058	Second Degree	2.6652	-0.5090
KD050+KD065	First Degree	7.9718	-0.6162
KD057+KD061	Second Degree	0.0637	-3.2328
KD057+KD062	Second Degree	0.0718	-2.0480
KD058+KD062	Second Degree	2.2536	-0.1100
KD058+KD065	Second Degree	2.5362	-0.3284
KD045+KD047	Second Degree	0.1992	-1.8503
KD045+KD058	Second Degree	0.0747	-1.7479

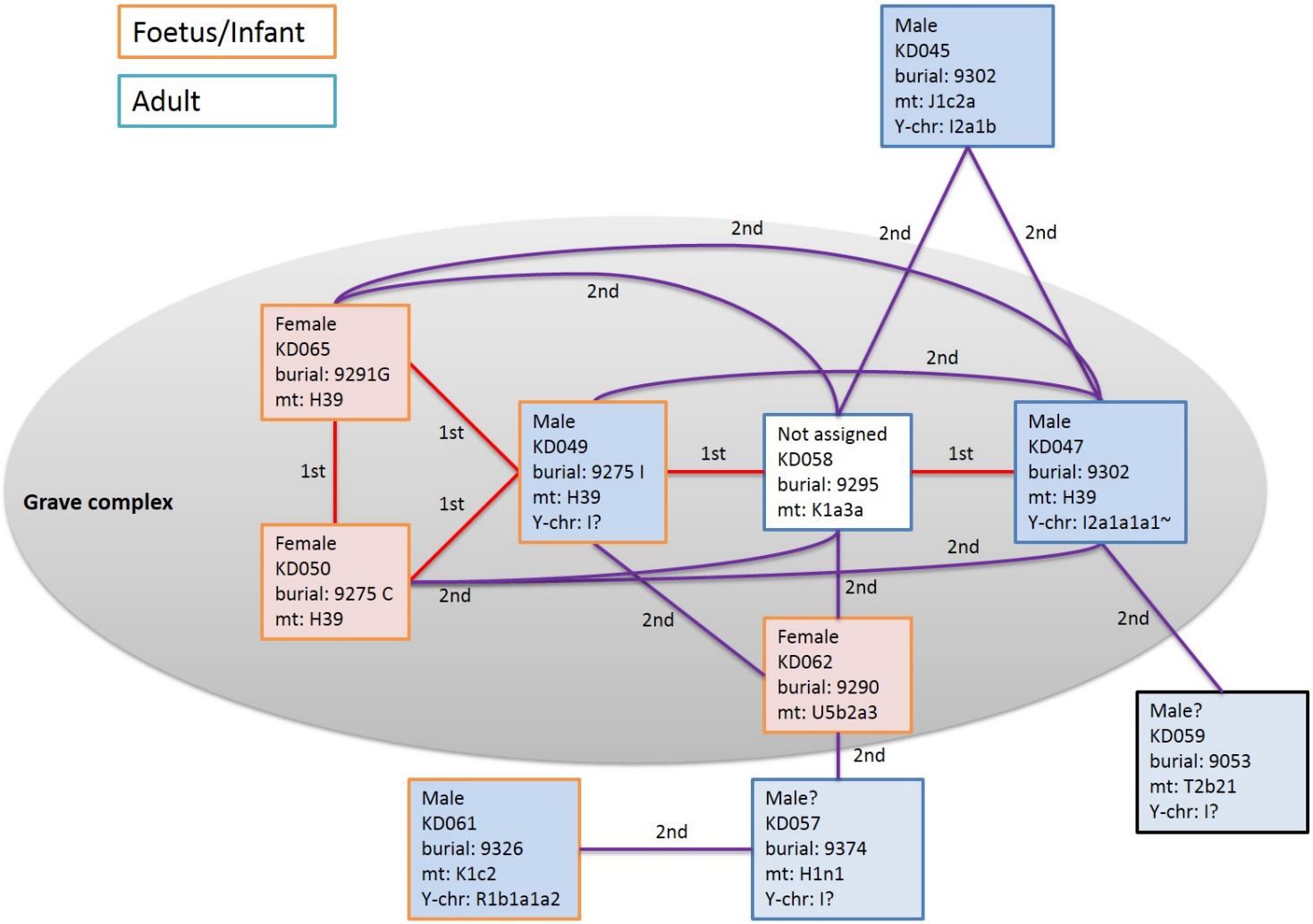

FIGURE 4.2: Schematic overview of the kin relationships between the analysed individuals of the Bronze Age cemetery.

TABLE 4.6: Overview of the kinship degrees between the Bronze Age LoN individuals identified as related to each other. Green indicates first-degree kin relationships, such as parent-offspring and siblings, and yellow indicates second-degree kin relationships, such as grandparent-grandchild and aunt/uncle-nephew/niece.

Code	Haplotype	Sex	Age	Grave	KD045 J1c2a M adult 9395	KD047 H39 M adult 9302	KD049 H39 M juvenile? 9275	KD050 H39 F juvenile? 9275	KD057 H1n1 M adult 9374	KD058 K1a3a M adult 9295	KD059 T2b21 M ? 9053	KD061 K1c2 M baby 9326	KD062 U5b2a3 F baby 9290	KD065 H39 F baby 9291G
KD045	J1c2a	M	18-25 years	9395		2nd	X	X	X	2nd	X	X	X	X
KD047	H39	M	>46 years	9302	2nd		2nd	2nd	X	1st	2nd	X	X	2nd
KD049	H39	M	juvenile?	9275	X	2nd		1st	X	1st	X	X	2nd	1st
KD050	H39	F	juvenile?	9275	X	2nd	1st		X	2nd	X	X	X	1st
KD057	H1n1	M	>46 years	9374	X	X	X	X		X	X	2nd	X	X
KD058	K1a3a	M	26-35 years	9295	2nd	1st	1st	2nd	X		X	X	2nd	2nd
KD059	T2b21	M	?	9053	X	2nd	X	X	X	X		X	X	X
KD061	K1c2	M	neonate	9326	X	X	X	X	2nd	X	X		2nd	X
KD062	U5b2a3	F	foetus	9290	X	X	2nd	X	X	2nd	X	2nd		X
KD065	H39	F	foetus	9291G	X	2nd	1st	1st	X	2nd	X	X	X	

4.3.5.2 Knowe of Skea

The only kin relationship that was identified among the Iron Age KoS samples was that the middle-aged adult woman (KD042) and the older adult man (KD043) are likely second-degree related (Table 4.7).

TABLE 4.7: Estimated genetic kin relationships of the Knowe of Skea individuals. Higher and lower class of genetic difference (Z_upper and Z_lower) are given as standard errors from the normalised mean *P*0 score.

Sample pair	Relationship	Z_upper	Z_lower
KD042+KD043	Second-degree	1.02717578217	-6.75304406054
KD042+KD004	Unrelated	NA	-13.42318498
KD043+KD004	Unrelated	NA	-113.803369616

4.3.6 Principal component analysis

Samples that had more than 10,000 SNPs in common with the published modern and ancient dataset (Lazaridis et al., 2016) were used for principal component analysis. Figure 4.3 shows that the Bronze Age individuals from Links of Noltland cluster closest together with the European Late Neolithic/Bronze Age individuals.

4.3.7 ADMIXTURE analysis

Samples with less than 10,000 SNPs were excluded from this analysis. The best value of *K* (lowest CV error; see Chapter 2 and Appendix H for details) generated for this ADMIXTURE analysis was *K*=12 (Figure 4.4).

4.4 Discussion

4.4.1 Genome coverage

Two samples from the Bronze Age cemetery at Links of Noltland were excluded from the analysis due to failures in the library preparation and sequencing stages. All remaining 23 samples were successfully sequenced, although to a very low coverage of less than 0.1x on average. Surprisingly, 20 of them had endogenous contents of over 10%, seven of which were above 40%. All three samples from the Iron Age cemetery at Knowe of Skea had endogenous contents over 10% and were sequenced successfully.
Due to the low coverage of the LoN Bronze Age samples, the genetic sex of

FIGURE 4.3: Principal component analysis (PCA) of the Links of Noltland (LON) and Knowe of Skea (KOS) samples merged with ancient European and Near Eastern populations (Lazaridis et al., 2016).

only 16 individuals could be determined using the Skoglund script (Skoglund et al., 2013). Using karyotype plots and X-chromosome percentages (Skoglund et al., 2015), it was possible to assign genetic sex to a further five of the otherwise undetermined individuals. Therefore, there were 10 females and 11 males (including the three individuals that were consistent with XY but not XX). The genetic sex of two individuals could not be assigned using either method as the coverage was too low, and the Skoglund score was inconclusive (Appendix E).

4.4.2 Estimated kin relationship

Based on the radiocarbon dates from the site, the skeletons were likely interred contemporaneously. This is especially interesting in the case of the boat-shaped grave complex containing 17 articulated and disarticulated individuals. Six of these were analysed genetically in this study (Figure 4.2), and five were identified as having a first-degree relationship. This supports the archaeological interpretation that the boat-shaped grave complex contained family-based burials. The two individuals 9414P (KD052) and 9414S (KD053) that were buried together in another grave did not produce enough

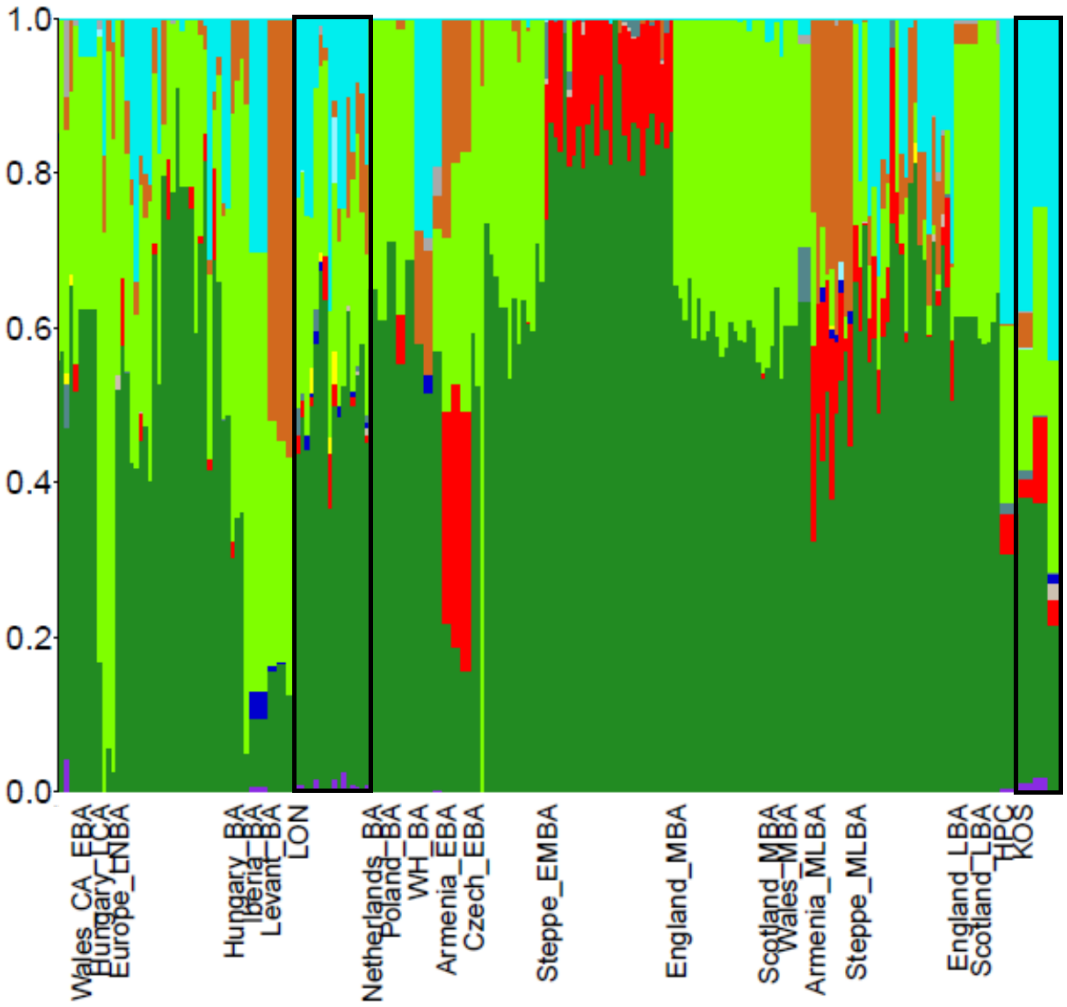

FIGURE 4.4: ADMIXTURE plot of the Bronze Age and Iron Age samples from Orkney together with ancient European and Near Eastern individuals from the same time period from Lazaridis et al. (2016) and Olalde et al. (2018). Abbreviations used: E: Early, M: Middle, L: Late, CA: Copper Age, BA: Bronze Age, LON: Links of Noltland, WH: West Heslerton, HPC: High Pasture Cave, KOS: Knowe of Skea.

variants to estimate any degree of kinship, and the coverage of KD053 was even too low to determine the mitochondrial haplogroup. It was, therefore, not possible to assign even a maternal relationship to these individuals.

Individuals buried in the boat-shaped grave complex and a few individuals buried in close proximity to it were identified as second-degree kin (see Figure 4.2). The adult male in burial 9374 (KD057) was found to be second-degree related to the infant male KD061 in burial 9326, as well as to the infant female KD062 in burial 9290. All three belong to very different

mitochondrial haplogroups, namely H1n1, K1c2 and U5b2a3 respectively. The Y-chromosome coverage of the adult male was too low to predict a Y-chromosome haplogroup, and so it was not possible to identify whether the adult male KD057 and the infant male KD061 were related through the paternal line. Based on the burials being contemporaneous, and the age of the individuals (KD057 was an old man, whilst the two related individuals, KD061 and KD062, were foetuses or infants), it is possible that KD061 may have been their grandfather or uncle.

The infant female KD062 was one of the individuals buried inside the boat-shaped grave complex, being also second-degree related to both the possibly juvenile male KD049 in burial 9275I and the adult male KD058 in burial 9295. She could have been the niece of KD049 and the granddaughter of KD058. Individuals KD058 and KD049 were first-degree related and most likely father and son, but, unfortunately, analysis of their Y-chromosome haplogroups was not possible due to low coverage. KD049 shared mitochondrial haplogroup H39 with the juvenile female KD050 in burial 9275C and the infant female KD065 in burial 9291G. All three were estimated to be first-degree related to each other, which suggests that they were siblings. The degree of kinship between adult male KD058, to KD050 and KD065, was determined as second-degree, which is contrary to the previously described scenario of KD058 being the father of all of them and not just KD049. However, it is possible that the degree of kin relationship was misclassified as second-degree rather than first-degree due to low coverage.

Within the boat-shaped grave, KD058 was first-degree related to the older adult male KD047 and they share mitochondrial haplogroup H39. Based on the age difference between them (KD058 was 26-35 years old, while KD047 was over 46), it seems most likely that KD047 was the father of KD058. This fits nicely with the second-degree kinship between KD047 and the three children KD049, KD050, and KD065, who were likely his grandchildren. Two further individuals that were buried outside the grave complex were identified as being second-degree related to individuals within the boat-shaped grave complex. An adult male KD045 was second-degree related to both males KD047 and KD058. Based on the age differences of the individuals, KD047 may have been the father of KD058 and the grandfather of KD045. KD058 would have then been the uncle of KD045 with his sister being the mother of KD045. The old adult male KD047 was also determined to be second-degree related to another individual from outside the grave complex, male KD059. Neither the age nor the Y-chromosome of KD059 is unknown,

so it is not possible to identify which of the two individuals would be the grandfather/uncle of the other, but there is a kin relationship.

4.4.3 Patrilocality

Overall, there appears to be patrilocality within the cemetery as, with the exception of the siblings and their grandfather belonging to mitochondrial haplogroup H39, no connection based on the female line of descent could be observed. All adult individuals that showed degrees of kinship were males. Samples KD046, KD048, KD051, KD052, KD053, KD059, and KD066 produced less than 10,000 SNPs and were, therefore, not suitable for READ analysis to estimate the degree of kinship. However, the analysis of all individuals within the Bronze Age cemetery based on their mitochondrial haplogroups indicated shared maternal lineages. Individuals KD006, KD046 and KD066 shared mitochondrial haplogroup T2a1b1a. The adult female KD044 shared haplogroup U5b2a3 with the female foetus KD062, which possibly makes it her child or grandchild. The young adult male KD045 shares the haplogroup J1c2a with the adult female KD055. The two adult males KD057 and KD060 share haplogroup H1n1, and male KD059 and neonate male KD064 both belong to haplogroup T2b21. The sharing of mitochondrial haplogroups suggests a shared maternal lineage, which is supported by the general finding that individuals within this cemetery were closely related. Similarly, shared Y-chromosome lineages indicated shared paternal lines of descent. Generally, the Y-chromosome coverage was very low and a low number of predictive markers was retrieved for most males, which made it impossible to determine the haplogroup more precisely than to haplogroup I for the majority. However, KD006, KD045 and KD064 had a higher number of derived predictive markers, and it was possible to identify that they shared Y-chromosome haplogroup I2a1b. Based on their individual ages at death, KD006 was possibly the senior relative.

4.4.4 Iron Age individuals from the Knowe of Skea

Mitochondrial genome analysis of the three Iron Age individuals from Knowe of Skea showed that the two male individuals were likely related through the maternal line, as they both belonged to haplogroup H1b. Although, their predicted Y-chromosome assignments showed that they both belonged to the same major Y-chromosome haplogroup R1b, their haplotypes were different, thus excluding a kinship through the paternal line. The predicted

Y-chromosome haplogroup R1b1a1a2a1a2c1a2 of the adolescent (KD004) is more derived than the predicted Y-chromosome haplogroup R1b1a1a2 of the older adult (KD043). Additional kinship analysis revealed an estimated second-degree relationship between the woman (KD042) and the older man (KD043). As the middle-aged adult woman belongs to a very different mitochondrial haplogroup than the two men, namely U5a1b1a, a maternal line of descent can be excluded. Therefore, the second-degree kin relationship suggests a grandfather-granddaughter relationship or an uncle-niece relationship, although it can't be excluded that they were half-siblings, sharing the same father.

4.4.5 Conclusions

In general, it appears that both the Bronze Age and Iron Age cemeteries on Westray, Orkney, represented familial use as, in both cases, most of the analysed individuals at each site were closely related to each other. However, it should be noted that the number of individuals from the Iron Age cemetery (n = 3) may not be representative of that time period, and the identified kin relationships might be an exception, especially since the vast majority of the individuals at Knowe of Skea were infants. Further genetic analysis of the remaining individuals would give information about the degree of kinship at the site as a whole, and allow for a more direct comparison to the Bronze Age cemetery at Links of Noltland.

PCA and ADMIXTURE analyses showed that the Bronze Age individuals from the Links of Noltland, as well as the Iron Age individuals from Knowe of Skea, showed very similar patterns to Late Neolithic/Bronze Age individuals from Europe. This striking similarity between the Bronze Age and Iron Age on Westray is highly suggestive of population continuity over at least 2,000 years.

Chapter 5

An Iron Age burial at the High Pasture Cave, Isle of Skye

5.1 Abstract

Human remains from the Iron Age in Atlantic Scotland are rare, which makes the assemblage of an adult female and numerous foetal bones at High Pasture Cave, on the Isle of Skye, particularly noteworthy. Archaeological evidence suggests that the female had been deposited as an articulated skeleton, as the cave entrance was blocked off, possibly as part of a 'closing deposit'. Particularly intriguing is the deposition of disarticulated remains from a perinate and a foetus close to the adult female. This intentional burial suggests that the female might have been the mother of both of the children. Shotgun genome sequencing was used in order to analyse the mitochondrial genomes of all three individuals, in order to investigate their relationships. The results revealed that, while the female could be the mother of the foetus, the perinate was not her child.

5.2 Introduction

High Pasture Cave (Uamh an Ard Achadh) is situated in the Parish of Strath on the Isle of Skye (Figure 5.1).
Stonework found in the surface trench was associated with a series of entrance arrangements, which led down a stone stairwell into the cave, whose system provides direct access to an underground stream. The passage had been deliberately blocked with boulders and sediments (Birch et al., 2013), with deposition of human remains from three individuals into the top of the blockage (Figure 5.2), along with remains of a foetal pig and perinatal dog.

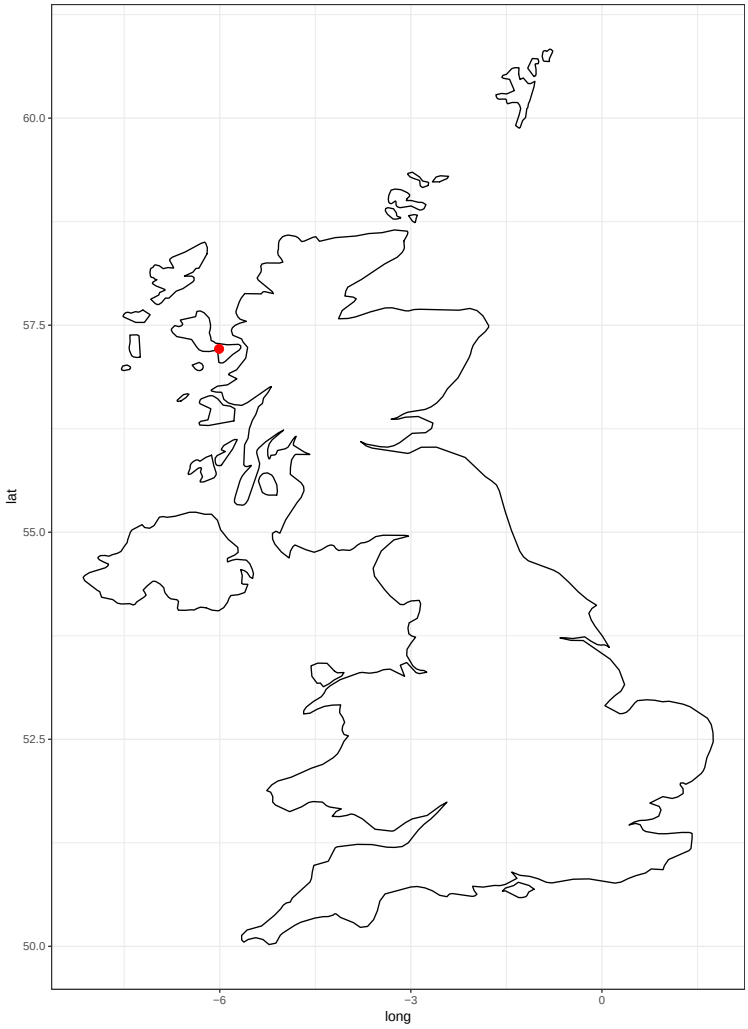

FIGURE 5.1: The location of the High Pasture Cave within the British Isles.

The remains found at this site display distinct ritual aspects, such as the burial of animals after butchery, and the presence of both disarticulated human remains and complete human inhumations, including foetal and infant burials (Birch et al., 2013; Livarda et al., 2017). The unusually high amount of pig remains also suggests a feasting deposit (Livarda et al., 2017).
The human remains consisted of an adult female, aged between 25-40 years, found together with two infants: one perinate that died during the last month of foetal development or during the first two weeks of life, and one foetus that died between 12 and 26 weeks of gestation. Further down the stairwell, the proximal half of a left tibia relating to a second perinate was found. These deposits might have moved down through the voids, away from the main burial environment (Birch et al., 2013) or else have been buried at a different time. Additional fragments of human bone, teeth and partially burnt

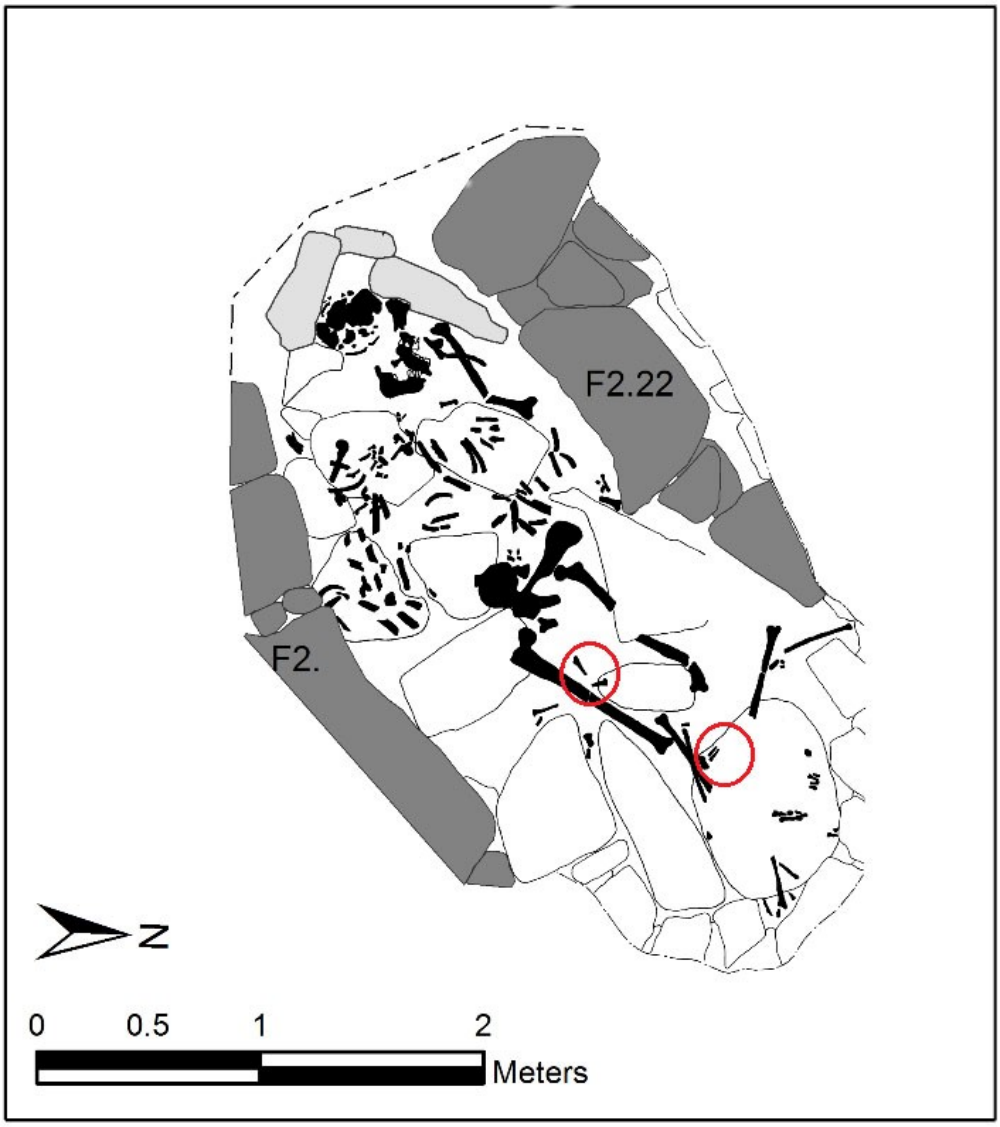

FIGURE 5.2: Schematic of the stairwell backfill with burials in the High Pasture Cave. The red circles indicate the position of the foetal bones.

human skull fragments were also discovered within the cave system, as well as outside the cave entrance (Birch et al., 2013). The woman appears to have been buried intact under the stones and this has led some to surmise that she may have been deliberately crushed to death (Armit, in press).

The wider environment surrounding the site revealed a diverse archaeological landscape of monuments and sites relating to prehistoric and historic periods, including a number of funerary monuments, and hut-circles combined with well-preserved field systems (Birch et al., 2013). Radiocarbon dates of samples from different layers within the cave, and from the complex deposits

outside the cave entrance, indicate periodic activity at High Pasture Cave from the 13th century *cal*. BC to *cal*. AD 1st century. This, and the diversity of small finds, indicates a complex occupation history with several phases of intense activity during the Iron Age (Birch et al., 2013). The archaeological material found in the cave indicates a typical domestic assemblage comprising bone, stone and iron tools, ceramics and residues of metalworking. Furthermore, a well-preserved faunal assemblage has been recovered, including animal and fish bones, shellfish and charred plant remains. Among the animal bones, a high percentage could be assigned to domesticated pig, which display evidence of unusual butchery practices, such as deliberate division of skeletons into left and right parts and unusual cut-marks across several pig ribs (Drews, 2005). The human remains of the High Pasture Cave adult female yielded a calibrated radiocarbon date of *cal*. AD 28-230 (2σ; SUERC-14946), which corresponds to the Iron Age in this region. The infant remains were radiocarbon dated to *cal*. AD 63-219 (SUERC-66371) for the foetus and *cal*. AD 76-228 (SUERC-66372) for the perinate (Figure 5.3).

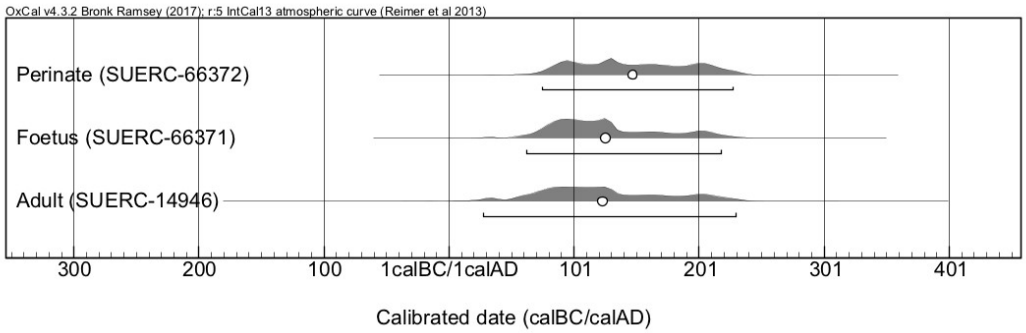

FIGURE 5.3: Calibrated radiocarbon dates for the analysed human remains of the High Pasture Cave stairwell.

During the Iron Age, unusual and specific burials for women, foetuses and infants were rather common (Finlay, 2000; Armit et al., 2007). The general archaeological interpretation has been that the women must have died in childbirth and were thus buried together with their foetuses. In order to validate this assumption the genetic relationship of the adult female to the foetus and the perinate was analysed.

5.3 Materials and Methods

5.3.1 Sampling and DNA extraction

A petrous bone from the female adult, and infant bone fragments from the perinate (ischium) and the foetus (scapula), were sampled for genetic analysis. The sampling was carried out at the Ancient DNA Facility at the University of Huddersfield under dedicated clean-room conditions. Laboratory researchers wore full body suits, hairnets, gloves and face masks throughout the drilling, extraction and library preparation processes. All tools and surfaces were constantly cleaned with LookOut® DNA Erase (SIGMA Life Sciences), as well as with bleach, ethanol and long exposures to UV light.
The bone surfaces were decontaminated by UV radiation for 30 minutes on each side, followed by cleaning with 5 µm aluminium oxide powder using a compressed air abrasive system. For the adult sample, the densest part of the petrous bone (Pinhasi et al., 2015) was excised using a circular saw. Bone powder was obtained by crushing this excised petrous portion, and the complete bone fragments from the two infants, in a MixerMill (Retsch MM400). DNA was extracted from ∼150mg of the petrous powder and 50-100mg of each infant bone, following the protocol by Yang et al. (1998) with modifications by MacHugh et al. (2000). Blank controls were included throughout the extractions, library preparation and amplification reactions to monitor for possible modern DNA contamination.

5.3.2 Library preparation and sequencing

Next-generation sequencing libraries were constructed from DNA extracts using the method described in Meyer et al. (2010), with modifications outlined in Gamba et al. (2014) and Martiniano et al. (2014). Briefly, the changes were as follows: T4 DNA polymerase buffer (Thermo Scientific) was used instead of Tango buffer in the blunt-end repair step and the *Bst* polymerase activity was heat-inactivated by incubating the libraries for 20 minutes at 80°C. Indexing amplifications were set up using Accuprime *Pfx* Supermix (Life Technology), primer IS4 (0.2µM), a specific indexing primer (0.2µM) and 3 µl of sample library, with a total reaction volume of 25 µl. All DNA purification steps were performed using the QIAQuick MinElute purification kit (Qiagen) following the manufacturer's protocol, with the modification of adding 0.05% Tween 20 (Fisher BioReagents) to the elution buffer.

Single-end libraries were prepared for both of the infant samples. Amplification was undertaken using the following thermal cycling conditions: 5 minutes at 95°C; 11 cycles of 15 seconds at 95°C, 30 seconds at 60°C and 30 seconds at 68°C; with a final extension step of 5 minutes at 68°C. The resulting amplification products were purified using the QIAQuick MinElute purification kit (Qiagen) as described above.
As the individual was part of an earlier study, four dual-index libraries (Kircher et al., 2012) were prepared for the adult sample. Amplification was set up using AmpliTaq Gold (5U/µl), Thermopol reaction buffer (10x), dNTPs (10mM each), both indexing primers (10µM each) and 10µl DNA sample library. The following thermal cycling steps were used for amplification: 12 minutes at 95°C; 10 cycles of 20 seconds at 95°C, 30 seconds at 60°C, and 40 seconds at 72°C; with a final extension step of 5 minutes at 72°C. We purified the amplification product using the QIAQuick MinElute PCR purification kit as described above. For the second amplification round, Accuprime *Pfx* Supermix (Life Technology) was used, together with primers IS5 (10µM), IS6 (10µM) and 2.5 µl of sample library. The amplification was carried out under the following thermal cycling conditions: 30 seconds at 98°C; 10 cycles of 20 seconds at 98°C, 30 seconds at 60°C, and 40 seconds at 72°C; with a final extension of 5 minutes at 72°C. The resulting amplification product was purified as described above.
All amplification reactions were quantified using a Qubit 3.0 Fluorometer, and additionally the libraries were quantified using an Agilent 2100 Bioanalyzer High Sensitivity DNA kit prior to pooling them equimolarly for next-generation sequencing (NGS) purposes. For the adult, four dual-indexed libraries were sent for 100-bp (base-pair) single-end sequencing on a HiSeq2500 (NBAF Liverpool, United Kingdom), whereas the infant remains were assigned one single-indexed library each, which were sent for 100-bp paired-end sequencing on a HiSeq4000 (Macrogen, South Korea).

5.3.3 Bioinformatic analysis

5.3.3.1 Single-end sequencing analysis

The NGS reads were trimmed and aligned to the human mitochondrial genome (rCRS, NC_012920.1; Andrews et al. 1999) and the human reference genome (UCSC hg19). The reads were trimmed using cutadapt (v.1.13; Martin 2011), allowing a minimum overlap of 1 bp between read and adapter (adapter sequence: AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC), discarding

reads shorter than 30 bp. BWA aln (v.0.7.12-r1039; Li et al. 2009a) was used to map reads to both the rCRS and hg19, filtering by base quality 15, and disabling seed length, as recommended for ancient DNA data (Schubert et al., 2012), with a maximum edit distance of 0.01 and maximum number of gap opens set to 2. Samtools version 1.3 was used to sort and filter reads, and to remove PCR duplicates, and DNA damage patterns were assessed using mapDamage 2.0 (Jónsson et al., 2013) and bamdamage (Malaspinas et al., 2014). Picard tools (v.2.18, `http://broadinstitute.github.io/picard`) was used to add read groups to the sequencing reads aligned to hg19, and the sequencing quality was assessed using FastQC (version 0.11.5; Andrews 2016) and qualimap (v.2.2.1; Okonechnikov et al. 2016). Samtools mpileup and bcftools (v.1.3) were used to create a consensus FASTQ file of the alignment to the rCRS, and SEQTK (v.1.2-r102-dirty) was used to convert the FASTQ file into FASTA format. Schmutzi (Renaud et al., 2015) was used to detect possible modern mitochondrial contamination, and mitochondrial haplotypes were assigned using Haplogrep 2 (Weissensteiner et al., 2016).

5.3.3.2 Paired-end sequencing analysis

The only difference in the treatment of the paired-end reads compared to single-end reads was the use of AdapterRemoval (v.2.1.7; Schubert et al. 2016), which trimmed the adapters from the read pairs and merged the reads.

5.3.3.3 Mitochondrial haplotype calling

The consensus sequence of the mitochondrial genome was created using samtools mpileup and bcftools; additionally GATK (v.3.7) was used to call the variants against the rCRS for the two infant samples. As the coverage of the foetus and perinate mtDNA genomes was very low, this approach enabled identification of as many polymorphisms from the mitochondrial genome as possible. All other bioinformatic steps were the same for both read types.

5.3.3.4 Phylogenetic tree reconstruction

All available published data from complete mitochondrial genomes was used to reconstruct the phylogenetic tree of the determined haplogroup, as well as data generated as part of a larger British Isles study. MtPhyl v4.015 (Eltsov et

al. 2011; `https://sites.google.com/site/mtphyl/home`) was used to reconstruct the phylogeny, applying the discrete-character method maximum parsimony. The reconstructed tree was reviewed based on mtDNA PhyloTree build 17 (`http://www.phylotree.org/tree/index.htm`).

5.4 Results

DNA was successfully extracted and sequenced from all three samples. The endogenous DNA content of the adult, perinate and foetus was quite low, with 11.89%, 8.01% and 7.57%, respectively. The genome of the adult female was sequenced to a coverage of 0.276x, while her mitochondrial genome was sequenced to a coverage of 58.068x. For the two infants, the genome coverage was 0.014x for the perinate and 0.006x for the foetus, with mitochondrial genome coverages of 4.140x and 2.241x, respectively. Due to the low coverage of the genomes of the foetus and the perinate and the extremely low number of retrieved overlapping SNPs (<10,000), no additional autosomal analysis was possible.

5.4.1 Contamination estimates

Contamination estimates ranged from 0.165 to 0.185% for the adult female and 0 to 0.005% for both infants. No exogenous contamination was observed in extraction or library blanks, supporting the authenticity of the results, and there were no matches between the mitotypes from the three ancient samples and those of the researchers.

5.4.2 Genetic sex identification

The genetic sex was identified using the tool published by Skoglund et al. (2013). The osteological determination that the adult was female was supported by the genetic sex identification. No sex could be assigned for the two infants due to low genomic coverage.

5.4.3 Mitochondrial haplotype determination

The mitochondrial haplotypes of the individuals show that only one of the infants shares the same haplogroup as the adult female. Due to gaps and low coverage in the foetal mitochondrial genome, the haplotypes do not fully match. However, the female and the foetus both belong to haplogroup H7a1b

(Table 5.1), whereas the perinate was assigned to haplogroup J1c3 (Table 5.2). To securely determine mitochondrial haplotypes for the infants, two bioinformatic approaches (that is, GATK Variant Call versus calling the consensus sequence, as discussed in the methods section) were compared, and the same haplotypes were assigned using both methods. Table 5.1 shows the called mutations for the adult female and her putative child, the foetal individual. Mutations that are likely due to deamination, polymorphisms or 'unspecified' are highlighted in Table 5.3. A more thorough overview of the rationale behind every disregarded mutation can the found in Appendix J.1.

TABLE 5.1: Mutations defining haplotype H7a1b, and their presence or absence in the adult female and foetus. N marks the positions that were not covered in the sequence. Dots indicate similarity to reference sequence.

	263	750	1393	1438	1719	4769	4793	8860	15326	16261	16519
rCRS	G	G	A	G	A	G	G	G	G	T	T
adult female	.	.	.	.	.	.	.	.	.	.	.
foetus	.	.	.	N	.	.	A	.	N	Y	.

TABLE 5.2: Mutations defining haplotype J1c3 and their presence in the perinate.

	185	228	295	462	489	3010	10398	12612	13708	13934	14798	16069
rCRS	G	G	C	C	T	G	A	A	G	C	T	C
diagnostic for J1c3	A	A	T	T	C	A	G	G	G	T	C	T
Perinate	A	A	T	T	C	A	G	G	G	T	C	T

TABLE 5.3: Comparison of scored mutations of all three mitochondrial genomes. Italic: changes vs reference up to the H2 root. Bolded: H7 to H7a1b diagnostic mutations and J1 to J1c3 diagnostic mutations.

Sample	Haplogroup	Scored mutations	Deamination	Gaps
Adult female	H7a1b	*263G 750G* **1393A** *1438G* **1719A** *4769G* **4793G** *8860G 15326G* **16261T** 16519C	– –	–
Foetus	H7a1b	*263G 750G* **1393A 1719A** *4769G 8860G* **16261Y** 16519C	633M 911K 5239K 6272R 6359M 6842Y 7022K 7025R 8576K 10674K 10750R 11629R 12308R 12766Y 12822R 12864K 12873W 13207W 14864K 15601K	1438G 15326G
Perinate	J1c3	*73G* **185A 228A** *263G* **295T 462T 489C** *750G 2706G* **3010A 10398G** *11719A* **12612G 13708A 13934T** *14766T* **14798C** *15326G 15452A* **16069T** *16126C*	152Y 539Y 3276R 4892M 6243R 9233Y 9477R 9814Y 10567W 11251R 12210W 12835R 12998Y 15854W 16022Y	–

5.5 Discussion

5.5.1 Novel approach

This is the first time that complete ancient mitochondrial DNA (mtDNA) sequences from foetal-aged bone fragments have been reported. Previous studies have used PCR-based methods to determine the genetic sex of neonates (Faerman et al., 1998; Waldron et al., 1999; Irish et al., 2008), but no mitochondrial genomes or genome-wide data have been published to date.
Due to the low genome coverage retrieved from the infants, it was not possible to determine the molecular sex of either the perinate or the foetus, and no genome-wide analysis could be performed. It is possible that a PCR-based approach, such as in Palomo-Díez et al. (2018) might have succeeded, but the aim of this study was to investigate the relationship between the three individuals, and shot-gun sequencing promised the best approach to cover as much of the mitochondrial genome as possible. Based on previous sequencing results of ancient samples, it is known that the mitochondrial genome can be sequenced to high coverage as a byproduct of sequencing the entire genome, even if the nuclear genome sequencing is not covered well enough to allow for WGA. Sequencing the whole mitochondrial genome results in a much more detailed analysis of the haplotype than a PCR-based approach where typically only fragments of the non-coding region are amplified.

5.5.2 Iron Age burials of infants

The separation of infant burials from the rest of the population appears to have been widespread across the Iron Age and into the Romano-British period (Redfern et al., 2011). At some sites, this separation is thought to be a special treatment involving age-specific funerary rites (Millett et al., 2015), while at others it appears to be due to deposition without any ceremony (e.g. Saville et al. 1994; Finlay 2000; Armit et al. 2007). In the British Iron Age, women that died in childbirth, as well as foetuses and young infants, were given special burials, highlighting the importance of the link between woman and child (Millett et al., 2015). Human remains of infants or foetuses were often placed into cave systems, and the number and spread of cave burials during this period in Scotland is relatively high – especially given that burials from this period are generally not abundant (Saville et al., 1994). There is also evidence of infants being buried in unstable, abandoned buildings or disused settlement sites. For example, an excavation at the site of

Howe in Orkney revealed an Iron Age burial of an adult male, aged between 35 and 45 years, along with a perinate and an infant, all of whom had been placed within accumulated rubble in a midden in a disused yard (Armit et al., 2007). The bones showed gnaw marks, and their layout was interpreted as an informal burial; that is, they were deposited without ceremony (Armit et al., 2007). Another example is the Iron Age site at the Knowe of Skea, also in Orkney, where over one hundred, mainly infant, bodies were found deposited in domestic rubble contexts. Cremations and inhumations of adult women with foetal/neonatal remains in reused, separate burial areas are also known from Ireland (Finlay, 2000).

5.5.3 High Pasture Cave context

The context of the High Pasture Cave appears different to those described above, as a domestic setting is not indicated by any of the structural features (Livarda et al., 2017). The trauma evident on the female, and the lack of any wider tradition of inhumation in the region, has led to the suggestion that this burial was a deviant one, or even a sacrifice connected to the closure of the cave stairwell, rather than a ceremonial burial of a mother and her child (Armit, in press). The burial raises several questions, especially now that the genetic results show that only the foetus was likely to be the female's child, but not the perinate.

5.5.4 Maternal lineages

Mitochondrial genomes are used to trace the female line of descent, as they are passed on from mother to child without recombination. The only changes that occur over generations are naturally-occurring mutations, and these can be used for phylogenetic reconstructions over wide geographic areas and long timespans. The main defining mutations for haplogroup H7a1b can be found in both adult and foetus (bolded in Table 5.1); i.e. the transitions at positions 1719, 16261 and 1393 which define H7a, H7a1 and H7a1b, respectively. The read at the defining mutation for H7 at position 4793 shows the ancestral, instead of the derived, state in the foetus, and the majority of the mutations in this mitochondrial genome are most likely due to deamination (Dabney et al., 2013) or environmental factors, such as radiation and alkylation (Willerslev et al., 2005). These positions were, therefore, counted as being less significant when determining the haplogroup (Appendix J.1). Haplogroup H occurs at ∼44% of all British mtDNA lineages, but haplogroup H7

is a minor European subclade found at ∼1.6% in Britain, and its nested subclade H7a1b occurs at only ∼0.3% (Frequencies of H7a1b are given in Table 5.4, and additional information on haplogroup H7 can be found in Appendix C).

TABLE 5.4: Frequency table of H7a1b in Europe based on complete sequencing data and SNP capture data.

	Number of complete sequences	Frequency based on complete sequences in percent	Number of SNP capture based haplotypes	Frequency based on SNP capture in percent
England	1	11.1	4	22.2
Scotland	2	22.2	13	72.2
Ireland	-	-	1	5.5
Denmark	4	44.4	-	-
Finland	1	11.1	-	-
Sardinia	1	11.1	-	-

5.5.4.1 Haplogroup H7a1

In the database of extant whole-mitogenome lineages, H7 is commonly found in Mediterranean Europe, whereas H7a is more focused on north-western Europe (Britain, n = 10; Denmark, n = 16; Germany, n = 13), with some individuals in the central Mediterranean and others as far east as Finland and Russia (Behar et al., 2012a; Li et al., 2014; Raule et al., 2014; Olivieri et al., 2017; Dulias, 2019). Its major subclade, H7a1, dates to ∼6.4 ka based on maximum likelihood calculations. The earliest instance of H7 discovered so far is from the Early Neolithic *impressa* culture of Croatia, dating to ∼7.6 ka (Mathieson et al., 2018), while a Middle Neolithic individual from Hungary dates to ∼7.1 ka (Lipson et al., 2017). The earliest H7a1 individual known is from the Balkans Eneolithic Varna culture, dating to ∼6.6 ka. Various H7 lineages have also been seen from the Late Neolithic, through the Corded Ware and Bronze Age of central Europe (Mathieson et al., 2018; Olalde et al., 2018). The overall distribution pattern suggests that H7a lineages may have been dispersed north from the Mediterranean with the spread of the Neolithic. The subhaplogroup of H7a1b dates to ∼4.5 ka and is seen in modern-day Scotland (Dulias, 2019), England (Behar et al., 2012a) and Denmark (Behar et al., 2012a; Li et al., 2014; Raule et al., 2014), as well as Finland and Sardinia, but the individuals from High Pasture Cave are the earliest known H7 lineages from north-west Europe (Figure 5.4).

Given the geographic diversity and young age of H7a1b, it may have reached Britain more recently than the settlers carrying the Bell Beaker culture, when the genome-wide pattern of the British gene pool began to take on its modern form – possibly in the Bronze Age or in the Iron Age itself.

5.5.4.2 Haplogroup J1c

In contrast, the perinate belongs to haplogroup J1c (subclade J1c3). Haplogroup J1c is an almost entirely European subclade of haplogroup J, dates to ∼17 ka, and is the most frequent subclade of J, with J1c3 dating to ∼11.5 ka (Pala et al., 2012; Pereira et al., 2017). The mitochondrial genome of the perinate had better coverage than the foetus, and all of the defining positions of J1c3 subclades were covered. As none of the defining mutations of any of the J1c3 subclades could be observed, we can securely say that the mitogenome of this individual is basal to J1c3. Over 100 ancient J sequences have been published to date, of which >60% fall within J1c and are mainly found in the Neolithic and Bronze Age in Germany, the British Isles and Hungary (Gamba et al., 2014; King et al., 2014; Allentoft et al., 2015; Günther et al., 2015; Haak et al., 2015; Mathieson et al., 2015; Cassidy et al., 2016; Hofmanová et al., 2016; Lazaridis et al., 2016; Martiniano et al., 2016; Juras et al., 2017; Knipper et al., 2017; Lipson et al., 2017; Margaryan et al., 2017; Neparáczki et al., 2017; Saag et al., 2017; Schuenemann et al., 2017; Brace et al., 2018; Mathieson et al., 2018; Mittnik et al., 2018; Olalde et al., 2018). J1c has been present in the British Isles since the Early Neolithic (Olalde et al., 2018), although the earliest J1c3 individuals to date are both post-Beaker: a Chalcolithic/Early Bronze Age individual from Kent (Olalde et al., 2018) and one Roman individual from Yorkshire (Martiniano et al., 2016). The earliest J1c3 individual was found in Early Neolithic Spain (Haak et al., 2015), but it was present as far north and east as the Ukraine by 4.8 ka, prior to the spread of the Corded Ware (Mathieson et al., 2018). Thus, although the Iron Age lineage might be a survival from the earlier Neolithic in the British Isles, it could also potentially trace its ancestry to the rise of the Beaker culture on the continent, and its time of arrival in Britain is not clear.

5.6 Conclusions

The results of this study show that only one of the two infants buried alongside the adult female was possibly her child, whereas the perinate did not share the maternal lineage with her. This is an interesting find considering the common archaeological understanding of these type of female/foetus burials, where the children are assumed to be related to the women. So far archaeological interpretations of these burials have indicated either a mother-child relation or a midwife buried in the same place as foetuses and infants. In the case of the High Pasture Cave, where the pregnant woman was likely

crushed to death, archaeological questions still remain, such as why she was killed and why she was buried with a likely unrelated child? The most favoured explanation is that she was interred as part of a sacrificial closing burial. The genetic analysis showed that a general assumption of mother-child relations within such burials can be misleading and that genetic analysis can assist quite drastically the analysis and interpretation of archaeological contexts.

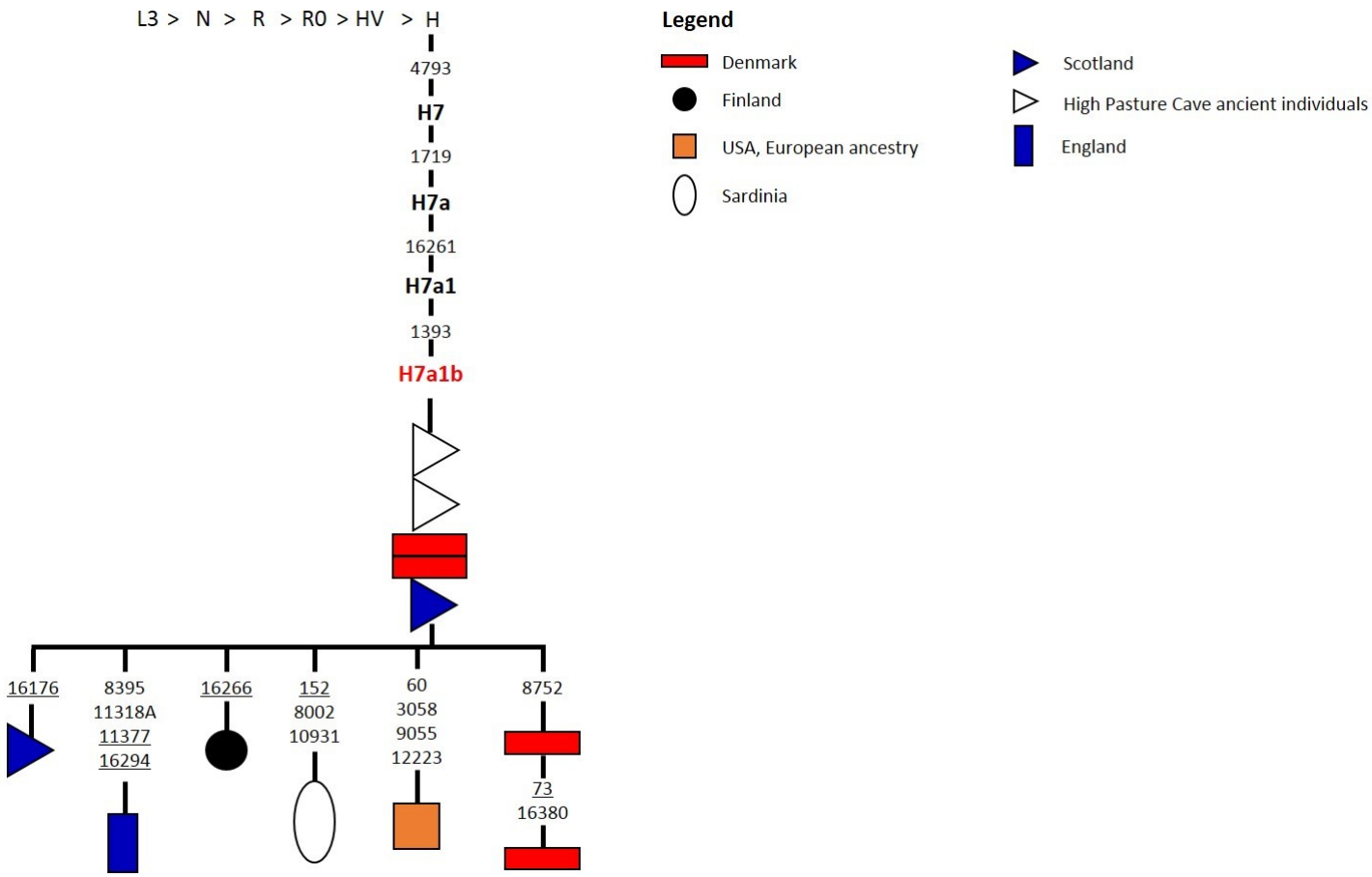

FIGURE 5.4: Schematic of the branch H7a1b of the mitochondrial DNA phylogenetic tree of haplogroup H7. The legend shows the origin of the samples. Recurrent mutations are underlined, transversions are indicated as postfixes. Scottish ancient samples from the High Pasture Cave are indicated as blank triangles.

Chapter 6

Rosemarkie Man - an unusual local or a foreigner in Pictish Scotland?

6.1 Introduction

Since 2006, the Rosemarkie Cave Project has been investigating the archaeology of 19 caves in the southeast of the Black Isle, Scotland (Figure 6.1). First results indicate that some caves were visited or occupied during the 7th-9th centuries AD. Further open-area excavations of this site uncovered iron-working activity and a *c.* 5th-7th century male inhumation burial.
Based on the archaeological layers, it is likely that the iron-working took place before or during the early medieval period and the cave continued to be used throughout the medieval and post-medieval periods (Birch et al., 2017). Radiocarbon dates of animal bones from across the site gave an overlapping time frame of 600-941 AD. Later investigations provided radiocarbon dates from archaeological sequences of worked bone and antler from the 2nd-3rd century AD. The male inhumation burial was found below a post-medieval cobbled floor and midden deposits in a dark alcove, and comprised the well-preserved remains of a young male (pers comm. Prof. Sue Black) who had suffered severe, multiple trauma to the head. The grave was unmarked, with stones weighting down the limbs, and butchered animal bones were placed over the position of the head (Birch et al., 2017). The burial is quite remarkable as its position, and the condition of the individual at the time of death, is unusual. The skeleton was aligned in a northwest-southeast position with the skull pointing to the southeast. The arms were by the side of the torso, while the lower limbs/feet were in a "cross-legged" position (Figure 6.2).
Forensic analysis showed that the individual was very healthy and of a robust build. He died due to multiple trauma wounds to his skull, caused by more than one weapon (Birch et al., 2017). A fragment of the left rib was taken for radiocarbon analysis and provided a date of *cal*. AD 430-631

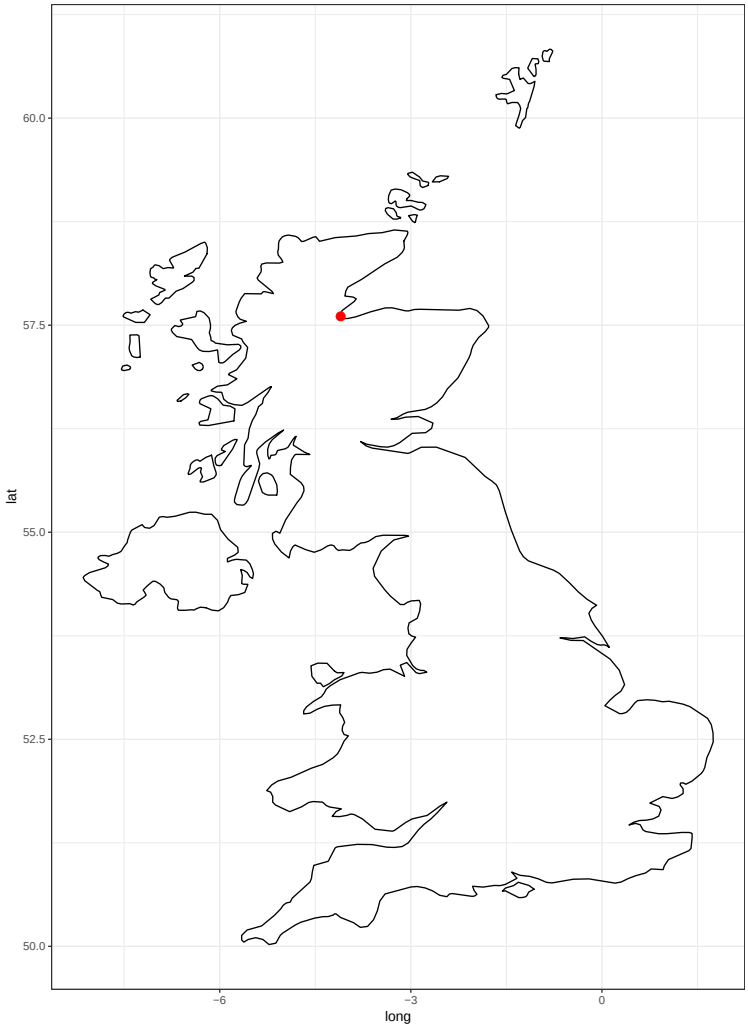

FIGURE 6.1: Location of the Rosemarkie Cave system within the British Isles (adapted from Birch et al. 2017).

(SUERC-70721). Strontium isotopic analysis gave a mean result of 0.70995 $^{87}Sr/^{86}Sr$ from two maxillary teeth, which fits with local values. Results of oxygen isotopic analysis are not available yet. Based on the unusual position and setting of the burial, it indicates a deliberate interment.

The stones that were placed on top of the body may represent an attempt to hold down the body after death, with the animal bones representing an offering, which may indicate a ritual act. The selection of the cave may have formed an important part of the burial process (Birch et al., 2017). Given the date range of the early 5th to 7th centuries, the burial could potentially represent Pictish, Viking, early Christian or other influencing factors, due to the many cultural influences and developments of that time in this region (Birch et al., 2017). Further to the forensic analysis of the burial, facial reconstruction has been undertaken (see Figure 6.3) (Birch et al., 2017).

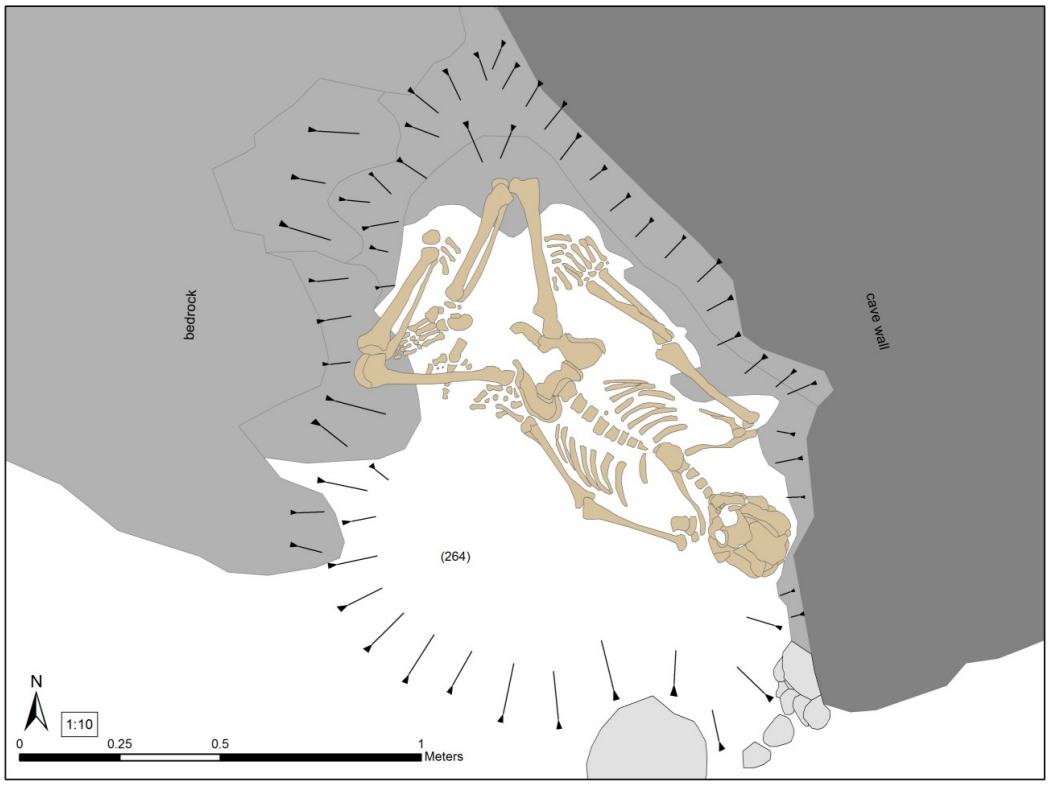

FIGURE 6.2: Schematic of the analysed inhumation after excavation (adapted from Birch et al. 2017).

6.2 Materials and Methods

6.2.1 Ancient sample processing

Ancient DNA sample processing was done at the Ancient DNA Facility of the University of Huddersfield (England) and the Ancient DNA lab, Smurfit Institute, Trinity College Dublin (Ireland), under dedicated clean-room conditions. The same protocol for sample processing and ancient DNA extraction was used in both labs.

6.2.2 Sampling and DNA extraction

The petrous bone of the Rosemarkie skull was detached from the intact cranium using a cutting saw attached to a hobby drill at the University of Huddersfield and shipped to the Ancient DNA lab at Trinity College Dublin (TCD), where the surface of the bone was decontaminated by UV radiation for 10 minutes on each side of each sample. The surface of the petrous was removed using a standard drill bit, followed by the excision of a piece of the densest petrous parts using a circular saw (Pinhasi et al., 2015). Bone

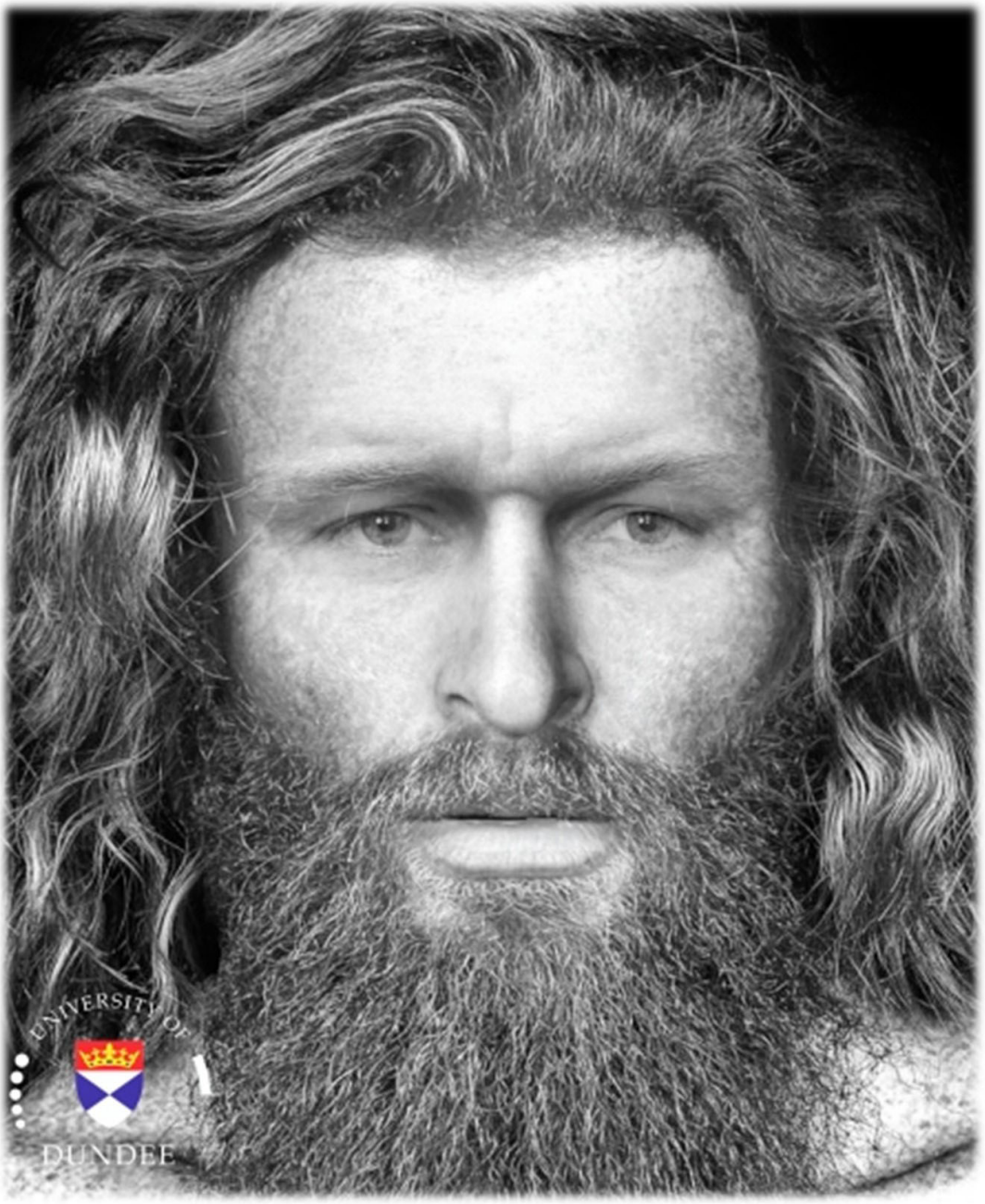

FIGURE 6.3: Facial reconstruction of the Rosemarkie individual (reproduced with permission of Prof. Dame S. Black).

powder was obtained by crushing the excised bone pieces with a Mixer-Mill (Retsch MM400). DNA was extracted from approximately 150 mg of the sample powder produced, following the protocol by Yang et al. (1998) with modifications by MacHugh et al. (2000). Blank controls were included throughout extractions to allow for quatification of contamination. Further sampling for bone powder was done following the exact same protocol at the Ancient DNA Facility of the University of Huddersfield.

6.2.3 Library preparation and sequencing

6.2.3.1 Library preparation

Library preparation followed the protocol described in Chapter 2. For the Rosemarkie sample one single-indexed library was prepared at TCD. Five further libraries were prepared at the University of Huddersfield. Three libraries were dual-indexed and two libraries were single-indexed. The same protocol was used in both labs, with the only difference being the single- and dual-indexing.
Amplification reactions performed at TCD were quantified using an Agilent 2200 TapeStation and pooled equimolarly. Amplification reactions performed at the University of Huddersfield were quantified using an Agilent 2100 Bioanalyzer High Sensitivity DNA kit and pooled equimolarly.

6.2.3.2 Sequencing

The NGS library prepared at TCD was sequenced in an Illumina HiSeq 2000 (single-index, 100 base pair, single-end sequencing; Macrogen, South Korea), library code KaD1. NGS libraries prepared at the University of Huddersfield were sequenced in an Illumina HiSeq 2000 (dual-index, 100 base pair, single-end sequencing; NBAF Liverpool, United Kingdom), library codes KD001.1, and KD001.2, and in an Illumina HiSeq 4000 (single-index, 100 base pair, paired-end sequencing; Macrogen, South Korea), library codes K23 and K34. Dual-indexing was due to requirements of NBAF Liverpool and single-end versus paired-end sequencing differed at Huddersfield between the two sequencing companies.

6.2.4 Data processing and read mapping

The NGS reads of libraries KaD1, KD001.1, and KD001.2 (all single-end sequenced) were trimmed using cutadapt version 1.13 (Martin, 2011), allowing a minimum overlap of 1 bp between read and adapter (adapter sequence: AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC), and discarding reads shorter than 30 bp. The NGS reads of libraries K23 and K34 (paired-end sequenced) were trimmed using AdapterRemoval version 2.1.7 (Schubert et al., 2016), to remove adapters from read pairs, trim Ns, remove low quality bases from the reads and merge reads (`-trimns -trimqualities -collapse -adapter1` AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC `-adapter2` AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT). Alignments were run

per library as described in Chapter 2, with the additional step of merging the separate alignments at the bam stage with `samtools merge`. After merging, the sample ID was KD001. All further analyses, such as sexing, analysis of uniparental marker, PCA and ADMIXTURE were run as described in Chapter 2. For the genome-wide analysis the previously described adult individual from High Pasture Cave (Chapter 5) was included, as it dates to the same period and was found geographically close to the Rosemarkie individual.

6.3 Results

6.3.1 Sequencing results

Sequencing results are displayed in Tables 6.1 and 6.2. Information on the negative controls can be found in Appendix D.

TABLE 6.1: Sequencing results of all combined libraries sequenced of the Rosemarkie individual. 'SNPs' denotes the number of SNPs called with the published datasets as reference panel.

Sample ID	Read count	Endogenous content in %	Duplication rate in %	Genome coverage	SNPs
KD001	3,652,867-105,144,736	60.91-90.05	0.68	0.7519x	544,787
KD005	10,040,467-64,122,526	11.87-13.25	1.02	0.276x	223,914

TABLE 6.2: Read count and endogenous contents of KD001 and KD005 libraries. Indexing type indicates whether a library was single or dual indexed. Read type indicates single-end (SE) and paired-end (PE) sequencing.

Lab ID	Read count	Endogenous content in %	Indexing type	Read type
		KD001		
KaD1	22,461,152	87.48	single	SE
KD001.1	11,790,087	89.99	dual	PE
KD001.2	3,652,867	90.05	dual	PE
K23	77,619,946	60.91	single	PE
K34	105,144,736	83.28	single	PE
		KD005		
KD005.1	10,040,467	13.25	dual	PE
KD005.2	64,122,526	11.87	dual	PE
KD005.3	18,991,631	13.16	dual	PE
KD005.4	20,104,930	12.40	dual	PE

6.3.2 Mitochondrial contamination estimates

In order to detect possible modern mitochondrial contamination, schmutzi (Renaud et al., 2015) was used. Estimates of mitochondrial contamination ranged from 0.24-0.35%, showing very little modern mitochondrial contamination.

6.3.3 Genetic sex determination

To determine the sex of the analysed ancient sample, software published by Skoglund et al. (2013) was used, which calculates the ratio of reads aligning to the X and Y chromosome. The individual studied was determined as male, supporting the results of the osteological analysis.

6.3.4 Mitochondrial genome analysis

The haplotype of the presented ancient sample was determined aligning the sample sequence reads to the revised Cambridge Reference Sequence (rCRS; NC_012920.1; Andrews et al. (1999)) using BWA filtering (with -b and -q 30), and duplicate reads were removed. The mtDNA consensus sequence was called using Samtools mpileup, as in Jones et al. (2015), Cassidy et al. (2016), and Martiniano et al. (2016). Haplofind (Vianello et al. 2013) was used to identify haplogroup defining mutations and assign the haplogroup of the sample (Results in Table 6.3). The mtDNA coverage was estimated using samtools. The individual was found to have all the defining mutations of haplogroup J1b1a1a.

TABLE 6.3: Mitochondrial DNA haplogroup and mutations identified

Sample	Haplogroup	Average coverage	Polymorphisms
KD001	J1b1a1a	96.2727x	73G 263G 462T 489C 750G 1438G 2706G 3010A 4216C 4769G 7028T 8557A 8860G 10398G 11251G 11719A 12007A 13708A 13879C 14766T 15326G 15452A 16069T 16126C 16145A 16172C 16192T 16222T 16261T

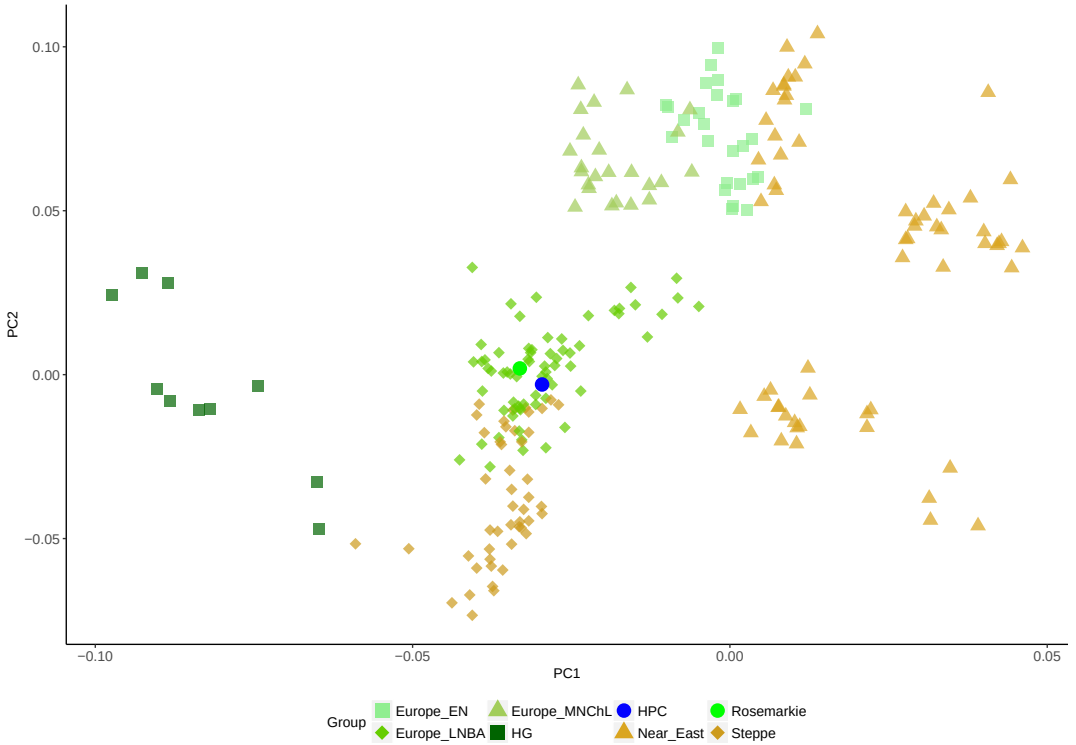

FIGURE 6.4: Principal component analysis of the Rosemarkie and High Pasture Cave individuals, with previously published ancient samples of Lazaridis et al. (2016) projected onto the first two principal components.

6.3.5 Y-chromosome analysis

The Y-chromosome analysis gave R1b1a1a2a1a as the predicted haplogroup with L151 and L11 being the most downstream predictive markers. For more details see Appendix G Table G.14.

6.3.6 Genome-wide analysis

For the genome-wide analysis, the Rosemarkie individual was analysed alongside the Pictish adult female from the High Pasture Cave, Isle of Skye (see Chapter 5 for more details), and the dataset of modern and ancient individuals used in Lazaridis et al. (2016). The individual from the High Pasture Cave had 223,914 overlapping SNPs with the published dataset.

6.3.6.1 Principal component analysis

The PCA of the two individuals from Rosemarkie and High Pasture Cave shows them plotting with the European Late Neolithic/Bronze Age samples (Figure 6.4).

6.3.6.2 ADMIXTURE

The ADMIXTURE plot is given in Figure 6.5. Further information on the ADMIXTURE analysis can be found in Appendix H. Both, the Rosemarkie Cave and High Pasture Cave, individuals appear most similar to the Iron Age individuals from Knowe of Skea, Orkney.

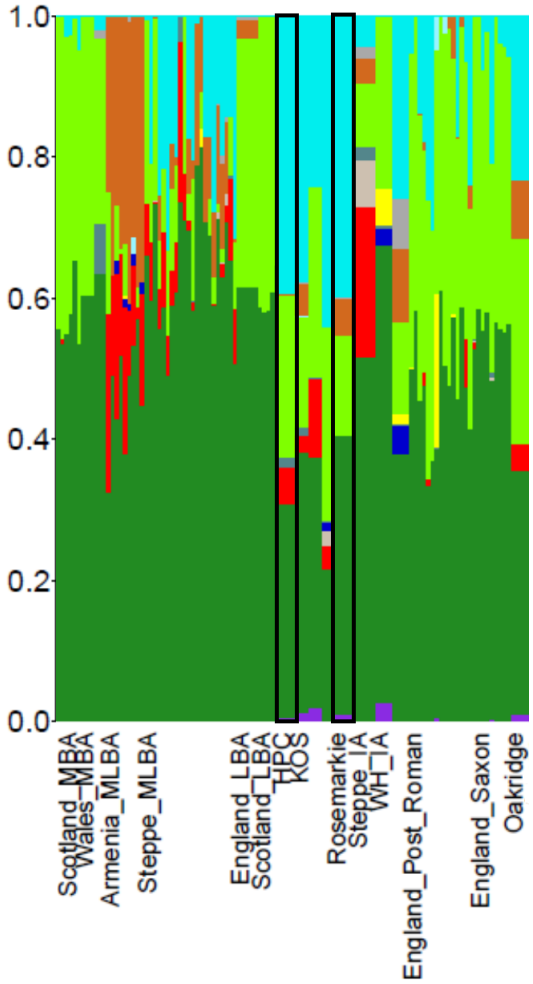

FIGURE 6.5: ADMIXTURE analysis of the Rosemarkie and High Pasture Cave individuals, with previously published ancient samples of Lazaridis et al. (2016) and Olalde et al. (2018) for K=12. Abbreviations used: E: Early, M: Middle, L: Late, BA: Bronze Age, IA: Iron Age, WH: West Heslerton, HPC: High Pasture Cave, KOS: Knowe of Skea.

6.4 Discussion

6.4.1 Genome coverage

The genome coverage of the Rosemarkie individual was >0.7% and allowed genome-wide analysis with published modern and ancient genomes, as well as a detailed analysis of his uniparental markers. The mitochondrial genome coverage was close to 100x and therefore, the complete mtDNA haplotype was recovered and the individual was securely determined to belong to haplogroup J1b1a1a, which is almost restricted to the British Isles (and Norway).

6.4.2 Maternal lineage

Mitochondrial haplogroup J originated in the Fertile Crescent of the Near East (Pala et al., 2012). Based on its geographic distribution haplogroup J likely diversified in the Near East around the time of the LGM. Subhaplogroups of J are nowadays present in both Near East and Europe, with J1b, J1c and J2 being mainly Near Eastern and J1c solely European. J1b is predominantly Near Eastern, except for J1b1a, which is considered European (Pala et al., 2012). J lineages are thought to have entered Europe during the Late Glacial period, prior to the Neolithic in Europe. The maternal lineage of the Rosemarkie individual is, therefore, one of the older European haplogroups. Another analysed contemporaneous male from Milla Skerra at Sandwick on Unst, Shetland Islands, was determined to belong to subhaplogroup J1b1a, which is basal to the clade the Rosemarkie individual sits in. Unfortunately, the autosomal genome coverage of this individual was very low and no Y-chromosome haplogroup could be predicted and no genome-wide analysis was possible for the same reason (for supplementary information on the Shetland individual please refer to Appendix K).

6.4.3 Paternal lineage

Within Europe, typically >50% of men belong to Y-chromosome haplogroup R. The most common subclade within R in Europe is R1, which shows a distinctive geographic distribution, where R1a is essentially eastern and R1b western European (Myres et al., 2011). The majority of western European men belong to Y-chromosome haplogroup R1b, as it is the most common male lineage cluster in this region, with highest frequencies of almost 100% in Ireland and the Basque country. Several studies have linked R1b to the

spread of the Corded Ware culture (Silva et al., 2019). During the Late Neolithic and Bronze Age, common Y-chromosome lineages in early European farmers almost disappear and were largely replaced by R1b in western Europe (Haak et al., 2015).

6.4.4 Genome-wide analysis

As the Rosemarkie Cave, and the High Pasture Cave (discussed in Chapter 5) are both in Scotland, and excavated individuals in both caves belong to the Scottish Pictish period and the Iron Age, their genomic data was analysed alongside each other. The PCA of the autosomal data showed that both the Rosemarkie and the High Pasture Cave individuals cluster closest with Late Neolithic/Bronze Age individuals from Europe (Figure 6.4). ADMIXTURE analysis showed a more nuanced picture (Figure 6.5). Both individuals appear most similar to the Iron Age individuals from Knowe of Skea, Orkney. A comparison of the genomes of the Rosemarkie and High Pasture Cave individuals with the genome of the Milla Skerra individual would be very interesting and might shed light on the population movements and admixture events across Scotland and the Scottish Isles, but further sequencing of the Milla Skerra human is required. The availability of more Iron Age individuals from the British Isles would further facilitate this analysis.
Because the burial of the Rosemarkie man was peculiar, it raised several questions, such as whether the man was a Scottish local or a foreigner. Genome-wide analysis showed that the individual was a local and that his terrible death cannot be explained by him being a foreigner.

Chapter 7

The Post-Roman cemetery at Worth Matravers, Dorset

7.1 Introduction

The archaeological site of Worth Matravers is located in Dorset (Figure 7.1). The area around Worth Matravers shows human occupation from the Neolithic/Early Bronze Age period to the late 8th century AD (Ladle, 2018b). One of the sites dating to the post-Roman period is a cemetery that was used for about 100 years during the 7th and early 8th centuries AD.

FIGURE 7.1: The location of Worth Matravers within the British Isles.

The graves in the cemetery are oriented in an east-west direction (Figure 7.2), organised in rows, and they generally lack grave goods, which is consistent with post-Roman/Early Medieval cemeteries in the west of Britain. The rare

find of a limestone anchor used as a head rest in one grave has been recognised in burials from the Romano-British and Early Medieval periods (Ladle, 2018a). The cemetery held 26 individuals in 21 graves, of which 16 were analysed genetically. There is a possibility of close kinship and marital relationships, which are likely reflected by the occurring double and triple burials. The excavated graves are of seven different types, varying from simple earth-cut to complex cist-type, and all except the earth-cut type had varying amounts of limestone incorporated. The differences in burial types may reflect social structures, such as familial bonds or social status. Genetic analysis of the 16 individuals was undertaken to identify possible kin relationships, and dietary stable isotope data was analysed to help identify social differences within the cemetery.

7.2 Materials and Methods

Teeth from 16 individuals excavated in the post-Roman cemetery in Worth Matravers were processed for DNA extraction and dietary stable isotope analysis as described in Chapter 2. DNA extracts were shipped to collaborator David Reich (Harvard Medical School) for enriched 1.2M SNP and mitogenome capture and data processing (Olalde et al., 2018). Kinship analysis was carried out using READ (Relationship Estimation from Ancient DNA) (Kuhn et al., 2018). Genetic sex determination was undertaken using the script by Skoglund et al. (2013), and Y-chromosome haplogroup determination using Yleaf (Ralf et al., 2018). In order to compare the newly generated ancient dataset to modern human populations, known informative SNP positions of the publicly available Lazaridis et al. (2016) dataset were used and SNPs were filtered as described previously (Mathieson et al., 2018; Olalde et al., 2018). Principal Component Analysis was run using smartpca of the EIGENSOFT package (Patterson et al., 2006; Price et al., 2006) and plotted with R version 3.4.4 (R Core Team, 2018). A combined dataset of the post-Roman genotypes, together with the reference panel, was run using ADMIXTURE (Alexander et al., 2009). Calibrated radiocarbon dates in cal. AD from six burials were calibrated using the IntCal13 atmospheric curve (Reimer et al., 2013) in OxCal v.4.3.2 (Ramsey, 2017). Calibrated radiocarbon dates were provided by our collaborator Bob Kenyon (Kruz, 2018). Stable isotope analysis followed as described in Chapter 2.

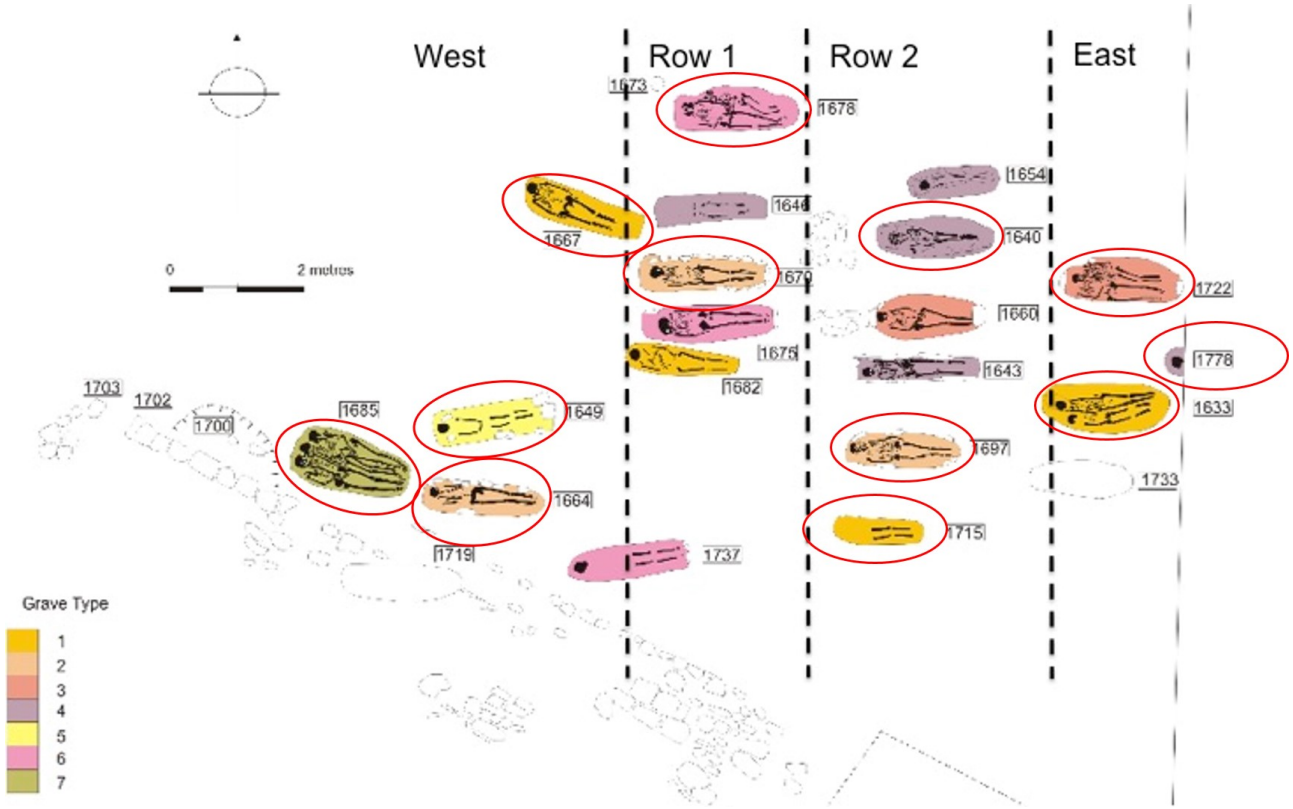

FIGURE 7.2: Schematic of the post-Roman at cemetery Worth Matravers in Dorset, England. Grave types: 1 - plain earth-cut; 2 - rubble lining or kerb alongside; 3 - stone lumps/slabs at head and foot position; 4 - cist-type with split limestone slab walls and occasional lids; 5 - stone lumps at either side of head and irregular arrangement of slabs/rubble at the foot end of the grave; 6 - one vertical stone at either head or foot; 7 - horizontal slabs on top of grave. Dotted lines indicate the distinction between the different rows as West, Row 1, Row 2 and East. Red circles indicate sampled graves. Figure modified from Ladle (2018b).

7.3 Results

All DNA samples passed the quality control and were sequenced successfully. The endogenous content ranged from 0.001% to 0.8%, the coverage ranged from 0.006x to 0.3x, and the overall SNP hits on autosomes ranged from 7,359 to 261,880 (Table 7.1).

TABLE 7.1: Sequencing results of the 1,240K SNP capture of the individuals from the post-Roman cemetery at Worth Matravers. Coverage, number of SNP hits on autosomes and the number of Y-chromosome markers are given. Letters A, B and C indicate different individuals within one burial. 'SNPs' denotes the number of SNPs called with the published datasets as reference panel.

Lab ID	Burial No.	endogenous content in %	genome coverage	SNPs
KD007	1633A	0.0042	0.0779x	88,395
KD008	1640	0.8076	0.0273x	30,870
KD009	1649	0.0185	0.2689x	261,880
KD010	1633B	0.0015	0.0302x	35,653
KD011	1664	0.0237	0.2518x	249,788
KD012	1667	0.0073	0.0455x	53,089
KD013	1670	0.0011	0.0491x	56,904
KD014	1678A	0.0050	0.0615x	70,422
KD015	1678B	0.0014	0.0063x	7,359
KD016	1685A	0.0012	0.0831x	90,738
KD017	1685B	0.0053	0.0727x	82,904
KD018	1685C	0.0026	0.0413x	48,089
KD019	1697	0.0490	0.3041x	290,996
KD020	1715	0.0032	0.0223x	26,417
KD021	1722	0.0079	0.1926x	200,243
KD022	1778	0.0026	0.0214x	25,389

7.3.1 Age range of the cemetery

Radiocarbon dates were obtained from six burials across the cemetery (data provided by Bob Kenyon; analysis carried out by Anthony Kruz at the Scottish Universities Environmental Research Centre (SUERC)). They show that the cemetery was in use for about 100 years. Three different periods were

classified: early phase (∼500 AD), intermediate phase (∼630 AD) and late phase (∼680 AD). One date each falls into the early and late phases, whilst four dates fall within the intermediate phase. The distribution of the phases across the cemetery is shown in Figure 7.3 A. An overview of the radiocarbon dates is given in Table 7.2.

TABLE 7.2: Radiocarbon dates of selected graves from the post-Roman cemetery at Worth Matravers.

Grave ID	Dating code	Uncal. dates in bp	cal. AD (95.4%)	Classified time period
1677	SUERC-61177	1340 ± 31	644 – 766	late phase
1633	SUERC-61178	1393 ± 31	600 – 673	intermediate phase
1640	SUERC-61179	1396 ± 31	599 – 671	intermediate phase
1678	SUERC-61180	1413 ± 31	583 – 665	intermediate phase
1685	SUERC-61182	1425 ± 31	575 – 660	intermediate phase
1697	SUERC-61181	1547 ± 29	424 – 575	early phase

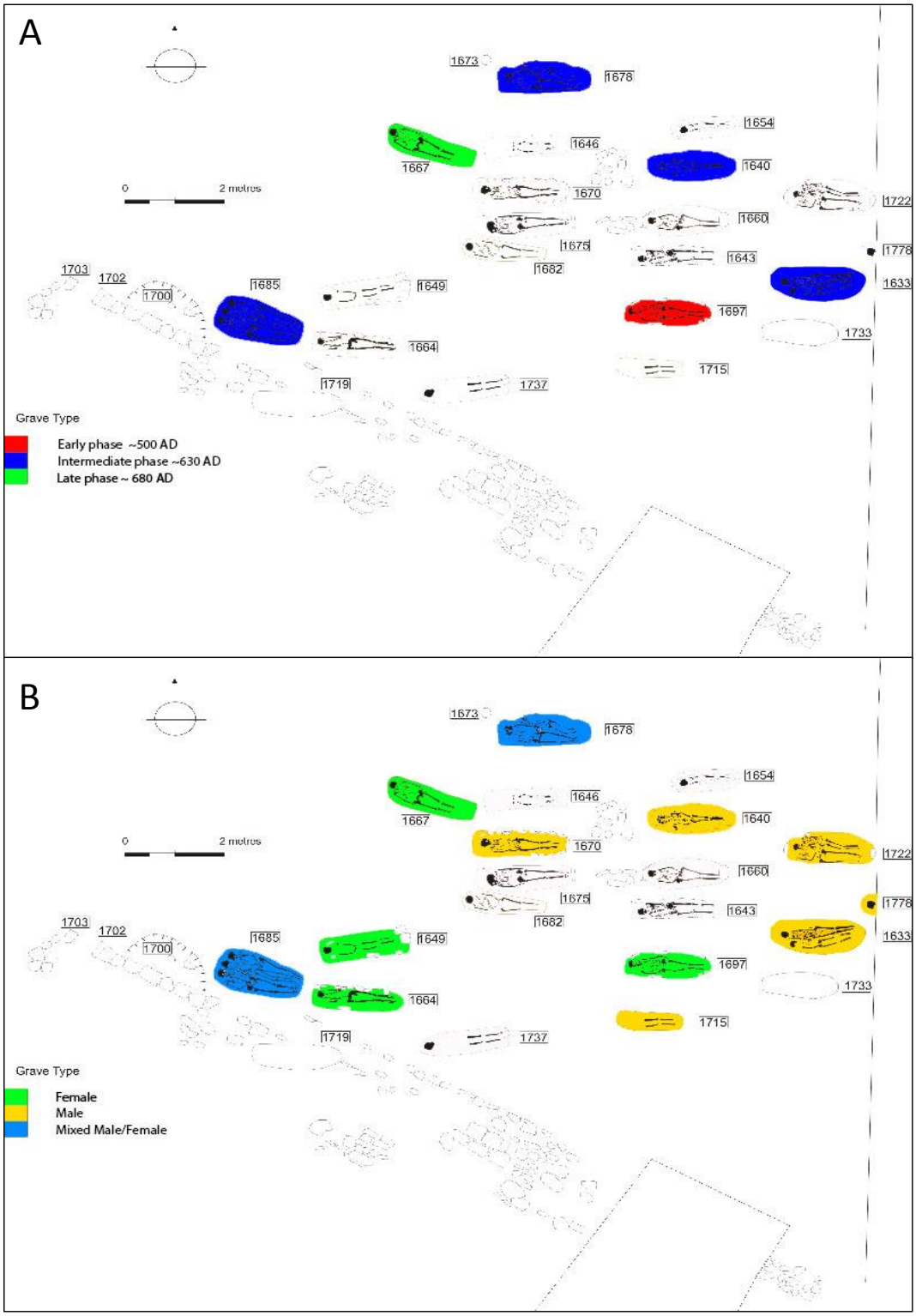

FIGURE 7.3: A - Dated individuals given as rounded mean: early phase ~500 AD; intermediate phase ~630 AD; late phase ~600 AD. B - Genetic sex: green - females (including undetermined, possible female in grave 1667); yellow - males; blue - mixed burials with at least one male and one female.

7.3.2 Dietary stable isotope analysis

One sample did not produce enough tooth powder for both DNA and stable isotope analysis and was, therefore, excluded from the stable isotope analysis. To identify the diet of the remaining 15 individuals from Worth Matravers, dietary stable isotopes were analysed and compared to seven Romano-British sites from Dorset: Alington Avenue (n = 12), Albert Road (n = 5), Gussage All Saints (n = 1), Maiden Castle Road (n = 5), Old Vicarage (n = 6), Poundbury Camp (n = 1), and Tolpuddle Ball (n = 2) (Redfern et al., 2010). The resulting plot is shown in (Figure 7.4).

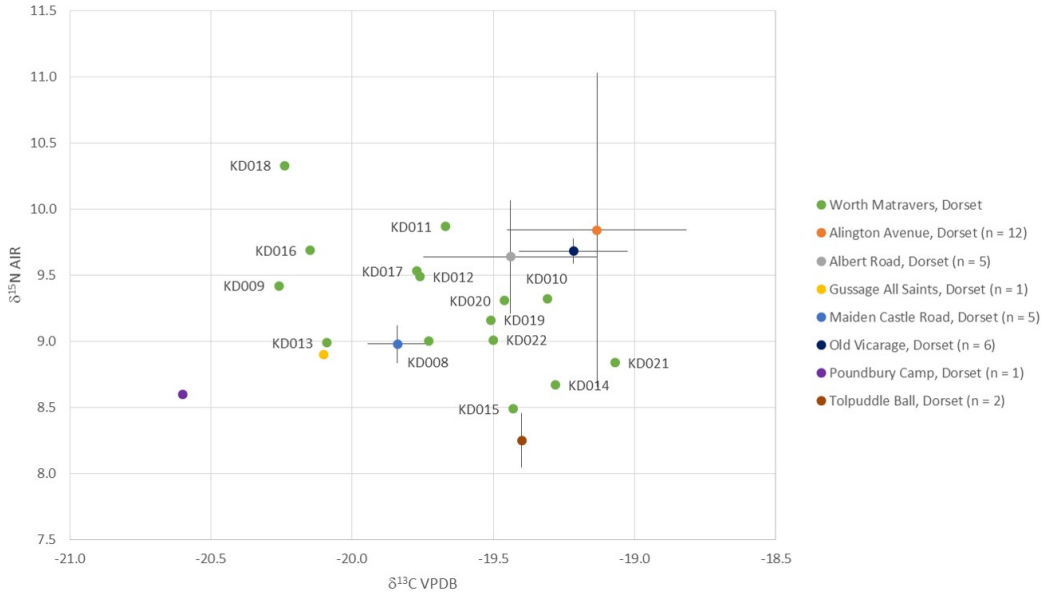

FIGURE 7.4: Dietary stable isotope analysis of the post-Roman individuals from Worth Matravers.

As can been seen, there is a huge variation seen in both carbon and nitrogen values at Worth Matravers, but this is reflected in the comparison sites. Given this variation, possible differences in nutrition in relation to burial position within the cemetery were studied (Figure 7.5). The group of individuals from the western extent of the cemetery (Figure 7.5 - West) have higher nitrogen levels than individuals in the other rows, suggesting that these people were eating a more protein-rich diet, which might indicate a higher social standing.

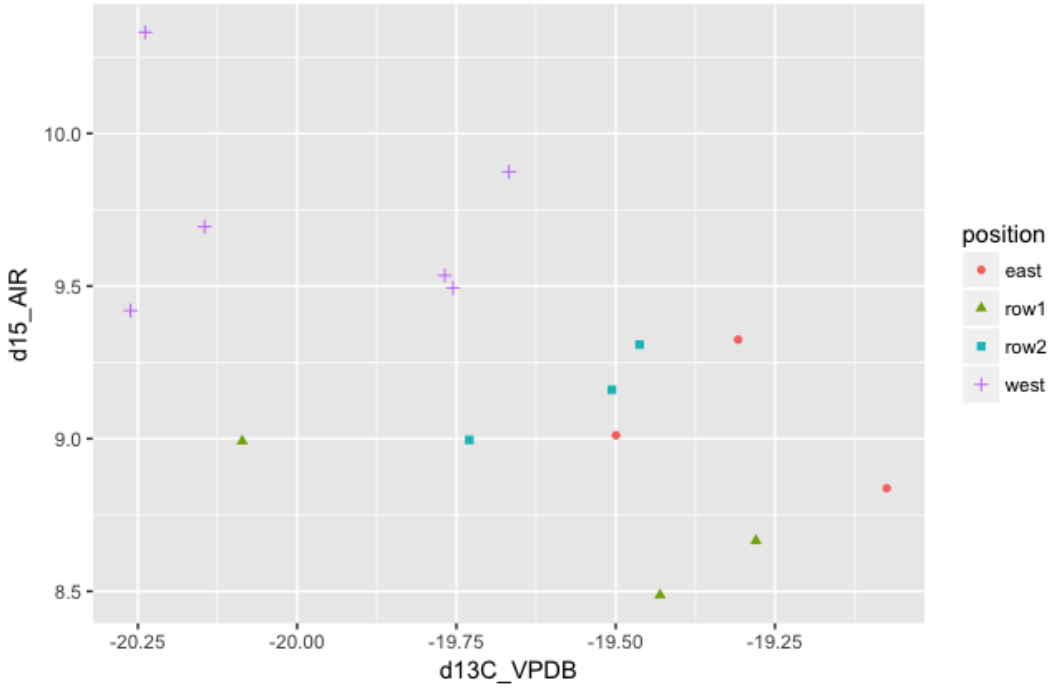

FIGURE 7.5: Dietary stable isotope data plotted by position within the cemetery.

7.3.3 Genetic sex identification

The genetic sex identification of the 16 genotyped individuals resulted in a total of ten males, five females and one individual whose sex could not be genetically identified (see Table 7.3). The layout of male and female burials is shown in Figure 7.3 B. The genetically undetermined individual is included in the females, as osteological analysis suggested that this was a female individual.

7.3.4 Mitochondrial genome analysis

Mitochondrial SNP capture was used to identify the mitochondrial haplotypes of the burials. The analysed individuals belong to nine different haplotypes (Table 7.3); four H, two U5, and one each for K1, T2 and HV. The distribution of mitochondrial haplotypes across the cemetery is shown in Figure 7.6 A.

TABLE 7.3: Uniparental haplotypes of the individuals within each grave from the post-Roman cemetery at Worth Matravers. Genetic versus osteological sex determination, and osteological age estimates of the individuals are given. Letters A, B and C indicate different individuals within one burial.

Burial number	Genetic sex	Osteological sex	Age estimate	mtDNA haplogroup	Y-chromosome haplogroup	chr Y markers
1633A	M	Undefined	adult	K1a2a	R1b1a1a2a1a2c1a2a2a1a1a	920
1640	M	Female?	17-25 years	K1a2a	I1a2a1a2∼	412
1649	F	Undefined	25-35 years	T2a1a	-	-
1633B	M	Undefined	17-25 years	U5b1	E1b1∼	382
1664	F	Female?	35-45 years	H11a	-	-
1667	U	Female?	25-35 years	HV6	-	-
1670	M	Female	17-25 years	HV6	R1b1a1a2a1a	538
1678A	M	Male	40-45 years	HV6	R1b1a1a2a1a	662
1678B	F	Undefined	15-16 years	HV6	-	-
1685A	F	Female	25-35 years	H1bb	-	-
1685B	M	Male	45-49 years	H1e1a	R1b1a1a2a1a	798
1685C	M	Undefined	15 years	H1bb	R1b1a1a2a1a	454
1697	F	Female?	25-35 years	U5b2c1	-	-
1715	M	Undefined	adult	T2a1a	R1b1a1a2a1a2c1a	274
1722	M	Female?	17-25 years	H5c	R1b1a1a2a1a1c2b2b1a	1899
1778	M	Female	16-17 years	T2a1a	I2a1b1	279

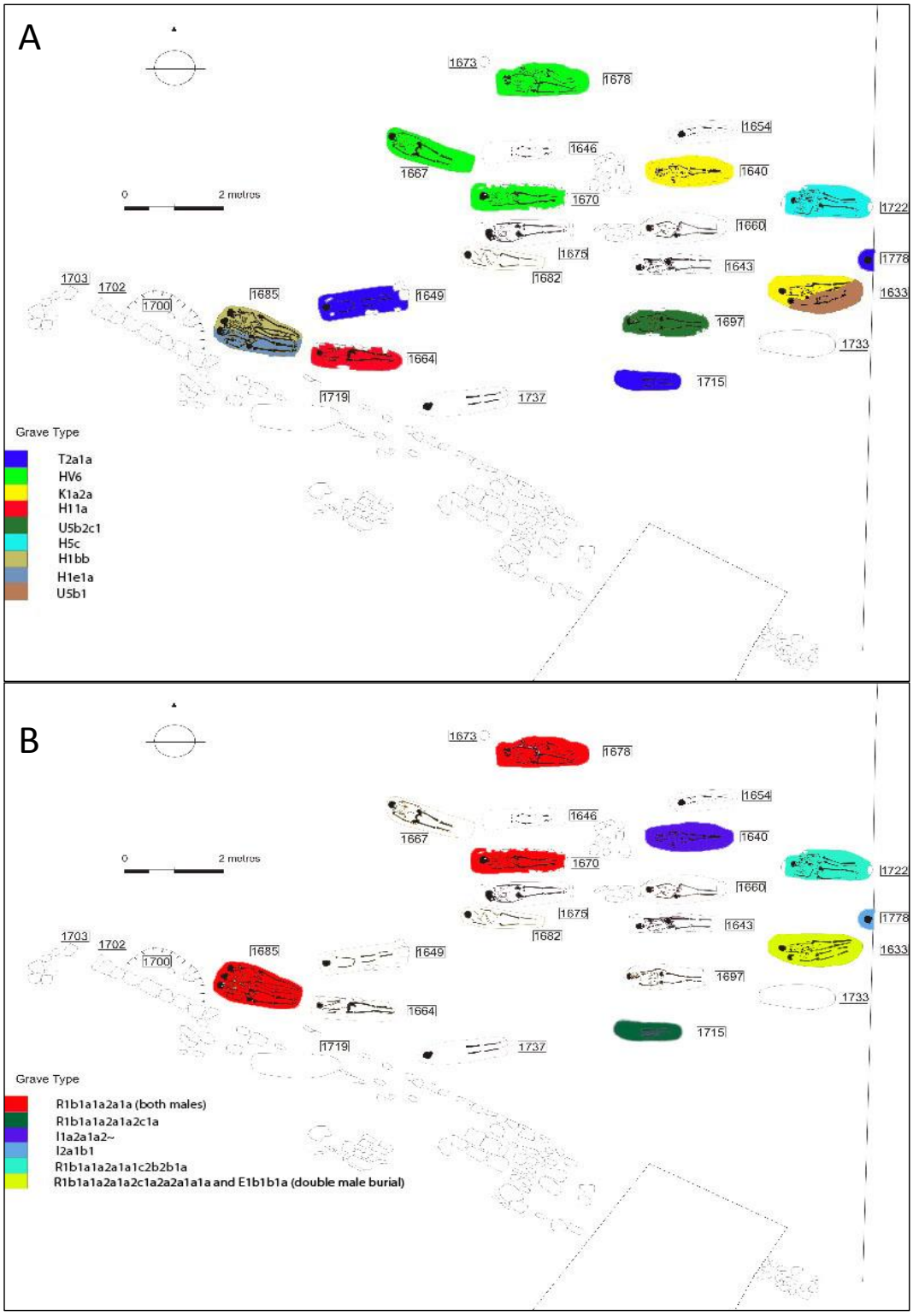

FIGURE 7.6: A - Mitochondrial haplotypes identified in the post-Roman cemetery, Worth Matravers. B - Y-chromosome haplotypes.

7.3.5 Y-chromosome analysis

Of the ten genetic males, seven were predicted to belong to different subclades of Y-chromosome haplogroup R1b, two were predicted to belong to haplogroup I, and one individual was predicted to belong to haplogroup E (Table 7.3). The distribution of Y-chromosome haplogroups within the cemetery is shown in Figure 7.6 B. Full list of all derived predictive markers can be found in Appendix G.

7.3.6 Kin relationship estimation

READ identified two first-degree (parent-offspring or siblings) and five second-degree (i.e. uncle/aunt-nephew/niece, grandparent-grandchild or half-siblings) kin relationships (Table 7.4).

TABLE 7.4: Estimated genetic kin relationships. Higher and lower class of genetic difference (Z_upper and Z_lower) are given as standard errors from the normalised mean *P*0 score.

Sample pair	Relationship	Z_upper	Z_lower
1633A+1640	second-degree	4.7258199089	-1.4743260222
1667+1670	second-degree	4.3994772251	-1.4064270971
1667+1678A	first-degree	3.126670294	-9.495248693
1678A+1670	second-degree	4.6555947222	-2.1036976561
1685A+1685C	first-degree	3.5705830828	-11.0245599904
1697+1715	second-degree	7.277960848	-0.5408975717
1715+1778	second-degree	0.1617568526	-2.6359153615

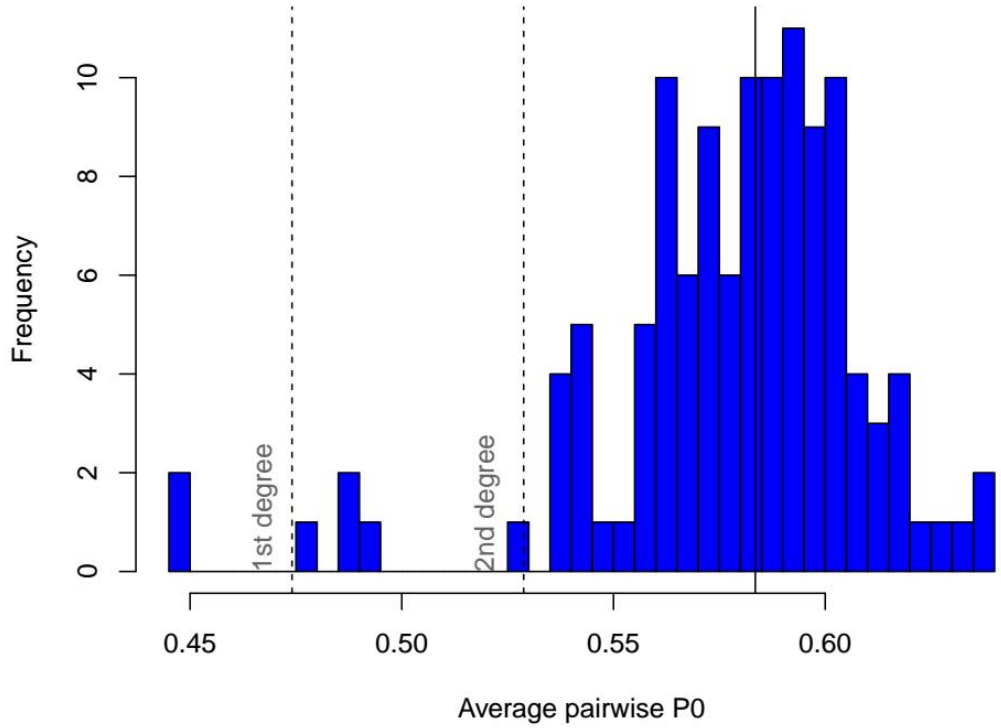

FIGURE 7.7: READ estimates for the post-Roman individuals. The histogram shows the non-normalised average *P*0 values. Vertical dashed lines show the cutoffs used to classify how the individuals were related. The solid line indicates the median value used for normalisation based on the expected value for a randomly chosen pair of unrelated individuals (Kuhn et al., 2018).

The frequency of the overall kin relationships identified using READ are shown in Figure 7.7. Burials connected through first- and second-degree kinships are shown in Figure 7.8 A and B.

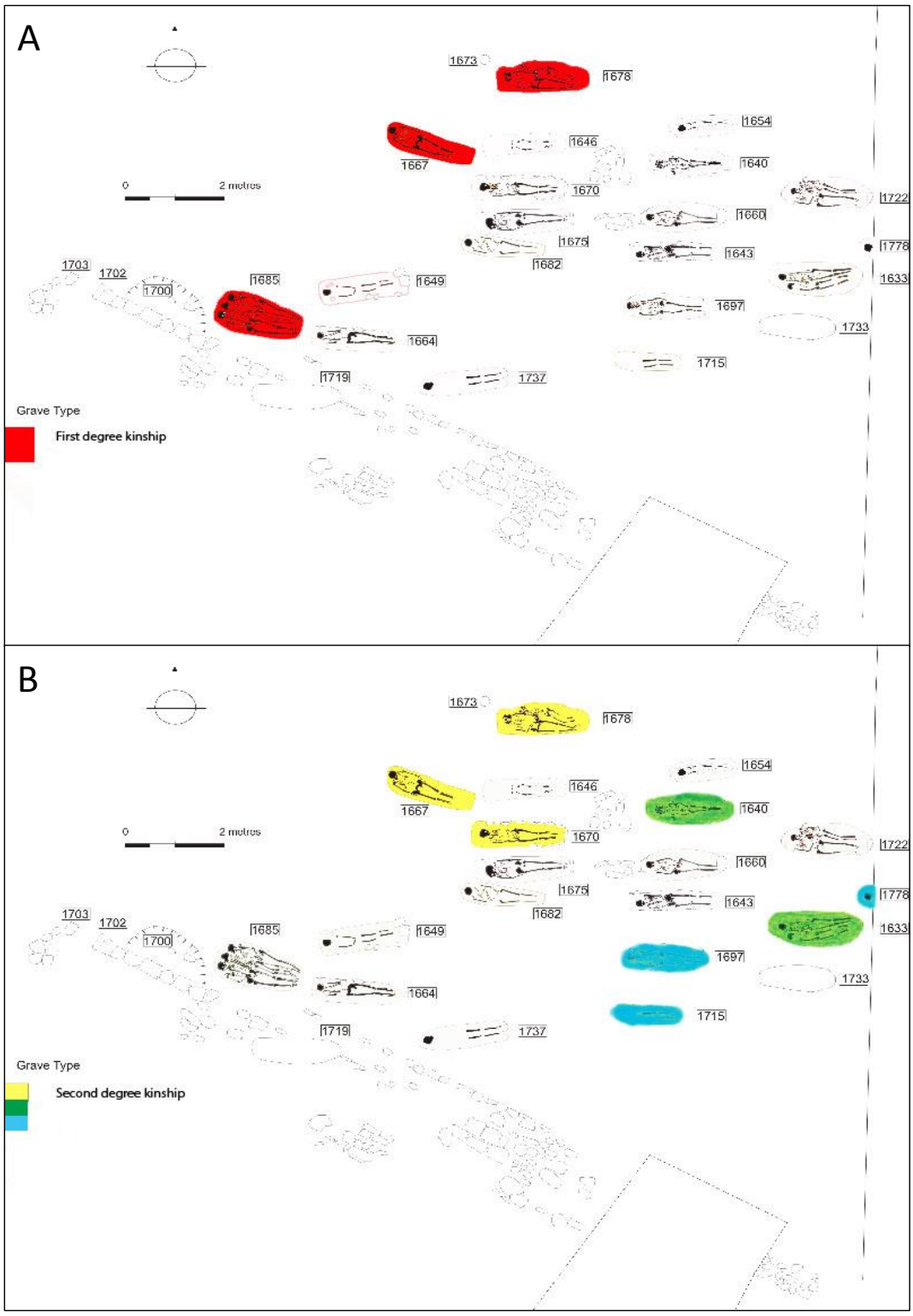

FIGURE 7.8: A - First-degree kinships. B - Second-degree kinships.

7.3.7 Principal component analysis

The post-Roman individuals fall within the modern European variation, with one sample being closer to southern Europeans and Near Eastern individuals

(Figure 7.9).

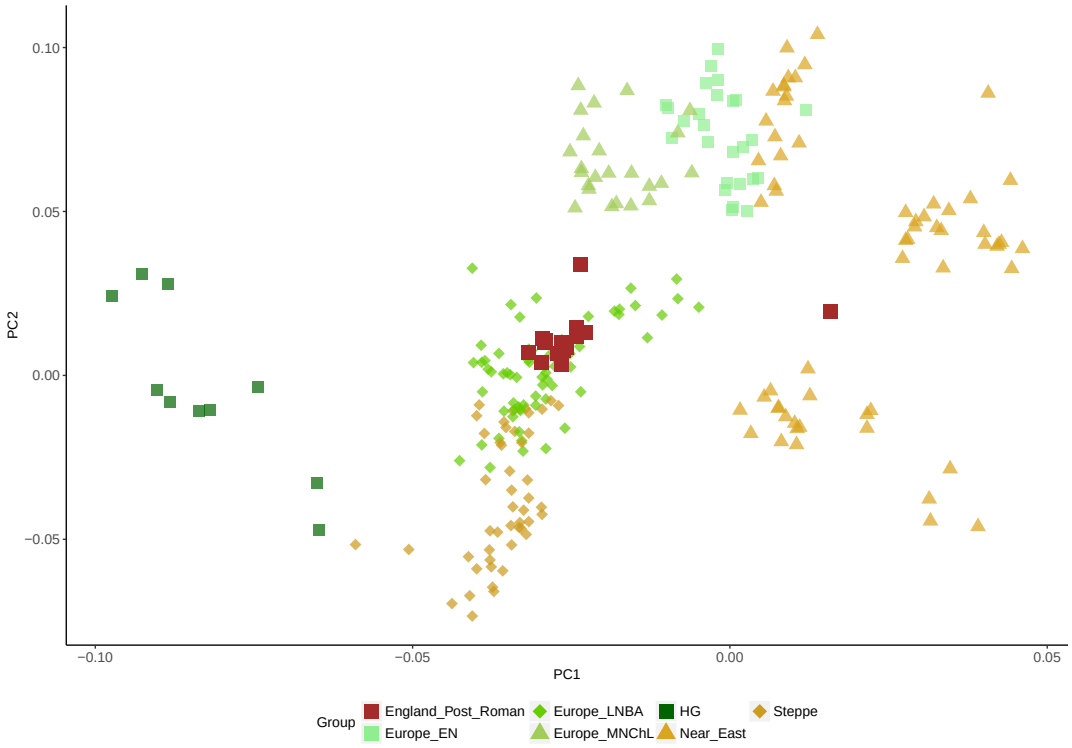

FIGURE 7.9: Principal component analysis of the 16 post-Roman individuals (green dots), with previously published modern and ancient samples of Lazaridis et al. (2016) projected onto the first two principal components.

7.3.8 ADMIXTURE analysis

The samples were merged with the publicly available datasets from Lazaridis et al. (2016) and Olalde et al. (2018). ADMIXTURE was run with cross-validation between K values 2 to 14, with K=12 giving the lowest standard error and therefore being the sensible modelling choice (Figure 7.10) (Appendix H).

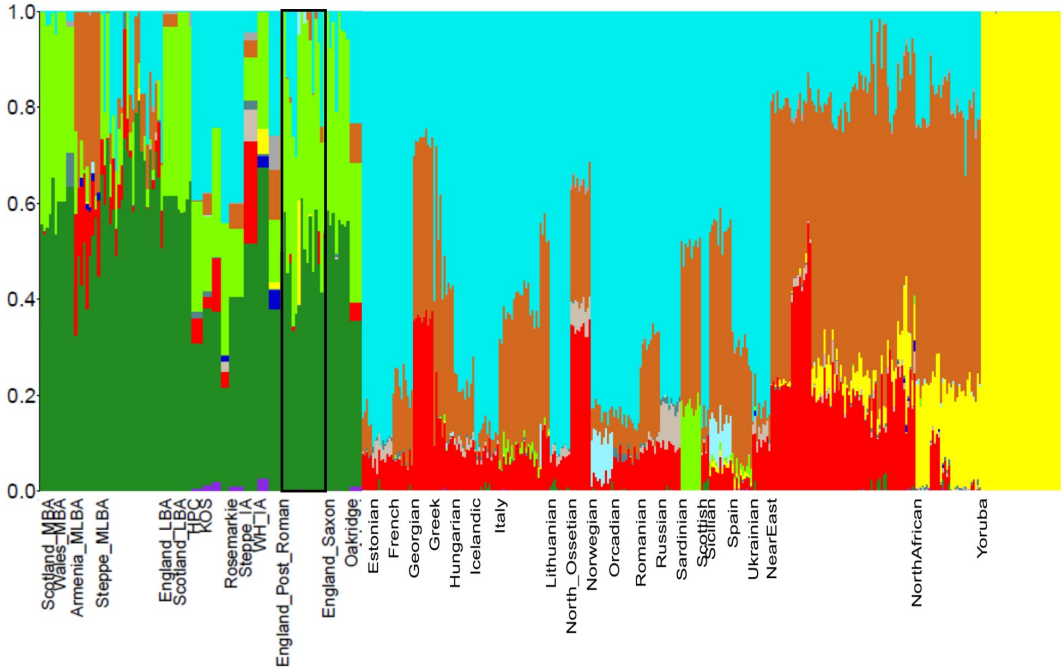

FIGURE 7.10: ADMIXTURE analysis of the post-Roman individuals with previously published modern and ancient samples of Lazaridis et al. (2016) and Olalde et al. (2018). Abbreviations used: E: Early, M: Middle, L: Late, BA: Bronze Age, IA: Iron Age, WH: West Heslerton, HPC: High Pasture Cave, KOS: Knowe of Skea.

7.4 Discussion

7.4.1 SNP capture success

The overall success rate of the 1,240K SNP capture was high, as all samples passed the quality controls, and all but one produced more than 20,000 SNPs (1678B produced <7,500 SNPs). The number of Y-chromosome markers retrieved was low, but nevertheless it was possible to predict the haplogroups as far as derived clades of the Y-chromosome tree. Similarly, the haplotyping of the mitogenomes allowed significant haplogroup assignments.

7.4.2 Radiocarbon dating

Based on radiocarbon dates (analysed by Kruz 2018), the cemetery was in use for less than 200 years and several of the dated burials have overlapping dates (Figure 7.11).

Burials belonging to the intermediate phase, with almost identical dates, contain the triple burial of a family (1685), the two double burials (1678 and 1633), and one single burial (1640). The case of, the multiple burial of three

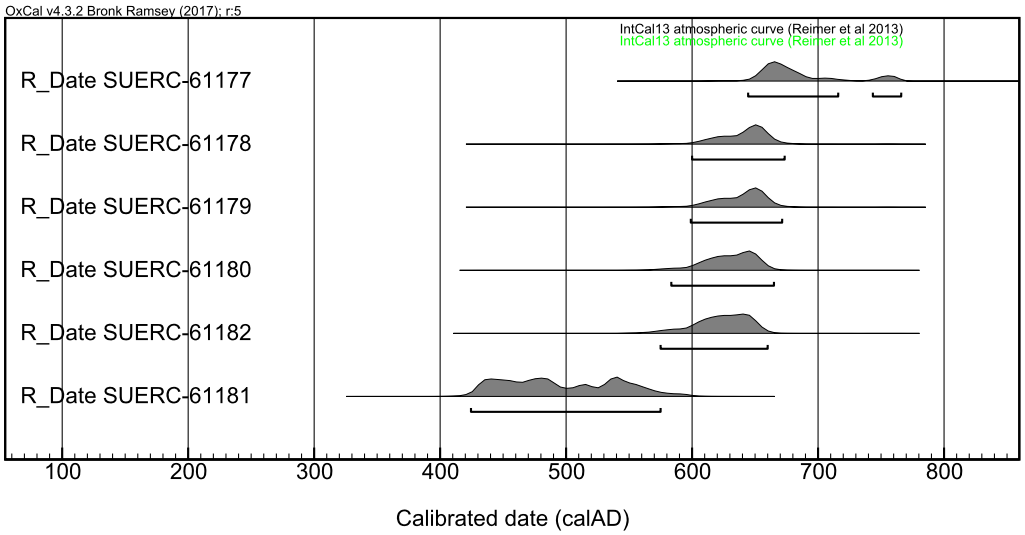

FIGURE 7.11: Calibrated radiocarbon dates of the dated individuals from Worth Matravers. The corresponding grave numbers to the SUERC codes are as follows: SUERC-61177 - grave 1667, SUERC-61178 - grave 1633, SUERC-61179 - grave 1640, SUERC-61180 - grave 1678, SUERC-61181 - grave 1697, SUERC-61182 - grave 1685.

individuals of different ages suggests that these people likely passed away contemporaneously, possibly due to disease.

7.4.3 Anglo-Saxon diet

The dietary stable isotopic analysis identified two main groups within the cemetery. Individuals buried on the western side of the cemetery have a higher nitrogen intake and so had a more terrestrial/meat-based diet. All the other individuals have a comparably lower nitrogen level, suggesting a slightly more marine influence; most likely from shellfish. Compared to published dietary stable isotope data from Anglo-Saxon sites in Dorset, the post-Roman individuals from Worth Matravers had a comparably more marine-based diet of low trophic animals and plants, showing lower trophic levels than individuals from Alington Avenue, Old Vicarage, and Albert Road (all Anglo-Saxon sites in Dorset). Their diet was possibly based on shell fish and sea weed, as they were close to the coast. They were most similar to Anglo-Saxon individuals from Maiden Castle Road (Dorset), Albert Road (Dorset), and Alington Avenue (Dorset). However, the range of $\delta^{13}C$ and $\delta^{15}N$ values of these individuals is quite large and implies different ways of subsistence among these populations, with rather strong differences in protein intake based on plants and/or animals.

7.4.4 Estimated kin relationship

The only triple burial is also the only excavated cist-burial (grave 1685), which includes one female aged 25-35 years, a male aged 45-49 years and a male adolescent. Based on the analysis of their uniparental markers and kinship analysis, it is very likely that this burial contains a family. Both the adult female and the adolescent man share the same mitochondrial haplotype (H1bb), suggesting a mother-son relationship, and the male and the adolescent share the same Y-chromosome haplotype (R1b1a1a2a1a), suggesting a father-son relationship. READ identified a first-degree (parent-offspring or siblings) kin relationship between the adult female and the adolescent, whereas no kin relationship between the adult male and young man could be identified. However, this might be a false negative assignment due to low coverage of the data.
The kin relation of the individuals in burials 1667 and 1670, and double burial 1678 is particularly interesting. All four individuals share the same mitochondrial haplogroup, HV6. The two men in burials 1670 and 1678 share the same Y-chromosome haplogroup (R1b1a1a2a1a), which is the same as the two men in the triple burial (1685). Based on the kinship analysis, the young woman in the double burial 1678 does not have first- or second-degree kinship to any of the other three individuals and will thus be excluded from further discussion of the degrees of relationship between the individuals in these three graves. The possible female individual in grave 1667 and the old man in grave 1678 are first-degree related, whereas the man in burial 1670 shares second-degree kinship with them both (Figure 7.12 - A). The potential scenarios of kinship among them are not very straightforward. Assuming the burial of the female is about 50 years younger (based on the age difference; the female is 25-35 and the male is 40-45) than burial 1678 of the older man, a mother-son relationship can be excluded. This leaves either a:

1. father-daughter kin relationship; or

2. sibling relationship

The sibling scenario is quite unlikely given the age difference of the individuals and the radiocarbon age difference of the graves. However, the standard errors of the radiocarbon dates are slightly overlapping, which means the burials could be contemporaneous. It is possible that, due to low coverage data, the kin relationship has been classified as first-degree, while in reality it is a second-degree kinship and the individuals have an uncle-niece

or grandparent-granddaughter relationship. As the samples had more than 1,000 overlapping SNPs the likelihood for a false positive degree classification is <2% (Kuhn et al., 2018). Otherwise this would suggest that the two individuals have a father-daughter relationship and the shared mitochondrial haplogroup is due to inbreeding in the maternal lineage.
The second-degree kinship of the man in burial 1670 with both these people suggests either that:

1. the man in burial 1670 is the uncle or grandfather of both the woman in burial 1667 and the man in 1678. This would mean that these two are siblings rather than father and daughter; or

2. the man in burial 1670 is the half-sibling of the man in burial 1678 and the uncle of the woman, so the first-degree kinship between 1667 and 1678 explains an offspring-parent relation.

Although, it is not possible to identify one specific kinship scenario, the shared mitochondrial haplogroup and shared Y-chromosome haplogroup between these individuals suggest close familial kin relationships. This is supported by the kinship analysis, even if inaccurate assignments of first- or second-degree kinship have to be accounted for, due to low coverage data.
Unfortunately, there are no stable isotopic or genetic data available for the three individuals that are buried in graves 1646, 1675 and 1682, which are neighbouring graves of the three genotyped ones (Figure 7.8). Based on the close proximity of these graves, and the familial relationships between the individuals analysed, it is possible that the people with missing data may share familial bonds with those analysed. Further sampling and genetic analysis would elucidate if the proximity of the graves of these individuals is based on familial bonds or by chance.
The three individuals in burials 1649, 1715 and 1778 share the same mitochondrial haplogroup, whereas the Y-chromosomes of the men in burials 1715 and 1778 are different, R1b1a1a2a1a2c1a and I2a1b1, respectively. Therefore, any familial kinship is more likely to be maternal. Kinship analysis could not pick up any first- or second-degree relation between the woman and either of the two men, but a second-degree kinship could be identified between the males. This means they are either grandfather and grandson or uncle and nephew (see Figure 7.12 - B). Furthermore, a second-degree kinship was identified between the man in grave 1715 and the woman in grave 1697. As they have very different mitochondrial haplotypes, T2a1a and

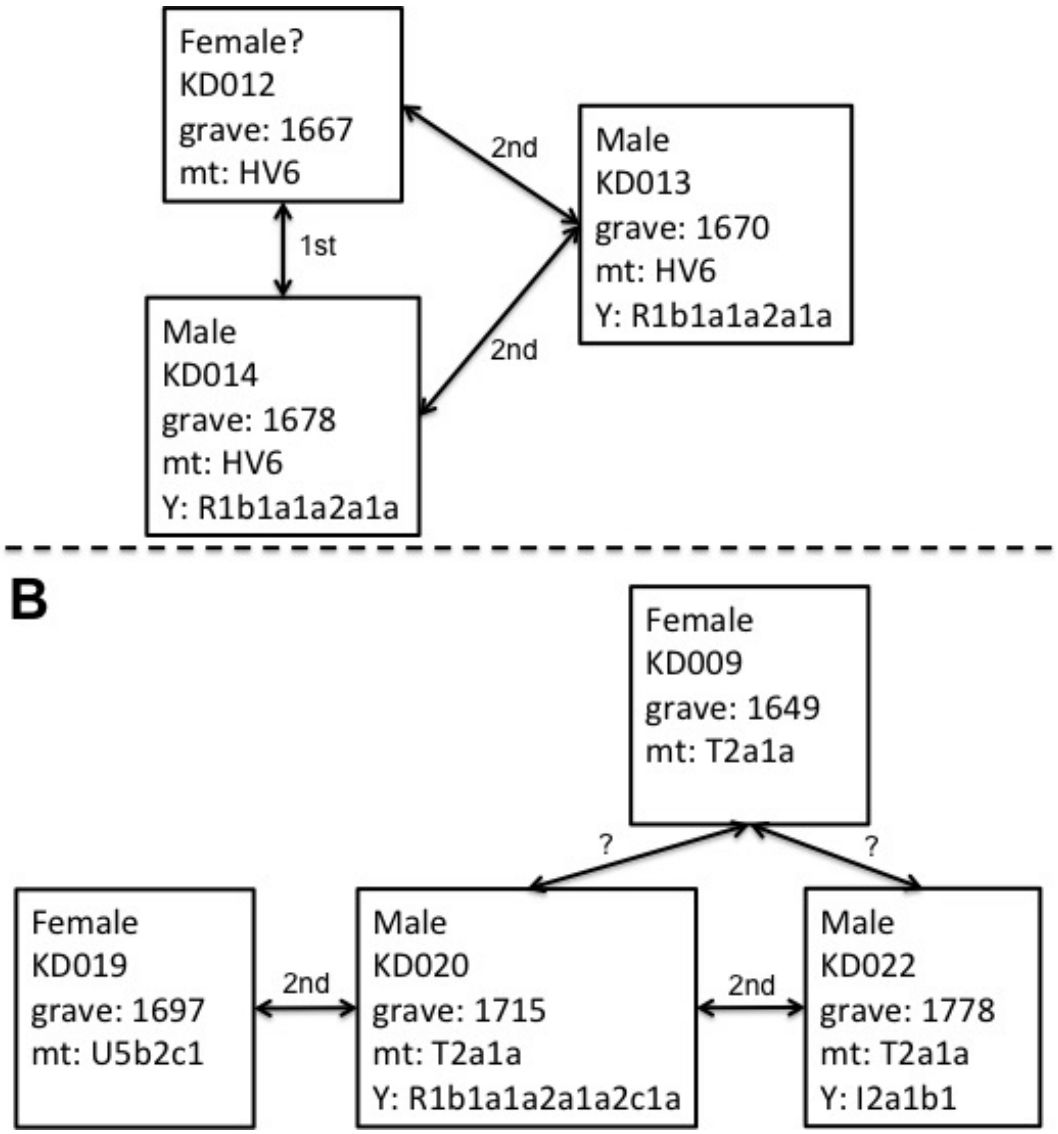

FIGURE 7.12: Schematic of estimated kin relationships within the post-Roman cemetery, Worth Matravers. A - Kin relationships between burials 1667, 1670 and 1678. B - Kin relationships between burials 1649, 1697, 1715 and 1778.

U5b2c1, respectively, they are very likely paternally related; either grandfather and granddaughter or uncle and niece; although another possibility would be that they shared the same father but had different mothers.
Other second-degree kinships were identified between the man in burial 1640 and the older man in the double burial 1633. Possible kin relationships between them are:

1. half-siblings – same mother, different fathers; or

2. uncle and nephew.

Interestingly, the older male in the earth-cut double burial (grave 1633), who had his head resting on a stone anchor, was buried with a young male who

has a mitochondrial haplotype belonging to U5b1 and a Y-chromosome haplotype belonging to E1b1∼, with the most predictive markers being L507, L535 (more downstream predictive markers CTS10775 and CTS3171, were inconclusive as they point to different branches). Although, it was not possible to predict a more downstream haplogroup of E1b1∼ due to low coverage of the Y-chromosome SNPs, no contradictory markers were identified, making the prediction of E1b1∼ most likely. Both uniparental markers of this young man have not been identified in any other person within the cemetery. As there seems to be no apparent kin relationship between him and the older man he was buried with, the type of burial with the anchor suggests a work relationship, such as an apprentice. Possibly they were both fishermen. Such a mixed burial of an immigrant and a local has not yet been documented.
The two individuals in single burials with no apparent kinship relations to anyone analysed samples are: a female in grave 1664 with mtDNA haplotype H11a and a male in grave 1722 with mtDNA haplotype H5c and Y-chromosome haplotype R1b1a1a2a1a1c2b2b1a.

7.4.5 Social structure found within the cemetery

The west-east gradient imposed by the stable isotope data on the schematic of the cemetery suggest that familial bonds were not the important factor for the proximity of familial burials in all cases. An example is burial 1649, which shared the maternal lineage with individuals buried further east in the cemetery. This might be due to interment at a different time, but could also be due to a different social status, which led to better access to food or they were very distant relatives, as the kinship analysis did not suggest any close kin relationship.
The analysis of uniparental markers suggests patrilocality at the site, and the overall layout of the cemetery seems to be based on familial bonds and kin relationships, as multiple burials contain either entire families or graves of family members are closely placed. Overall, no female to female first- or second-degree kinship could be observed, whereas male to male first-degree and second-degree kin relationships were prevelant. This supports patrilocality within the site.

7.4.6 Genome-wide analysis

Autosomal analysis of the 16 post-Roman individuals showed that they all clustered with ancient European Late Neolithic/Bronze Age people, with

the exception of one individual (KD010, 1633B) who clustered closest with ancient Near Eastern samples (Figure 7.9). ADMIXTURE analysis showed that this individual shared ∼50% of its ancestry with modern Near Eastern/North Africans. As he belongs to a typical European mitochondrial haplogroup, but has a Near Eastern/North African Y-chromosome haplogroup, and based on the proportions of European and North African ancestry components, it is most likely that he had a mother from northwest Europe and a Near Eastern/North African father. An individual with Near Eastern ancestry has been reported from a Roman cemetery in York, England (Martiniano et al., 2016). These people are evidence of long-distance migrations and diversity during the Roman Empire, which appears to have lasted into the post-Roman era.

Chapter 8

Multi-period West Heslerton, East Yorkshire

8.1 Introduction

The Early Anglo-Saxon or Anglian cemetery at West Heslerton, North Yorkshire, was discovered during mineral excavation operations at Cook's Sand Quarry (Haughton et al., 1999b) (Figure 8.1). The discovery of the site led to an archaeological rescue excavation that was conducted between 1977 and 1987 (Haughton et al., 1999b).

The first documented occupation at the site was found to date back to the Late Mesolithic, when the area was used for flint knapping, and there is evidence that people continued to live there up into the Bronze Age. During the Late Neolithic and Early Bronze Age, the site was used for rituals and the establishment of a field system was attempted, as well as the foundation of two Early Bronze Age (EBA) barrow cemeteries (Haughton et al., 1999b). One of the EBA cemeteries produced an important series of Food and Beaker vessels (Powlesland et al., 1986). Woodland regenerated in the area after the cemeteries went out of use during the late Bronze and early Iron Ages, but then the site was once again continuously occupied throughout the Roman and Anglo-Saxon periods, with evidence of settlement from the Late Iron Age onwards (Haughton et al., 1999b). The Anglo-Saxon, or Anglian, cemetery is the only case in the North of England where this combination of cemetery and settlement have been investigated together (Haughton et al., 1999b).

The rescue excavation examined 11 prehistoric burials and 201 out of an estimated 300 Anglian burials, consisting of both inhumations and cremations. The Anglian cemetery was established in the late 5th century within an area defined by an earlier Late Neolithic and Early Bronze Age ritual complex. The cemetery was in use for not more than 200 years until the 7th century

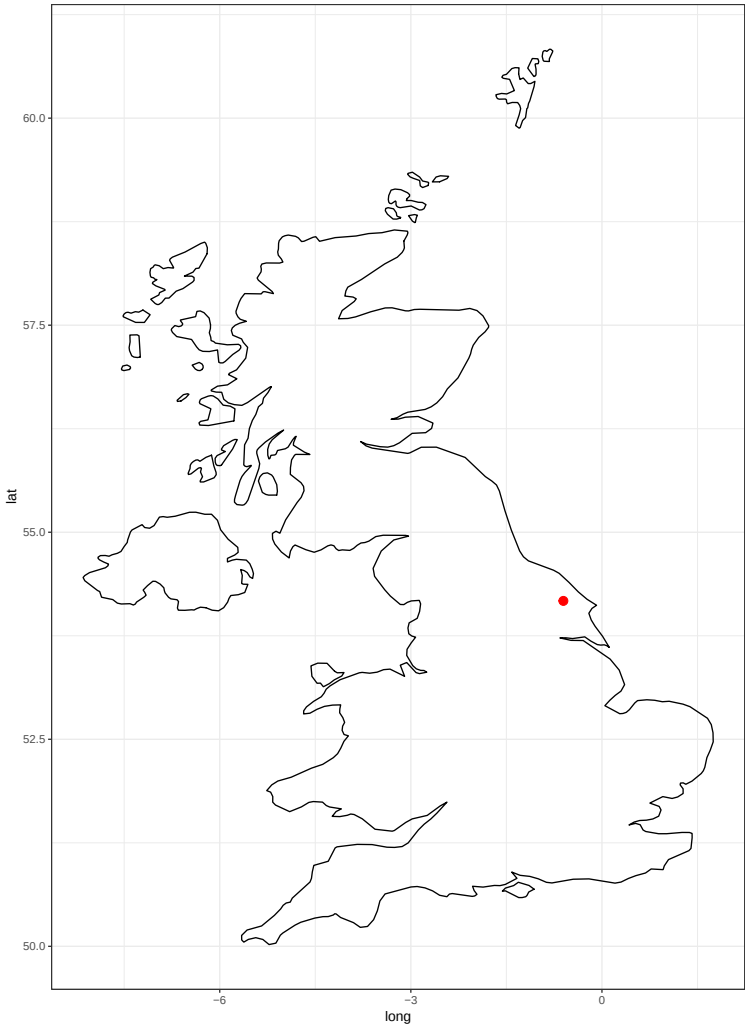

FIGURE 8.1: The location of West Heslerton within the British Isles.

AD, which does not cover the entire occupation period of the associated settlement (Haughton et al., 1999a; Haughton et al., 1999b).

The bone preservation of the Anglo-Saxon inhumations was very poor. Half of the graves contained either no skeletal material or only fragments of bone or tooth enamel, which is due to the well-drained sandy soil (Haughton et al., 1999b). Preservation was better only in a few graves in an excavation area with chalk soil. In contrast, the few Bronze Age and Iron Age burials had a much better preservation of skeletal material, which according to Haughton et al. (1999b) is possibly due to being buried in soil substrates with more chalk inclusions.

Nearby the Anglo-Saxon cemetery, two female Iron Age individuals were excavated from Arras burials at Knapton Wold, West Heslerton. The female skeletons were well preserved and assumed to be related to each other.

As their burials were associated with the settlement in West Heslerton, they were analysed together with the individuals from the Bronze Age and Anglo-Saxon cemetery. A total of 97 samples, mainly petrous bones, other cranial fragments and teeth, were selected. Of these, a total of the 15 best preserved individuals from the Bronze Age and Anglo-Saxon period were sampled for genetic analysis. The cemetery in West Heslerton offers the opportunity to analyse genomic data from individuals from different time periods in a continuously occupied area.

8.2 Materials and Methods

Skeletal remains were assessed for preservation and the best preserved samples from different time periods were sampled for genetic analysis. In total, four Bronze Age individuals, two Iron Age individuals and nine Anglo-Saxon individuals were analysed by sampling petrous bones.

TABLE 8.1: Processed samples from West Heslerton. Analysis types: L – NBAF Liverpool; M – Macrogen; R – Reich, Harvard Medical School.

DNA code	DNA sample	Analysis type	Era	Sample code
KD002	petrous	L	Anglo-Saxon	HP2BA100
KD003	petrous	L	early Bronze Age	HP2BA283 AN 1806
KD031	petrous	R	Anglo-Saxon	HP2BA790 AH 10139
KD032	petrous	R	Anglo-Saxon	HP2BA22 AP
KD033	petrous	R	Anglo-Saxon	HP2BA148 AN
KD034	petrous	R	Anglo-Saxon	HP2BA766 AM 11074
KD035	petrous	R	Anglo-Saxon	HP2BA775 AA 12206
KD036	petrous	R	Anglo-Saxon	HP2BA802 AJ 5146
KD037	petrous	R	Anglo-Saxon	HP2BA805 AO
KD038	petrous	R	Anglo-Saxon	HP2BA890 AI 10148
KD039	petrous	M	early Bronze Age	HP2BA589 AC 11035
KD040	petrous	M	early Bronze Age	HP2BA229 21
KD041	petrous	M	early Bronze Age	HP2BA241 AH
KD071	R petrous	M	Iron Age	82 CA [18]
KD072	L petrous	M	Iron Age	82 CA [27]

8.2.1 Ancient sample processing

Ancient DNA sample processing was undertaken at the Ancient DNA Facility of the University of Huddersfield, and the Ancient DNA Lab, Smurfit Institute, Trinity College Dublin (TCD) (Ireland), under dedicated clean-room conditions. The same protocol for sample processing and ancient DNA extraction was used in both labs.

8.2.2 Sampling and DNA extraction

Sampling and DNA extraction was undertaken at both laboratories as outlined in Chapter 6. Dietary stable isotope sampling of KD071 and KD072 followed as outlined in Chapter 2.

8.2.3 Library preparation and sequencing

8.2.3.1 Library preparation

Libraries for shot-gun sequencing of seven samples (four Bronze Age, two Iron Age and one Anglo-Saxon individual) were prepared. Library preparation followed the protocol described in Chapter 2. One single-indexed library was prepared at TCD for samples KD002 and KD003. Eleven further libraries were prepared at the University of Huddersfield. For KD002 and KD003, three libraries each were dual-indexed and one single-indexed library each was prepared for samples KD039, KD040, KD041, KD071 and KD072. The same protocol was used in both labs, with the only difference being the single- and dual-indexing. For further details see Chapter 6.

8.2.3.2 Sequencing

The NGS libraries prepared at TCD were sequenced in an Illumina HiSeq 2000 (single-index, 100 base pair, single-end sequencing; Macrogen, South Korea), with library codes KaD2 and KaD3 (equivalent to samples KD002 and KD003, respectively). NGS libraries prepared at the University of Huddersfield were sequenced in an Illumina HiSeq 2000 (dual-index, 100 base pair, single-end sequencing; NBAF Liverpool, United Kingdom) [with library codes KD002.1, KD002.2, KD002.3, KD003.1, KD003.2 and KD003.3], and in an Illumina HiSeq4000 (single-index, 100 base pair, paired-end sequencing; Macrogen, South Korea) [with library codes K1, K2, K3, K22 and K24, for samples KD039, KD040, KD041, KD071 and KD072], respectively. Dual-indexing was undertaken due to the requirements of NBAF Liverpool, and single-end versus paired-end sequencing at Macrogen was based on the offer given per lane.

8.2.4 Data processing and read mapping

The NGS reads of all single-end sequenced libraries were trimmed using cutadapt version 1.13 (Martin, 2011), whereas the NGS reads of paired-end sequenced libraries were trimmed using AdapterRemoval version 2.1.7 (Schubert et al., 2016). For further details please refer to Chapter 6. Read files for the multiple libraries of KD002 and KD003 were merged using samtools merge. The detailed steps of analysis are described in Chapter 2. Kinship analysis was carried out using READ (Relationship Estimation from Ancient DNA)(Kuhn et al., 2018). As the obtained data had very low coverage, and the vast majority of positions were only covered by a single read, all genotypes were converted to homozygous. Genetic sex determination was undertaken using the script by Skoglund et al. (2013). Y-chromosome haplogroup determination was carried out using Yleaf (Ralf et al., 2018). In order to compare the newly generated ancient dataset to modern human populations, bases to known SNP positions of the publicly available Lazaridis et al. (2016) dataset were used and SNPs were filtered as described previously (Lazaridis et al., 2016). Principal Component Analysis was run using smartpca of the EIGENSOFT package (Patterson et al., 2006; Price et al., 2006) and plotted with R version 3.4.4 (R Core Team, 2018). A combined dataset of the newly generated genotypes, together with the reference panel, was run using ADMIXTURE (Alexander et al., 2009).

8.2.5 SNP capture

DNA extracts of eight Anglo-Saxon samples were shipped to collaborator David Reich (Harvard Medical School) for enriched 1.2M SNP and mitogenome capture. Data processing followed the pipeline used in Olalde et al. (2018).

8.3 Results

Overall sequencing results are given in Table 8.2. Information on the negative controls can be found in Appendix D.

TABLE 8.2: Sequencing results of all shotgun-sequenced samples from West Heslerton and Knapton Wold. 'SNPs' denotes the number of SNPs called with the published datasets as reference panel.

Age	Lab ID	Read count	Endogenous content in %	Duplication rate in %	Genome coverage	Chr. Y haplogroup markers	SNPs
Bronze Age	KD003	3,893,102-23,948,074	9.36-28.4	1.1	0.1608	–	90,090
	KD039	75,915,892	9.74	0.1319	0.0078	4,914	4,311
	KD040	70,076,838	9.97	0.1105	0.0152	324	9,084
	KD041	81,678,236	13.38	0.1137	0.0267	672	16,544
Iron Age	KD071	116,208,258	7.86	0.1009	0.0186	–	12,134
	KD072	76,431,932	11.83	0.1111	0.0159	–	10,303
Anglo-Saxon	KD002	21,712,945-93,412,198	0.34-1.89	0.32	0.0265	–	16,294

Detailed endogenous contents of the two samples KD002 and KD003, consisting of multiple libraries, are given in Table 8.3.

TABLE 8.3: Read count and endogenous contents of KD002 and KD003 libraries. Indexing type indicates whether a library was single or dual indexed. Read type indicates single-end (SE) and paired-end (PE) sequencing.

Lab ID	Read count	Endogenous content in %	Indexing type	Read type
		KD002		
KaD2	93,412,198	0.34	single	SE
KD002.1	29,201,886	1.65	dual	PE
KD002.2	22,710,532	1.69	dual	PE
KD002.3	21,712,945	1.89	dual	PE
		KD003		
KaD3	23,948,074	9.36	single	SE
KD003.1	14,539,263	28.06	dual	PE
KD003.2	17,144,472	28.23	dual	PE
KD003.3	3,893,102	28.40	dual	PE

Because the endogenous content of the Anglo-Saxon sample KD002 was very low in all libraries, it was decided to try the SNP capture approach on the remaining Anglo-Saxon samples, rather than shot-gun sequence them. The SNP capture results of the eight additional Anglo-Saxon samples are given in Table 8.4.

TABLE 8.4: SNP capture results of the Anglo-Saxon samples.

Sample ID	Coverage	SNPs hit on autosomes	chr. Y haplogroup markers
KD031	0.6210	439,024	4,914
KD032	0.3823	340,551	3,641
KD033	0.6914	509,157	–
KD034	0.6412	476,232	5,404
KD035	0.9794	601,957	–
KD036	0.2141	221,669	2,448
KD037	1.8333	691,863	–
KD038	1.0694	605,796	7,610

8.3.1 Genetic sex identification

The genetic sex of the shot-gun sequenced individuals, as determined by the Skoglund script (Skoglund et al., 2013), the Skoglund score (Skoglund et al., 2015) and karyotype plotting, is given in Table 8.5. As the sex of three of the shot-gun sequenced samples could not be determined by using the script (Skoglund et al., 2013), karyotype plotting and X-chromosome percentage calculation were used as aids to identify the genetic sex of the individuals (for details see Appendix E). Overall, four individuals were female and three were male.

TABLE 8.5: Genetic sex determination of shot-gun sequenced samples from West Heslerton using the Skoglund script [1] and the Skoglund score [2]. Samples in date order.

Age	Sample ID	Sex deter. based on [1]	Chr. X in %	Sex deter. based on [2]
Bronze Age	KD003	XX	–	–
	KD039	Not assigned	55.2	XY
	KD040	consistent with XY but not XX	50.8	XY
	KD041	XY	–	–
Iron Age	KD071	Not assigned	78.2	XX
	KD072	Not assigned	75.0	XX
Anglo-Saxon	KD002	XX	–	–
	KD031	XY	–	–
	KD032	XY	–	–
	KD033	XX	–	–
	KD034	XY	–	–
	KD035	XX	–	–
	KD036	XY	–	–
	KD037	XX	–	–
	KD038	XY	–	–

8.3.2 Dietary stable isotope analysis

In order to identify the diet of the two Iron Age samples from Knapton Wold (KD071 and KD072), dietary stable isotopes were analysed and compared to four other Iron Age sites from East Yorkshire: Garton Slack (n = 2); Kirkburn (n = 1); Wetwang Slack (n = 27); and Wetwang Village (n = 1); (Jay et al., 2013). The two individuals were eating very different diets, as can been seen in Figure 8.2.

8.3.3 Mitochondrial genome analysis

The coverage of the shot-gun sequenced mitochondrial genomes ranged from 0.9055x to 26.358x. The coverage of KD039 was too low to determine the mtDNA haplogroup. For Anglo-Saxon samples KD031 to KD038 no genome coverage and standard deviation are available, as these samples were typed using a capture array. Detailed mutation lists per individual can be found in Appendix F.

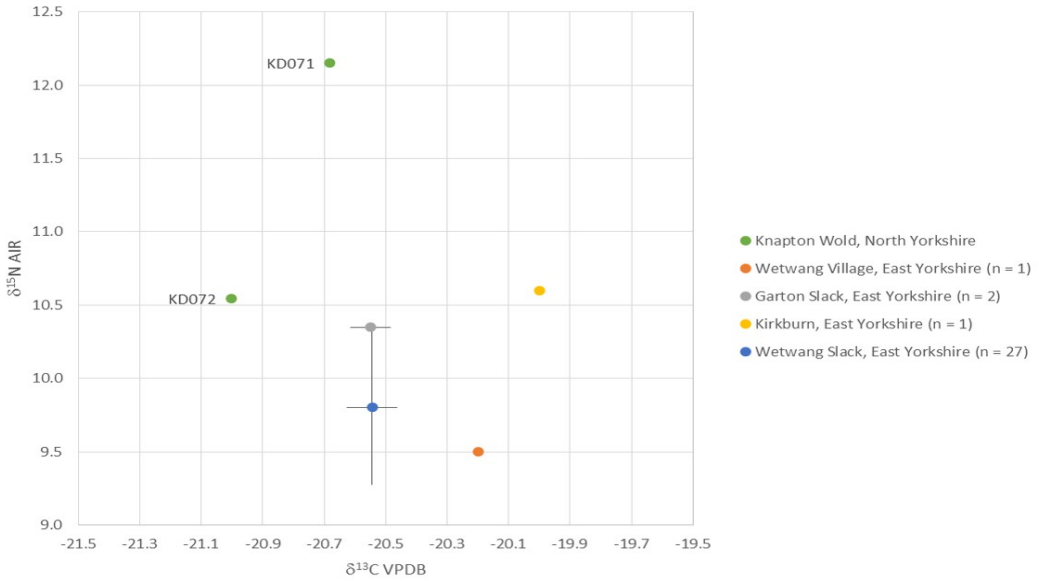

FIGURE 8.2: Dietary stable isotope analysis of the two Iron Age individuals from Knapton Wold (West Heslerton), compared to four other Iron Age human assemblages from East Yorkshire (Jay et al., 2013).

TABLE 8.6: Mitochondrial haplogroup assignment of the shot-gun sequenced samples from West Heslerton and Knapton Wold.

Sample ID	mitogenome coverage	std deviation	mtDNA haplogroup
KD003	26.358	5.2719	T2e
KD039	0.9055	1.7292	–
KD040	2.3820	1.9072	T2b4h
KD041	4.2561	2.2681	U5a1a2a
KD071	2.7945	2.0171	H1b1+16362
KD072	3.9670	2.5486	H1b1+16362
KD002	3.2841	1.8178	H6a1a
KD031	–	–	H1c1
KD032	–	–	H1q
KD033	–	–	H1b
KD034	–	–	X2b4a
KD035	–	–	U5a1a1e
KD036	–	–	K1d1
KD037	–	–	T2b2b
KD038	–	–	K1a4a1a+195

8.3.4 Y-chromosome analysis

One of the three shot-gun sequenced males from the Bronze Age did not have enough coverage to predict the Y-chromosome haplogroup (KD039). However, for the remaining samples, KD040 and KD041, and the five Anglo-Saxon males, the predicted Y-chromosome haplogroups are given in Table 8.7. Detailed lists of retrieved derived predictive markers per individual can be found in Appendix G.

TABLE 8.7: Predicted Y-chromosome haplogroups of the West Heslerton Anglo-Saxon individuals.

Sample ID	Predicted Y-chromosome haplogroup	Most downstream predictive markers
KD031	R1b1a1a2a1a1	F69 PF6438 L23 L51 L151 M405
KD032	R1b1a1a2a1a1	CTS894 PF6430 PF6438 M405
KD034	R1b1a1a2a1a2c1a4b2c	F69 PF6428 L23 S245 Z260 S471 S883
KD036	I1a2	CTS11534 CTS136 FGC2426 Z2727 FGC7747 S244
KD038	I2a2b	L758 PF3665 PF3666 FI4 L181 L35 L39 L272.3
KD040	R1b1a1a2a	M3637 M3647 P131 CTS5884 PF6435 L478
KD041	R1b1a1a2	P128 PF5887 P282 F47 L500 PF6485

8.3.5 Principal component analysis

Principal component analysis (PCA) shows the Bronze Age, Iron Age and Anglo-Saxon individuals plotting closest with the European Late Neolithic and Bronze Age individuals (Figure 8.3).

8.3.6 ADMIXTURE analysis

The lowest cross-validation error obtained for the ADMIXTURE run of the shot-gun sequenced individuals was K=12 (Figure 8.4). Using the shot-gun sequences, only two individuals (both Bronze Age; KD003 and KD041) had enough SNPs overlapping with the published dataset on which ADMIXTURE could be run. The remaining Bronze Age, the two Iron Age samples and the single Anglo-Saxon sample did not produce enough SNPs and were, therefore, excluded from the analysis.

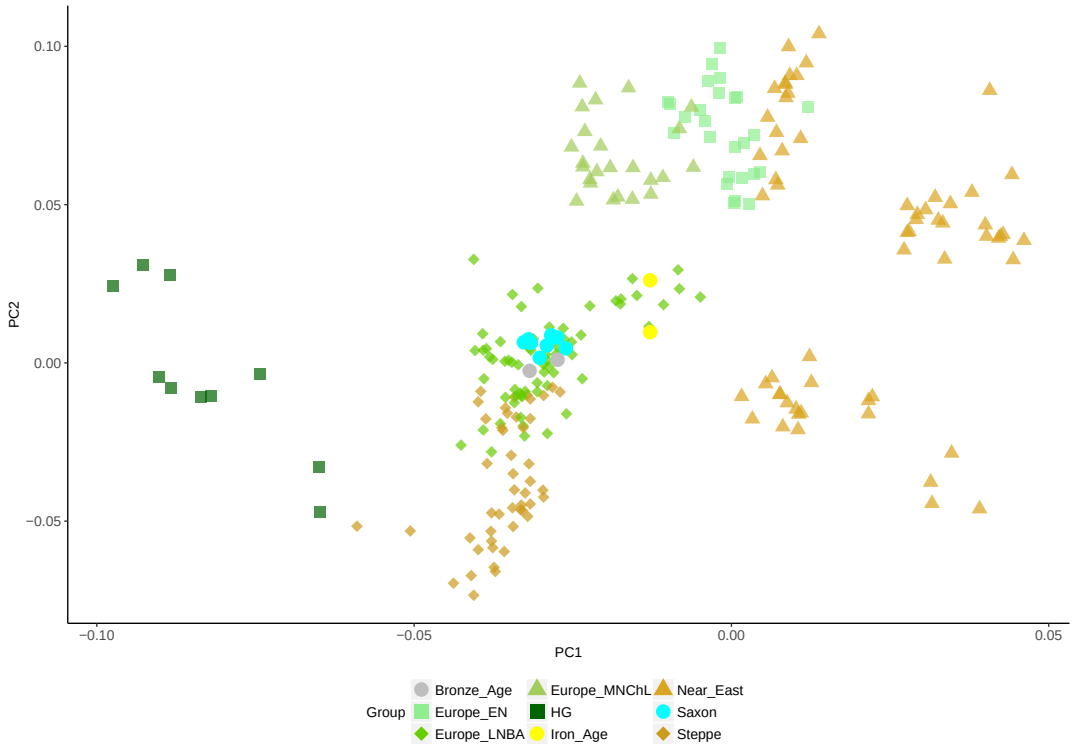

FIGURE 8.3: Principal component analysis of the individuals from West Heslerton, together with previously published ancient samples of Lazaridis et al. (2016).

8.4 Discussion

8.4.1 Sequencing success

All samples from all time periods of West Heslerton and Knapton Wold were successfully sequenced. Of the four Bronze Age individuals, three were determined as male and one as female. Although one sample (KD039) did not produce enough coverage of the mitochondrial genome to determine its haplogroup, the other three (KD003, KD040 and KD041) belonged to mitochondrial haplogroups T2e, T2b4h and U5a1a2a, respectively. The determined Y-chromosome haplogroups for samples KD040 and KD041 were predicted as R1b1a1a2a and R1b1a1a2.

8.4.2 The Bronze Age individuals from West Heslerton

The mitochondrial haplogroup T2e dates to ∼11 kya and appears to be predominantly Mediterranean and southern European (Pala et al., 2012). Its distribution ranges as far as Iceland and Scandinavia in northern Europe, as well as, with low frequency, the Near East. It has already been found in Mesolithic (Bramanti et al., 2009) and Neolithic Germany (Haak et al.,

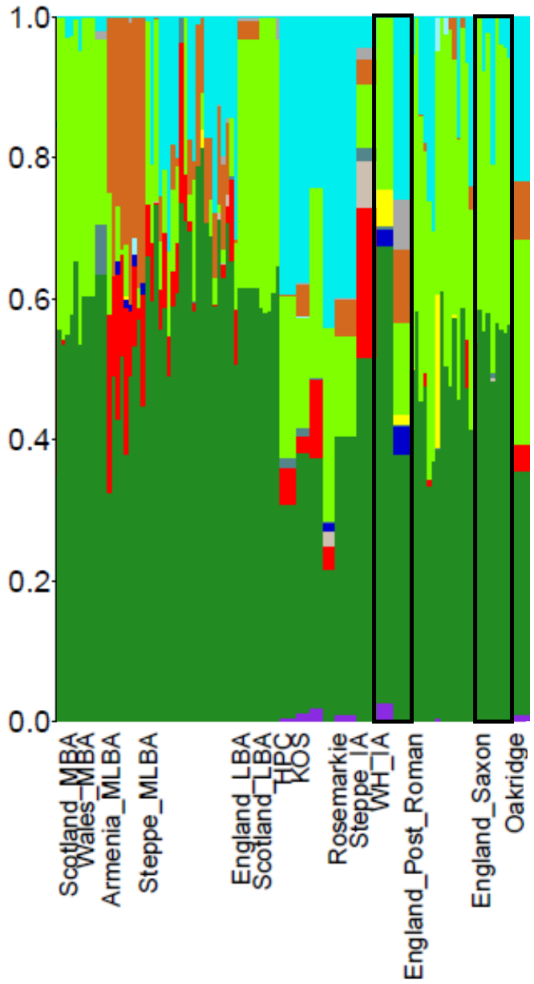

FIGURE 8.4: ADMIXTURE analysis of the individuals from West Heslerton, with previously published ancient samples of Lazaridis et al. (2016) and Olalde et al. (2018) for K=12. Abbreviations used: E: Early, M: Middle, L: Late, BA: Bronze Age, IA: Iron Age, WH: West Heslerton, HPC: High Pasture Cave, KOS: Knowe of Skea.

2015), Neolithic Scotland (Tulloch of Assery B, Highland), Middle Bronze Age England (Amesbury Down, Wiltshire) and in a Bell Beaker burial from Poland (Olalde et al., 2018). The Polish Bell Beaker individual also carried Y-chromosome haplogroup R1b1a1a2, similar to the two Bronze Age males from West Heslerton. Haplogroup T2b is considered to encompass about half of T2 amongst Europeans (Pala et al., 2012). It dates to ∼10 kya and is predominantly European. Although most of the larger subclades of T2b are predominantly found in Europe, gene flow into the Near East has been detected, with T2b4 being a clear example which includes small derived clades that have spread as far as Nepal (Pala et al., 2012). T2b is suggested to have

dispersed into Europe during the early Neolithic period and it has been identified in multiple Neolithic remains from Southern, Central and Northern Europe (Haak et al., 2005; Bramanti et al., 2009; Malmström et al., 2009; Haak et al., 2010; Lacan et al., 2011a; Lacan et al., 2011b; Allentoft et al., 2015; Haak et al., 2015; Mathieson et al., 2015; Lazaridis et al., 2017; Lipson et al., 2017; Nikitin et al., 2017; Fernandes et al., 2018; Mathieson et al., 2018; Mittnik et al., 2018; Olalde et al., 2018). To date, two ancient mitochondrial genomes belonging to T2b4 have been reported. One Late Bronze Age male from Srubnaya, Russia, with Y-chromosome haplogroup R1a1a1b2 (Mathieson et al., 2018), and one Middle Bronze Age male from Baston, Lincolnshire, who belonged to Y-chromosome haplogroup R1b1a1a2a1a2 (Olalde et al., 2018).
U5 is the most ancient European mitochondrial haplogroup and arose in the Upper Palaeolothic among the first settlers (Malyarchuk et al., 2010b). Subclade U5a dates to ∼16-20 kya. The origin of this cluster has been indicated to be eastern Europe, due to the presence of old U5a1 lineages there (Malyarchuk et al., 2010b). Four ancient individuals belonging to U5a1a2a have previously been reported. One Late Neolithic German individual (Haak et al., 2015), two Bronze Age individuals from England and Russia (Allentoft et al., 2015; Olalde et al., 2018), and one individual from the Hungarian conquest period (Neparáczki et al., 2017). The reported Bronze Age individual from Hasting Hill, Sunderland, was determined as male and its Y-chromosome haplogroup was R1b1a1a2a1a2.

8.4.3 The Iron Age individuals from Knapton Wold

The two analysed Iron Age individuals were thought to be related based on their burial arrangement. They were both determined as genetically female, which supports the osteological analysis, and both shared the same mitochondrial haplogroup, H1b1+16362. However, unfortunately their genome coverage was too low to use READ analysis to detect kin relationship. To date the only ancient individual reported to belong to the same mitochondrial haplogroup has been from Bell Beaker Germany (Olalde et al., 2018). The dietary stable isotope analysis of these two individuals showed that, as well as their diets differing from other Iron Age sites in Yorkshire, they themselves were eating surprisingly different diets from each other. Although the carbon values are very variable across the sites included in the comparison, apart from KD071 the nitrogen values are similar, which is indicative that all individuals had a similar amount of protein intake in their diet. The high nitrogen value seen in KD071 may be due to an increased marine input into

her diet. The variation in carbon is most likely explained by the amount of fish in the diet at each site, with KD072 having the least amount. The difference seen between KD071 and KD072 is very striking, but requires analysis of more individuals, as well as comparative faunal data, to be studied in more detail.

8.4.4 The Anglo-Saxon individuals from West Heslerton

Of the analysed Anglo-Saxon individuals (one shot-gun sequenced individual and eight SNP capture sequenced individuals), four were determined as females and five as males. The determined mitochondrial haplogroups were H1b, H1c1, H1q, H6a1a, K1a4a1a+195, K1d1, T2b2b, U5a1a1e, and X2b4a. Thus, the majority of the Anglo-Saxon mitogenomes belong to haplogroup H, and amongst those, with one exception of H6a1a (KD002), to H1. H1 is the most common and widespread subclade of haplogroup H in Europe (for more details on haplogroup H, specifically H1 and H6, please refer to Appendix C).

Haplogroup U5a1a1e (KD035) is a subhaplogroup of U5a that corresponds to the end of the LGM, ∼18 kya (Malyarchuk et al., 2010b), and so far no ancient samples belonging to this subclade have been reported.

For haplogroup K1 the only reported ancient sample basal to K1d has been found in a Bell Beaker from the Netherlands (Olalde et al., 2018). K1a4a1 on the other hand, has been reported a total of 14 times, from Neolithic Scotland (Raschoille Cave, Oban, Argyll and Bute) and Spain (Lipson et al., 2017; Mathieson et al., 2018; Olalde et al., 2018), and Bronze Age Czech Republic, England (Amesbury Down, Wiltshire), Germany, Poland, Portugal, and Spain (Allentoft et al., 2015; Lipson et al., 2017; Knipper et al., 2017; Olalde et al., 2018).

H1b has only been found in one Bronze Age individual from Poland (Mathieson et al., 2018) and one individual from Medieval Alemmanic Germany (O'Sullivan et al., 2018). H1c has previously been reported in Neolithic England (Totty Pot, Cheddar, Somerset), Scotland (Distillery Cave, Oban, Argyll and Bute), and Sweden (Skoglund et al., 2014; Olalde et al., 2018), as well as Early and Late Bronze Age in England (Windmill Fields, Stockton-on-Tees, North Yorkshire), Germany and Latvia, respectively (Knipper et al., 2017; Mittnik et al., 2018; Olalde et al., 2018). H1q has been reported from Bell Beaker Spain (Olalde et al., 2018).

Mitochondrial haplogroup X2b4a has been found in two Bell Beaker males from France, who also both belong to the Y haplogroup/type R1b1a1a2a1a2

(Olalde et al., 2018), whereas KD034 belongs to a more derived Y-chromosome lineage of R1b1a1a2a1a2, namely R1b1a1a2a1a2c1a4b2c.
Y-chromosome haplogroup R1b lineages are associated with the Bronze Age migrations from the steppe, whereas haplogroup I1a is thought to have originated and spread from France (Rootsi et al., 2004). With the exception of two individuals (KD036, KD038) the analysed Anglo-Saxon males all belonged to R1b1a1a2 or one of its subhaplogroups (KD031, KD032, KD034, KD040 and KD041). Subhaplogroup I2a2b is one of the less frequent clades of Y-chromosome haplogroup I, which is an entirely European haplogroup (Rootsi et al., 2004).

8.4.5 Genome-wide analysis

In the PCAs, all analysed individuals from West Heslerton clustered within the variety of the European Late Neolithic/Bronze Age samples. Admixture analyses showed a more nuanced picture. The two Bronze Age individuals that had enough coverage to run admixture looked very similar to the European Late Neolithic/Bronze Age individuals. However, the SNP captured Anglo-Saxon individuals looked more similar to English and Scottish Late Bronze Age individuals, as well as to the post-Roman individuals from Worth Matravers, Dorset (Chapter 7). The two Iron Age individuals showed similar ancestral components as the English and Scottish Late Bronze Age individuals, however one looked more similar to the Iron Age individuals from the High Pasture Cave, Scotland and Knowe of Skea, Orkney.

Chapter 9

Archaeological evidence for facial mutilation in Anglo-Saxon Oakridge, Hampshire

9.1 Introduction

There is extensive literate relating to body mutilation, intentional or incidental. The evidence covers a wide time span, from dynastic Egypt up to modern times. Locations are also widespread, including the Far East, Asia, North and South America, the Middle East, Africa and Europe. Such bodily mutilation applies to the limb extremities, the head and especially the face. The human face is defined by the morphology of underlying bone, muscle and fat. However, soft tissues such as those forming the ears, nose, and lips, coupled with eye colour and hair colour serve to make each face distinct. They also facilitate social communication, allowing expression of emotions. The significance of normal facial appearance is also reflected in ancient codes of law. These introduced many forms of bodily mutilation as punishment for various crimes including amputation of one or both hands, or the slicing or cutting off of noses, ears and lips.

Whilst there is some archaeological evidence for intentional mutilation of limbs (Verano et al., 2000), there is limited archaeological evidence for facial mutilation, largely because it can affect soft tissues without leaving traces on the skeleton. Early excavations on a Romano-British site at Lowbury Hill, Berkshire, uncovered a skeleton supposedly with mutilated facial bones (Atkinson, 1916), but this may have been due to the effects of taphonomic change (Bennike, 2003). Possible evidence of scalping has been reported at the Romano-British sites of St Albans (Mays et al., 1996) and Wroxeter (Barker, 1981). The Wroxeter case appears to be an example of possible

frontal scalping, whilst the St Albans case is now considered to be the result of post-mortem defleshing, on account of the large number (> 90) of presumed cut marks over the whole skull surface. Here, genetic analysis was undertaken on an archaeological human skull that appeared to exhibit evidence of facial mutilation in the form of total removal of the nose and partial removal of the upper lip.

9.2 Materials

A skull was recovered during rescue excavations at a housing development at Oakridge, Basingstoke, England.

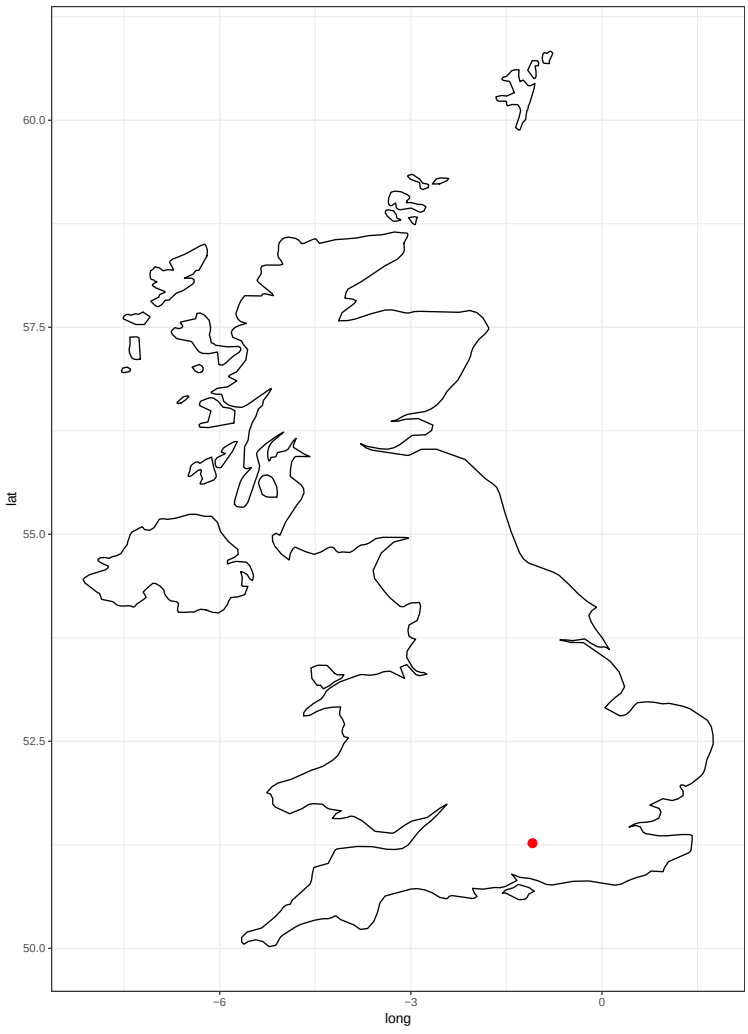

FIGURE 9.1: The location of the rescue excavation at Oakridge within the British Isles.

Only restricted archaeological intervention was permitted, although, after accidental discovery in a drainage pit, a Romano-British burial was fully excavated and recorded (Oliver, 1992). For some unknown reason, a skull recovered from the spoil heap was not analysed at the time, and it was not cleaned, so that the skull interior, the orbits, nasal cavity, and tooth sockets were completely filled with a very fine dark brown sandy silt. This was significant as it allowed the skull to be associated with a specific feature exposed in one of the drainage pit walls. The cleaned skull is shown in Figure 9.2.

9.3 Methods

The skull was assessed using standard osteological techniques to determine age at death and biological sex. The latter was tested by genetic analysis.
The left maxillary M1 was selected for ancient DNA analysis, in order to preserve the skull. Sample processing was undertaken at the Ancient DNA Facility of the University of Huddersfield under dedicated clean-room conditions and followed the standard protocol previously described in Chapter 2. The tips of the roots were cut off using a cutting saw attached to a hobby drill. A drill piece was used to powder the inside of the roots and the powder was collected for DNA extraction. DNA was extracted from approximately 150mg of the sample tooth powder produced, following the protocol by Yang et al. (1998) with modifications by MacHugh et al. (2000). Blank controls were included throughout extractions, library preparation and amplification steps to keep account for possible modern DNA contamination (Appendix D). Three dual-indexed next-generation sequencing libraries were constructed, quantified (Table 9.1) and pooled using the methods outlined in Chapter 2.

TABLE 9.1: DNA concentrations measured with the Qubit 3.0 Fluorometer.

ID	concentration in ng/µl
USER Blank	1.4
CE005.1	6.74
CE005.2	14
CE005.3	4.73
PCR Blank	0.882
Negative Control Library	0.992
PCR Blank	0.812

FASTQ files of the libraries were received with the indexes already trimmed off, a standard procedure of the sequencing service in Liverpool. The NGS reads were trimmed using cutadapt version 1.13 (Martin, 2011), allowing

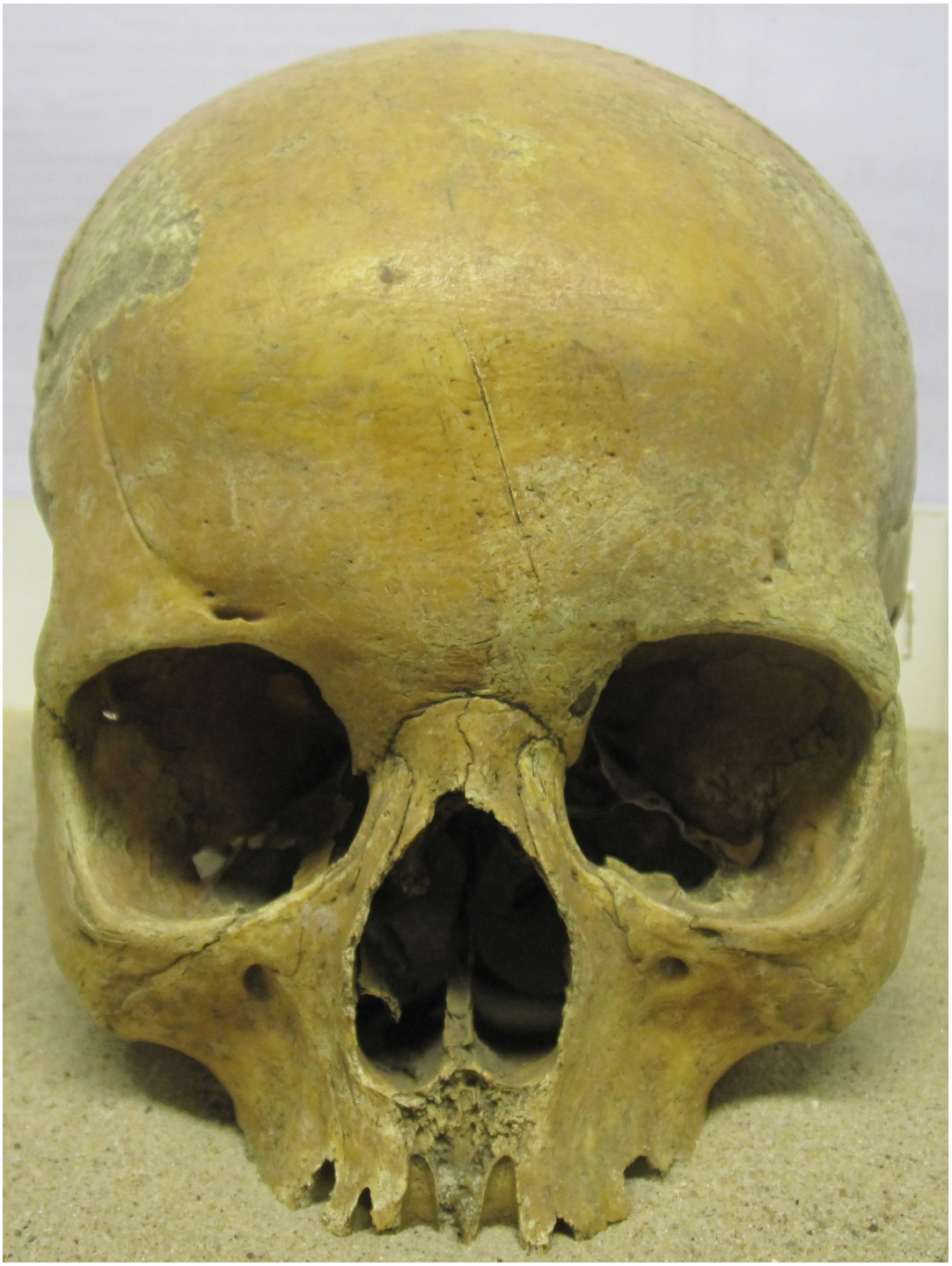

FIGURE 9.2: Norma frontalis of cleaned skull showing well preserved frontal zone, with taphonomic surface degradation on the external surface of the left and right parietals, and the occipital. Figure reproduced from an original by Garrard Cole.

a minimum overlap of 1 bp between read and adapter (adapter sequence: AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC), and discarding reads shorter than 30 bp. All further data processing steps followed the pipeline as described in Chapter 2. NGS libraries were sequenced on an Illumina

HiSeq2500 (1x100bp) by NBAF Liverpool. To assess the previously described anti-contamination measures, negative controls were introduced during each stage of sample processing as described in Appendix D. The two main patterns of typical ancient DNA post-mortem degradation, over-representation of C to T changes at the end of reads and short sequence length, were assessed using mapDamage 2.0 (Jónsson et al., 2013) and bamdamage (Malaspinas et al., 2014). Reads were only taken into account with a minimum mapping quality of 30. In order to detect possible modern mitochondrial contamination, schmutzi (Renaud et al., 2015) was used.
Genetic sex was determined using the software published by Skoglund et al. (2013), which calculates the ratio of reads aligning to the X and Y chromosome. The mitochondrial haplotype was determined by alignment of the sequence reads to the revised Cambridge Reference Sequence (rCRS; NC_012920.1; Andrews et al. (1999)), using BWA filtering (with -b and -q 30) and duplicate reads were removed. The mtDNA consensus sequence was called using Samtools mpileup, as in Martiniano et al. (2016). Haplofind (Vianello et al., 2013) was used to identify haplogroup defining mutations and assign the haplogroup of the sample. The mtDNA coverage was estimated using EAGER (Peltzer et al., 2016).
A powdered sample was extracted from the inner table of the occipital bone in order to determine a radiocarbon date for the specimen. This was processed using the standard Oxford Laboratory protocol (Ramsey et al., 2002; Ramsey et al., 2004b; Ramsey et al., 2004a), and the resultant data was calibrated using OxCal v4.1.7 (Ramsey et al., 2010). In addition, the same left maxillary M1 that had been selected for the ancient DNA analyses was also sampled for strontium, carbon, nitrogen and oxygen stable isotopic analyses.

9.4 Results

The skull was clearly that of a young person, as the maxillary third molars were just erupting from their crypts. The only *in situ* erupted tooth was the left first maxillary molar, all other sockets were present but with the relevant teeth lost post mortem. These features in combination indicate that the individual died at a developmental age of about 18 years. Given the young age of the individual and the absence of the pelvis, it was not possible to determine the sex of the individual with any confidence. The skull had rounded features, with prominent frontal bossing and a gracile zygomatic arch, which

is typical of females, but young males may also exhibit such features prior to full expression of sexual dimorphic traits.

There was a linear cut on the medial left frontal bone. The cut was found to have a v-shaped base, and was oriented obliquely to the right when facing the skull. It had clearly been made peri-mortem. Secondly, the removal of sediment revealed exposed trabecular bone at the base of the nasal aperture, cutting through the front of the maxillary central incisor sockets. The anterior nasal spine was also missing. The trabecular zone was surrounded by a margin of sharply defined cortical bone, especially at the base of the nasal aperture. The lower left and right lateral margins of nasal aperture were truncated. Overall there appeared to be one or more straight cuts through the nasal margin from mid way up the nasal aperture through the anterior nasal spine to the prosthion. Close examination of the lateral aspect revealed a sharp v-shaped nick on the left side of the nasal aperture with the cut through the lips at a slightly different angle to that through the nasal margins. This suggests that at least two cuts were made to inflict the injuries (see Figure 9.3).

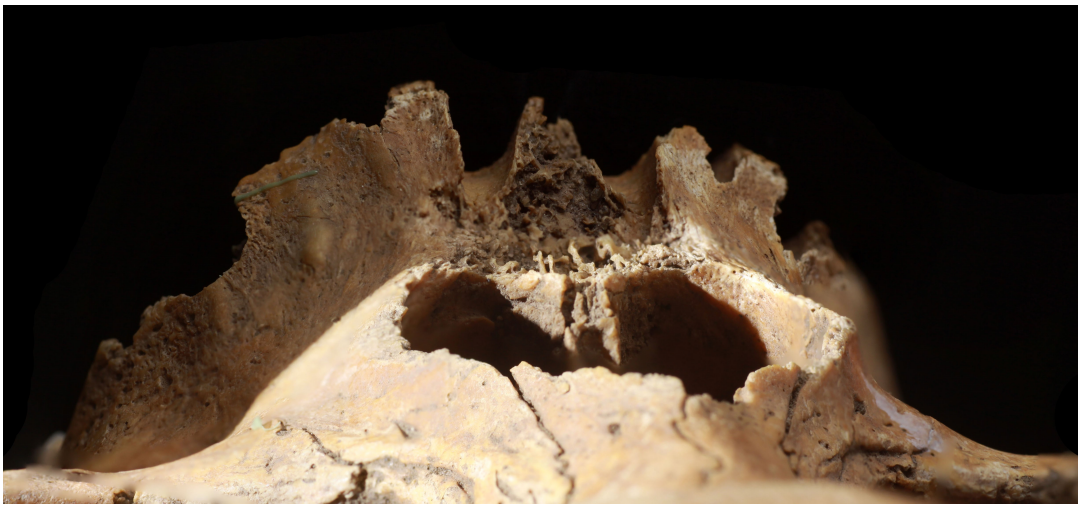

FIGURE 9.3: View along nasal aperture showing linearity of cut on the base and right side and sharp nicks around the apex of the left central incisor.

9.4.1 Sequencing results

Sequencing results are displayed in Table 9.2.

TABLE 9.2: Summary of sequencing results.

ID	CE005
Genome coverage	1.0143x
Genome std coverage	3.3005x
Mean mapping quality	12.029
GC content in %	45.61
No of reads	65698100
No of mapped reads	65698100
No of mapped bases	3.26E+09
Duplication rate in %	2.67

9.4.2 Radiocarbon dating

The uncalibrated radiocarbon date (OxA-26646) was determined to be 1173 ± 24 years bp and the calibrated 2σ date range was found to lie between 776-899 (87.3%) and 920-946 (8.1%) (see Figure 9.4).

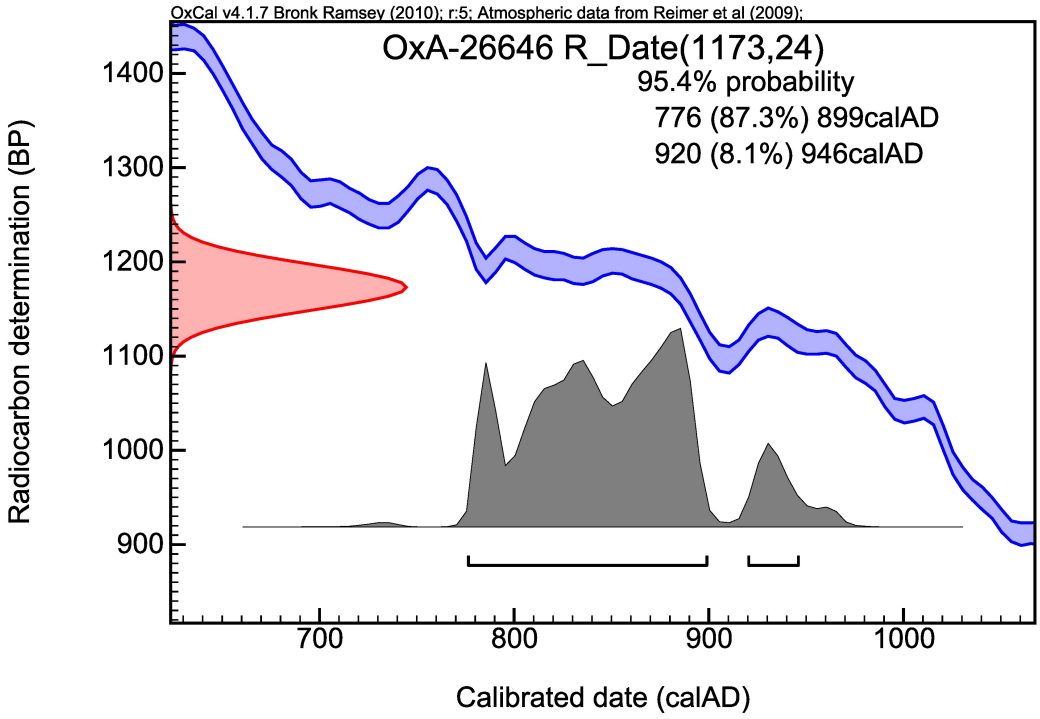

FIGURE 9.4: Calibrated radiocarbon dates based on an uncalibrated date of 1173 +/- 24 years bp.

9.4.3 Dietary stable isotope analysis

The collagen prepared from the tooth dentine gave a C/N ratio of 3.1, suggesting that the collagen was well preserved and likely to give a good measure of the in-vivo dietary carbon and nitrogen stable isotopic values (Van Klinken, 1999). The δ^{15}N value was 11.71‰ relative to AIR, and the δ^{13}C value was -19.03‰ relative to VPDB, both with error of ± 0.1‰. The δ^{13}Cc value was -12.78‰ ± 0.01 and the δ^{18}Oc value was -6.47 ± 0.02 both relative to VPDB. The $^{87/86}$Sr ratio was 0.710279 ± 0.000014 (Figure 9.5).

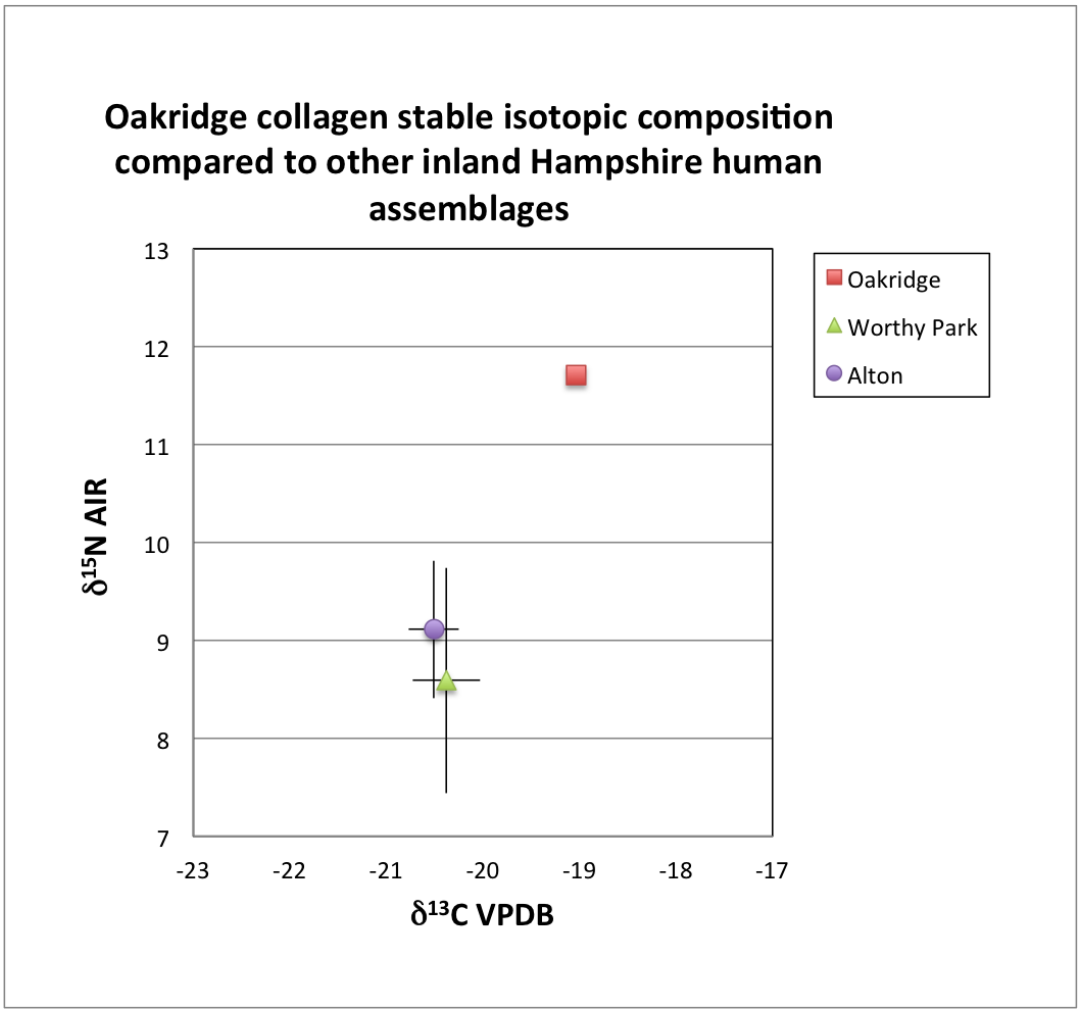

FIGURE 9.5: Dietary stable isotope estimates of the Oakridge individual in comparison to dietary stable isotope values of individuals from Worthy Park and Alton.

9.4.4 DNA damage and contamination estimation

The endogenous DNA content retrieved from the sample was 70.08%. A misincorporation pattern typical for aDNA was observed, with reads showing

an increase in C to T and G to A transitions towards the 5′ and 3′ ends, characteristic changes of aDNA molecules (Briggs et al., 2007; Brotherton et al., 2007). Estimates of mitochondrial contamination were very low, at a range from 0.12-0.17%.

9.4.5 Mitochondrial genome analysis

The individual was determined as female. The mitochondrial haplogroup was determined to be H3g1a (Results in Table 9.3).

TABLE 9.3: Summary of mitochondrial genome coverage, haplogroup assignment and haplotype of the Oakridge sample

Sample ID	Coverage	Haplogroup	Haplotype
CE005	42.2919x	H3g1a	152C 263G 750G 1438G 2523Y 3992T 4135C 4418C 4769G 6776C 8251A 8860G 10754C 15326G 15377G 16519C

The phylogenetic tree of mitochondrial haplogroup H3 was reconstructed for an analysis described in Appendix C. The branch of H3g1 was reconstructed using published data (Figure 9.6).

9.5 Discussion

9.5.1 Dating

The skull gave a date of late 9th century AD (between cal. AD 767 and 970). This date range covers that of a documented battle, the Battle of Basing, which occurred between Danes and local Saxons on the 22nd of January AD 871 (King et al., 2016). It is possible that the victim could either have been punished after being captured by the opposing side, or punished for collaborating with the enemy.

9.5.2 Stable isotope analysis

The calculated δ^{18}Ow value for the tooth enamel was -9.22‰ SMOW, whereas modern ground water δ^{18}O values for the area where the sample was found are around -7‰ SMOW (Darling et al., 2003). The conversion of measured

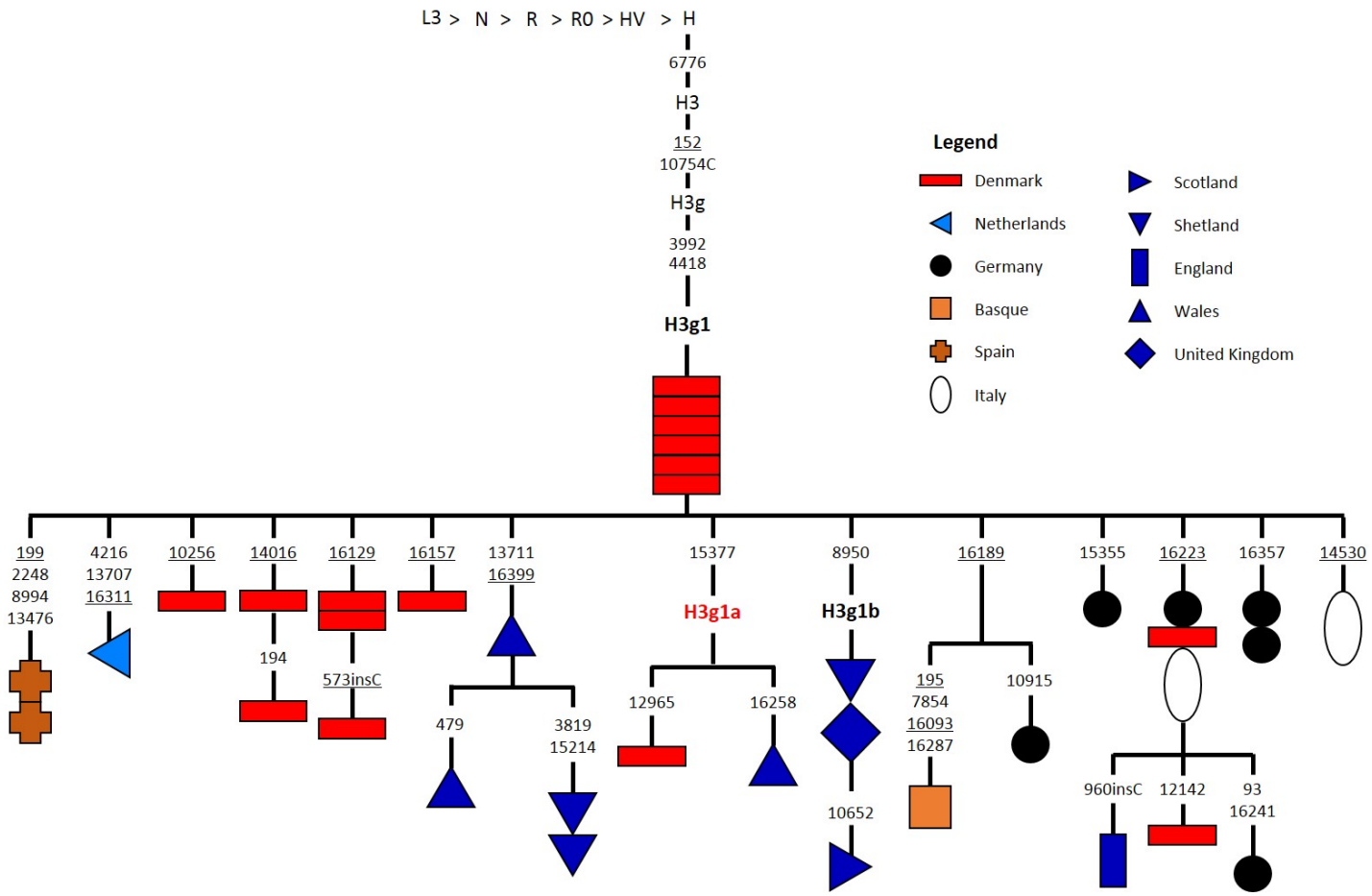

FIGURE 9.6: Schematic of the branch of H3g1 within the phylogenetic tree of H3. Samples originating in the British Isles are coloured in dark blue. The reconstruction of the phylogenetic tree only shows modern samples.

carbonate values to theoretical phosphate values and, thence to probable water values, is fraught with uncertainty (see Pollard et al. (2011) for discussion); however, in this instance, there appears to be a considerable difference between the likely drinking water composition at the time of tooth mineralisation, and the modern ground water composition at the location where the sample was excavated, suggesting that this individual was likely to be non-local. The strontium isotopic composition of tooth enamel is strongly influenced by the underlying geology at the time of tooth mineralisation (Chenery et al., 2010). The tooth enamel $^{87/86}$Sr ratio value measured from the Oakridge individual was 0.710279 ± 0.000014, which falls well outside of the range for the local chalk and just outside of the range for the nearby Tertiary sediments. This also suggests that this individual was not local; that is, they did not spend their early years, during the time of M1 mineralisation, in the immediate area around Basingstoke.

The carbon and nitrogen stable isotopic results, of δ^{13}C -19.03‰ (VPDB) and δ^{15}N of 11.71‰ (AIR) respectively, suggest a predominantly terrestrial diet with a minor marine protein component. However, in the absence of any associated faunal remains from the same location, it is impossible to give an absolute indication of what these values might mean in terms of the relative amounts of plant versus terrestrial animal versus marine protein consumed. Comparable local dietary isotopic data sets are available from Worthy Park near Winchester, Hampshire (ca. 25 kilometres southwest of Basingstoke), and Alton, Hampshire (ca. 15 kilometres to the southeast) (Hull, 2008). These sites have yielded mean δ^{13}C values for adult human bone collagen of -20.38‰ and -20.51‰, with a standard deviation of 0.35 (n = 26) and 0.26 (n = 25), respectively, and mean δ^{15}N values of 8.59‰ and 9.11‰, with a standard deviation of 1.15 (n = 26) and 0.70 (n = 25), respectively.

By comparison, the values for the Oakridge individual are higher in both δ^{13}C and δ^{15}N than the means reported for the Worthy Park and Alton Saxon burials (Figure 9.5). This suggests that the diet of the Oakridge individual was slightly anomalous compared to the average diet that might be expected for inland Hampshire in middle Saxon times, with the less negative value for δ^{13}C and the more positive value for δ^{15}N suggesting more of a marine component in the Oakridge diet than usual for inland Hampshire. Taken alongside the strontium and oxygen isotopic data, the dietary stable isotopic information is also compatible with a non-local origin for this individual.

9.5.3 Mutilation

The skeletal remains consisted of the partial skull of a juvenile individual exhibiting dramatic trauma. For males, this is most likely to have happened as a result of fighting. For females, it was a most likely a punishment for (alleged) adultery. As morphological indicators were unable to identify sex (mainly due to partial remains and young age of the individual), the sex was determined genetically as being from a female. If she was involved with the Battle of Basing, it is possible that she was punished for collaborating with the enemy. The battle option also raises the question of origins, Dane or Saxon? Unfortunately, it was not possible to analyse which of these populations she belonged to as Danish and Saxon people are very similar genetically, as the genomic coverage was too low. As the individual was female, then it seems more probable that she was a Saxon, as it would be less likely to have a 'foreign' female fighting in a battle.

9.5.4 Maternal lineage

She had mitochondrial haplogroup H3g1. Today, over 40% of all maternal lineages in Europe belong to haplogroup H. H3 has high frequency in modern-day Spain (Achilli et al., 2004) and in the Western Isles of Scotland (Brotherton et al., 2013), and is also present in the middle Neolithic period (including at Orkney) and the Bell Beaker culture in Europe (Brotherton et al., 2013; Olalde et al., 2018). H3g1 is generally seen in modern-day northwestern European people, with the root comprising solely of people from Denmark (n = 6; Figure 7), although English, Shetland, Scottish, Welsh, German and Dutch people also group within the H3g1 clade. The Oakridge sample, with a mitochondrial haplotype of H3g1a, groups together with modern people from Denmark and Wales (n = 1 for each), so, unfortunately, it is not possible to say anything about the origin of the female based on her maternal lineage.

9.6 Conclusions

This chapter presents evidence of facial mutilation to a young adult female around 18 years of age. She was probably not local to the Hampshire area, but perhaps was from northern England or southern Scandinavia. Since the physical evidence was observed only from the skull, and there is no evidence

for or against the infliction of injuries elsewhere on the body, the exact circumstances behind the infliction of the mutilating wounds is not known. The limited 'rescue' excavation means that there is no evidence whether the body was even deposited along with the skull. The injuries present are consistent with documented punishment of females for adultery. Based on historical evidence, other possible options include a brutal attack arising from jealousy, or abuse of a hostage. Irrespective of the circumstances behind the incident, this case appears to be the first archaeological evidence for this form of brutal mutilation in Anglo-Saxon England.

Chapter 10

Discussion

The following is a brief overall discussion of the findings reported in this thesis, as more detailed discussions are included in each chapter. A total of 77 bone and teeth samples, covering the Neolithic, Bronze Age, Iron Age, Pictish, post-Roman/Anglo-Saxon, and Viking periods (Table 10.1), were processed to investigate human dispersal by means of the maternal and paternal lineages of descent, as well as using autosomal data for admixture and kinship analyses. The samples analysed in the present study have given new insights into the genetic makeup of the British population at various time periods, and allowed a direct comparison across periods and locations (Figure 10.1).

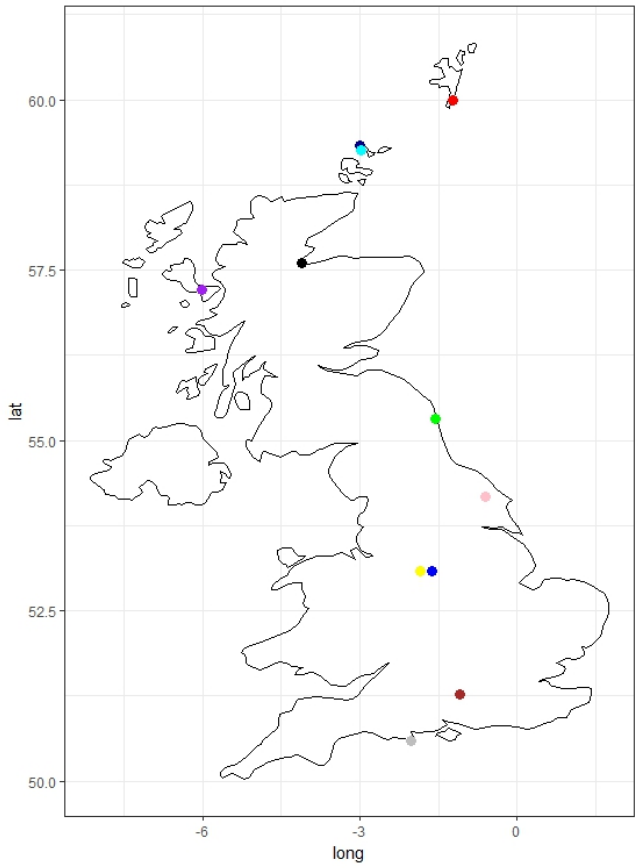

FIGURE 10.1: All sampling sites within the British Isles. Red - Milla Skerra, Darkblue - Links of Noltland, Cyan - Knowe of Skea, Black - Rosemarkie Cave, Purple - High Pasture Cave and Strathglebe, Green - Low Hauxley, Pink - West Heslerton, Grey - Worth Matravers, Yellow - Carsington Pasture Cave, Blue - Beeston Tor, Brown - Oakridge.

TABLE 10.1: Overview of main archaeological and genetic data of all analysed individuals. DNA sample gives the type of remains the DNA sample was taken from. Analysis type key: L - Liverpool, dual-index shot-gun sequencing; M - Macrogen, single-index shot-gun sequencing; R - Reich, Harvard Medical School, SNP capture. Sex denotes the genetic sex determination as XX/XY and the osteological sex determination as M, F, – for male, female and undetermined. Era key: N - Neolithic; BA - Bronze Age; IA- Iron Age; P - Pictish; pR - post-Roman; AS - Anglo-Saxon; V - Viking. Site code key: BT - Beeston Tor; CPC - Carsington Pasture Cave; HPC - High Pasture Cave; KOS - Knowe of Skea; KW - Knapton Wold/West Heslerton; LH - Low Hauxley; LON - Links of Noltland; MS - Milla Skerra; O - Oakridge; RC - Rosemarkie Cave; S - Strathglebe; WH - West Heslerton; WM - Worth Matravers. Site type funerary comp. denotes funerary complex.

DNA code	DNA sample	Analysis type	mtDNA	Sex	Y-chr.	Era	Site	Country	Site type
CE001	LRM2	L	T2c1d1	XY/M	I	N	CPC	England	cave
CE002	ULP2	L	U8b	XY/–	–	BA/IA	BT	England	cave
CE003	UM2	L	X2b4a1	XX/M	n/a	IA	CPC	England	cave
CE004	LRM2	L	H10b	XY/M	R1b1a1a2a1a	IA	CPC	England	cave
CE005	ULM1	L	H3g1a	XX/–	n/a	V	O	England	–
KD001	petrous	L/M	J1b1a1a	XY/M	R1b1a1a2a1a	P	RC	Scotland	cave
KD002	petrous	L	H6a1a	XX/F	n/a	AS	WH	England	cemetery
KD003	petrous	L	T2e	XX/F	n/a	BA	WH	England	cemetery
KD004	R petrous	L	H1b	XY/M?	R1b1a1a2a1a2c1a2	IA	KoS	Scotland	funerary comp.
KD005	petrous	L	H7a1b	XX/F	n/a	P	HPC	Scotland	cave
KD006	R petrous	L	T2a1b1a	XY/–	I2a1b	BA	LON	Scotland	cemetery
KD007	molar tooth	R	K1a2a	XY/–	R1b1a1a2a1a2c1a2a2a1a1a	pR	WM	England	cemetery
KD008	tooth	R	K1a2a	XY/F?	I1a2a1a2∼	pR	WM	England	cemetery
KD009	tooth	R	T2a1a	XX/–	n/a	pR	WM	England	cemetery
KD010	tooth	R	U5b1	XY/–	E1b1∼	pR	WM	England	cemetery

Table 10.1 continued from previous page

DNA code	DNA sample	Analysis type	mtDNA	Sex	Y-chr.	Era	Site	Country	Site type
KD011	tooth	R	H11a	XX/F?	n/a	pR	WM	England	cemetery
KD012	tooth	R	HV6	–/F?	–	pR	WM	England	cemetery
KD013	tooth	R	HV6	XY/F	R1b1a1a2a1a	pR	WM	England	cemetery
KD014	tooth	R	HV6	XY/M	R1b1a1a2a1a	pR	WM	England	cemetery
KD015	tooth	R	HV6	XX/–	n/a	pR	WM	England	cemetery
KD016	tooth	R	H1bb	XX/F	n/a	pR	WM	England	cemetery
KD017	tooth	R	H1e1a	XY/M	R1b1a1a2a1a	pR	WM	England	cemetery
KD018	tooth	R	H1bb	XY/–	R1b1a1a2a1a	pR	WM	England	cemetery
KD019	tooth	R	U5b2c1	XX/F?	n/a	pR	WM	England	cemetery
KD020	tooth	R	T2a1a	XY/–	R1b1a1a2a1a2c1a	pR	WM	England	cemetery
KD021	tooth	R	H5c	XY/F?	R1b1a1a2a1a1c2b2b1a	pR	WM	England	cemetery
KD022	tooth	R	T2a1a	XY/F	I2a1b1	pR	WM	England	cemetery
KD023	LRM1	M	J1c1	–/–	–	N	S	Scotland	cairn
KD024	LLM2	M	T2b	–/–	–	N	S	Scotland	cairn
KD025	LR?M1	M	H1c	–/–	–	N	S	Scotland	cairn
KD026	UL?M1?	M	U5b2c	XY/–	I	N	S	Scotland	cairn
KD027	L?L?M1?	M	U5b2c	–/–	–	N	S	Scotland	cairn
KD028	LR?M1	M	T2b	–/–	–	N	S	Scotland	cairn
KD029	LLM1	M	H5b	XX/–	n/a	N	S	Scotland	cairn
KD030	URM3	n/a	n/a	–/–	n/a	N	S	Scotland	cairn

Table 10.1 continued from previous page

DNA code	DNA sample	Analysis type	mtDNA	Sex	Y-chr.	Era	Site	Country	Site type
KD031	petrous	R	H1c1	XY/–	R1b1a1a2a1a1	AS	WH	England	cemetery
KD032	petrous	R	H1q	XY/F?	R1b1a1a2a1a1	AS	WH	England	cemetery
KD033	petrous	R	H1b	XX/F?	n/a	AS	WH	England	cemetery
KD034	petrous	R	X2b4a	XY/–	R1b1a1a2a1a2c1a4b2c	AS	WH	England	cemetery
KD035	petrous	R	U5a1a1e	XX/F	n/a	AS	WH	England	cemetery
KD036	petrous	R	K1d1	XY/–	I1a2	AS	WH	England	cemetery
KD037	petrous	R	T2b2b	XX/F	n/a	AS	WH	England	cemetery
KD038	petrous	R	K1a4a1a+195	XY/M	I2a2b	AS	WH	England	cemetery
KD039	petrous	M	–	–/–	–	BA	WH	England	cemetery
KD040	petrous	M	T2b4h	XY/M	R1b1a1a2a	BA	WH	England	cemetery
KD041	petrous	M	U5a1a2a	XY/M	R1b1a1a2	BA	WH	England	cemetery
KD042	R petrous	M	U5a1b1a	XX/F	n/a	IA	KoS	Scotland	funerary comp.
KD043	L petrous	M	H1b	XY/M?	R1b1a1a2	IA	KoS	Scotland	funerary comp.
KD044	LM3	M	U5b2a3	XX/–	n/a	BA	LON	Scotland	cemetery
KD045	U incisor	M	J1c2a	XY/F	I2a1b	BA	LON	Scotland	cemetery
KD046	molar	M	T2a1b1a	–/–	–	BA	LON	Scotland	cemetery
KD047	R petrous	M	H39	XY/M	I2a1a1a1∼	BA	LON	Scotland	cemetery
KD048	L petrous	M	H39	–/–	?	BA	LON	Scotland	cemetery
KD049	R petrous	M	H39	XY/–	I	BA	LON	Scotland	cemetery
KD050	R petrous	M	H39	XX/–	n/a	BA	LON	Scotland	cemetery

Table 10.1 continued from previous page

DNA code	DNA sample	Analysis type	mtDNA	Sex	Y-chr.	Era	Site	Country	Site type
KD051	R petrous	M	n/a	–/–	–	BA	LON	Scotland	cemetery
KD052	R petrous	M	K1a29a	–/–	–	BA	LON	Scotland	cemetery
KD053	R petrous	M	n/a	–/–	–	BA	LON	Scotland	cemetery
KD054	decidious molar	n/a	n/a	–/–	n/a	BA	LON	Scotland	cemetery
KD055	URM3	M	J1c2a	XX/M	n/a	BA	LON	Scotland	cemetery
KD056	UL incisor	no library	n/a	–/–	n/a	BA	LON	Scotland	cemetery
KD057	lower incisor	M	H1n1	XY/M	I	BA	LON	Scotland	cemetery
KD058	R petrous	M	K1a3a	XX/F	n/a	BA	LON	Scotland	cemetery
KD059	lower incisor	M	T2b21	XY/–	I	BA	LON	Scotland	cemetery
KD060	R petrous	M	H1n1	XY/–	I	BA	LON	Scotland	cemetery
KD061	L petrous	M	K1c2	XY/–	R1b1a1a2	BA	LON	Scotland	cemetery
KD062	R petrous	M	U5b2a3	XX/–	n/a	BA	LON	Scotland	cemetery
KD063	L petrous	M	H58a	XX/–	n/a	BA	LON	Scotland	cemetery
KD064	R petrous	M	T2b21	XY/–	I2a1b	BA	LON	Scotland	cemetery
KD065	L petrous	M	H39	XX/–	n/a	BA	LON	Scotland	cemetery
KD066	LLM1	M	T2a1b1a	XX/–	n/a	BA	LON	Scotland	cemetery
KD067	L petrous	M	H+195	XX/–	n/a	BA	LON	Scotland	cemetery
KD068	R ischium	M	J1c3	–/–	–	P	HPC	Scotland	cave
KD069	L scapula	M	H7a1b	–/–	–	P	HPC	Scotland	cave
KD070	L petrous	M	T2e1a	XY/M?	R1b1a1a2	BB	LH	England	cairn, cist

Table 10.1 continued from previous page

DNA code	DNA sample	Analysis type	mtDNA	Sex	Y-chr.	Era	Site	Country	Site type
KD071	R petrous	M	H1b1+16362	XX/F?	n/a	IA	KW	England	cemetery
KD072	L petrous	M	H1b1+16362	XX/F	n/a	IA	KW	England	cemetery
KD073	L petrous	M	J1b1a1	XY/M	–	IA	MS	Scotland	settlement

10.1 Genome-wide analysis

The distinctive geographic position of the British Isles at the north-western extreme of Europe led to later arrival times of agriculture and following cultural packages, such as the Bell Beaker culture, compared to continental Europe. With these cultural changes, major population replacements have been reported for the Mesolithic/Neolithic transition (Brace et al., 2018) and the arrival of the Bell Beakers (Olalde et al., 2018). Nevertheless, it is also clear that the population in the Bronze Age was not exactly the same as it is today (Olalde et al., 2018). Further investigations, especially at the transitions between the Bronze Age and Iron Age, Iron Age to Roman era, and Roman era to the Anglo-Saxon and Viking invasions, are required in order to more fully understand population turnover and admixture in the British Isles through time. The only ancient genome-wide data from these later, post-Bronze Age, periods are studies of a Roman cemetery in York (Martiniano et al., 2016) and Iron Age and Anglo-Saxon sites in south-east England (Schiffels et al., 2016).
ADMIXTURE and PCA plots were calculated using the protocol in Chapter 2, and the resulting figures are shown in Figure 10.2 and Figure 10.3. Only samples that had greater than 10,000 SNPs were included in this analysis, so the total number of individuals was reduced to 55.
Due to the fact that the sequencing data of the recently sequenced Mesolithic and Neolithic individuals of the British Isles (Brace et al., 2018) are not yet publicly available, no direct comparison between my samples and these individuals was possible. However, comparing the Neolithic individual from Strathglebe, Isle of Skye (KD026), with the available published data (Lazaridis et al., 2016; Olalde et al., 2018) and the additional Bronze Age individuals from this study, the greatest similarity of ancestral components appears between the Strathglebe individual and European Middle Neolithic individuals, all of which lack the Caucasus hunter-gatherer Steppe component that is clearly seen in the Bronze Age individuals.
Interestingly, the Bell Beaker individual from Low Hauxley (KD070), which otherwise showed genetic similarity to the Late Neolithic and Bronze Age individuals from Czech Republic, Denmark, England, Estonia, Germany, Hungary, Poland, Scotland, and Sweden (Lazaridis et al., 2016; Olalde et al., 2018), also exhibits a small proportion of North African ancestry (Figure 10.2). As the overall genome coverage of this sample was low, and thus the North African ancestry might conceivably be artificial, this result needs to be treated with caution. However, this interesting result could be verified by additional

sequencing to higher depth and repetitive runs of ADMIXTURE analysis.
Unfortunately, the genome coverage of the Iron Age individual from Sandwick, Unst, Shetland (KD073), was too low to allow any autosomal comparison with the other individuals of this study or the published modern and ancient data. In the PCA, the Neolithic individual from Carsington Pasture Cave in Derbyshire (CE001) (Appendix I) and Strathglebe on Skye (KD026) cluster closest together within the distribution of the European Middle Neolithic individuals, whereas the two Iron Age individuals (CE003 and CE004) cluster with Bronze Age individuals from the Links of Noltland and post-Roman individuals from Worth Matravers (Figure 10.3).
An individual from Strathglebe (KD027) was dated to ∼3,494-3,104 cal. BC, which corresponds to the Scottish Early Neolithic. Together with the Neolithic individual from Carsington Pasture Cave, Derbyshire (CE001), the Scottish Neolithic individual that had enough coverage to be examined in more depth (KD026) showed greatest similarities with published European Middle Neolithic individuals. Comparing the ancestral components of the Neolithic data, which comprise the two individuals presented here, as well as samples from Cassidy et al. (2016) and Olalde et al. (2018), with the Beaker and Bronze Age periods, the arrival of a new population with Steppe ancestry in Europe is evident in the Bronze Age. The impact of this Steppe component is especially strong in the Bell Beaker individual from Low Hauxley (KD070), Bronze Age individuals from the Links of Noltland (KD006, KD044-045, KD047, KD049-050, KD055, KD057-058, KD060-065, and KD067) and West Heslerton (KD003 and KD041), Iron Age individuals from Knowe of Skea (KD004, KD042 and KD043), most of the post-Roman individuals from Dorset (KD007-009, KD011-014 and KD016-022), the Yorkshire Anglo-Saxons (KD002 and KD033-038), and the Oakridge individual (CE005). One of the Iron Age individuals from West Heslerton has an exceptionally large Steppe component (KD072), whereas the Pictish Rosemarkie individual (KD001) had a comparably small portion of this ancestral component.
The individuals from the post-Roman cemetery at Worth Matravers in Dorset showed most similar proportions of ancestral components to modern northern Europeans (Figure 10.2). In general, these data showed a difference in the genetic composition during post-Roman times in England, with one outlier individual (KD010) having ∼50% North African or Near Eastern ancestry, which is shown in a PCA with modern samples (Figure 10.4).
The autosomal analysis of the Iron Age High Pasture Cave (KD005) and the

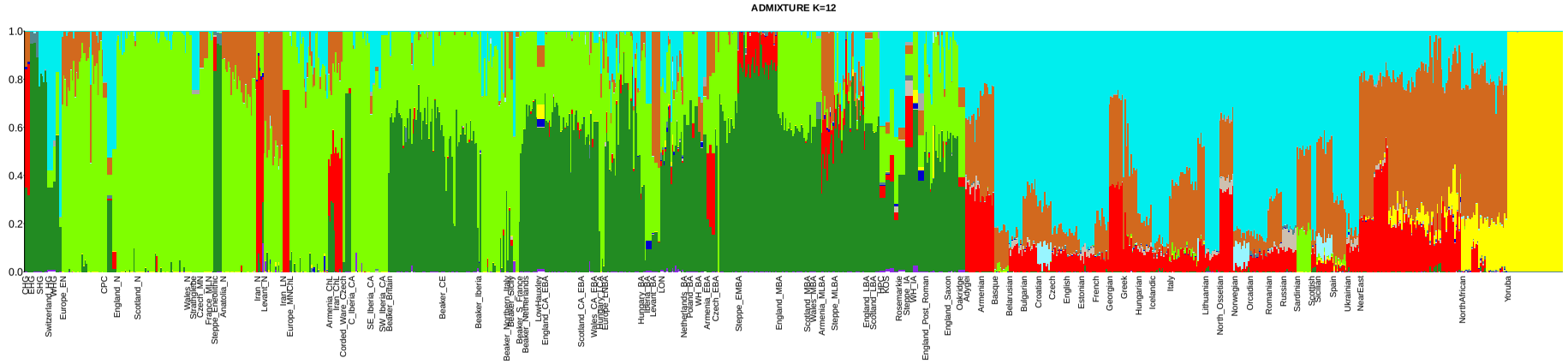

FIGURE 10.2: ADMIXTURE analysis of all studied ancient samples with a selection of the published ancient and modern samples of Lazaridis et al. (2016) and Olalde et al. (2018). Abbreviations used: E: Early, M: Middle, L: Late, HG: Hunter-Gatherer (C - Caucasus, E - Eastern, S - Scandinavian, W - Western), N: Neolithic, ChL: Chalcolithic, CA: Copper Age, BA: Bronze Age, IA: Iron Age, CE: Central Europe, CPC: Carsington Pasture Cave, LON: Links of Noltland, WH: West Heslerton, HPC: High Pasture Cave, KOS: Knowe of Skea.

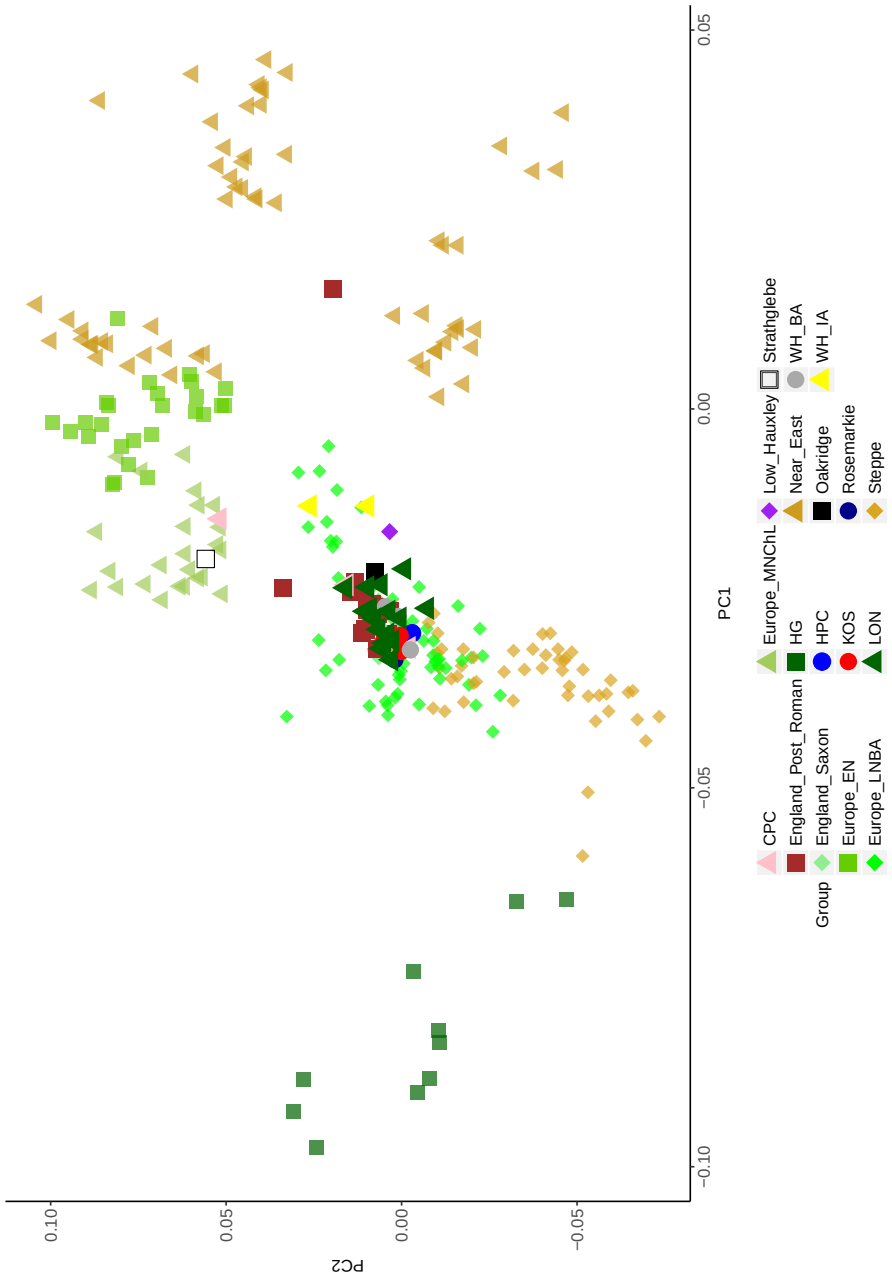

FIGURE 10.3: PCA of all analysed individuals with >10,000 overlapping SNPs projected on top of the published ancient data of Lazaridis et al. (2016).

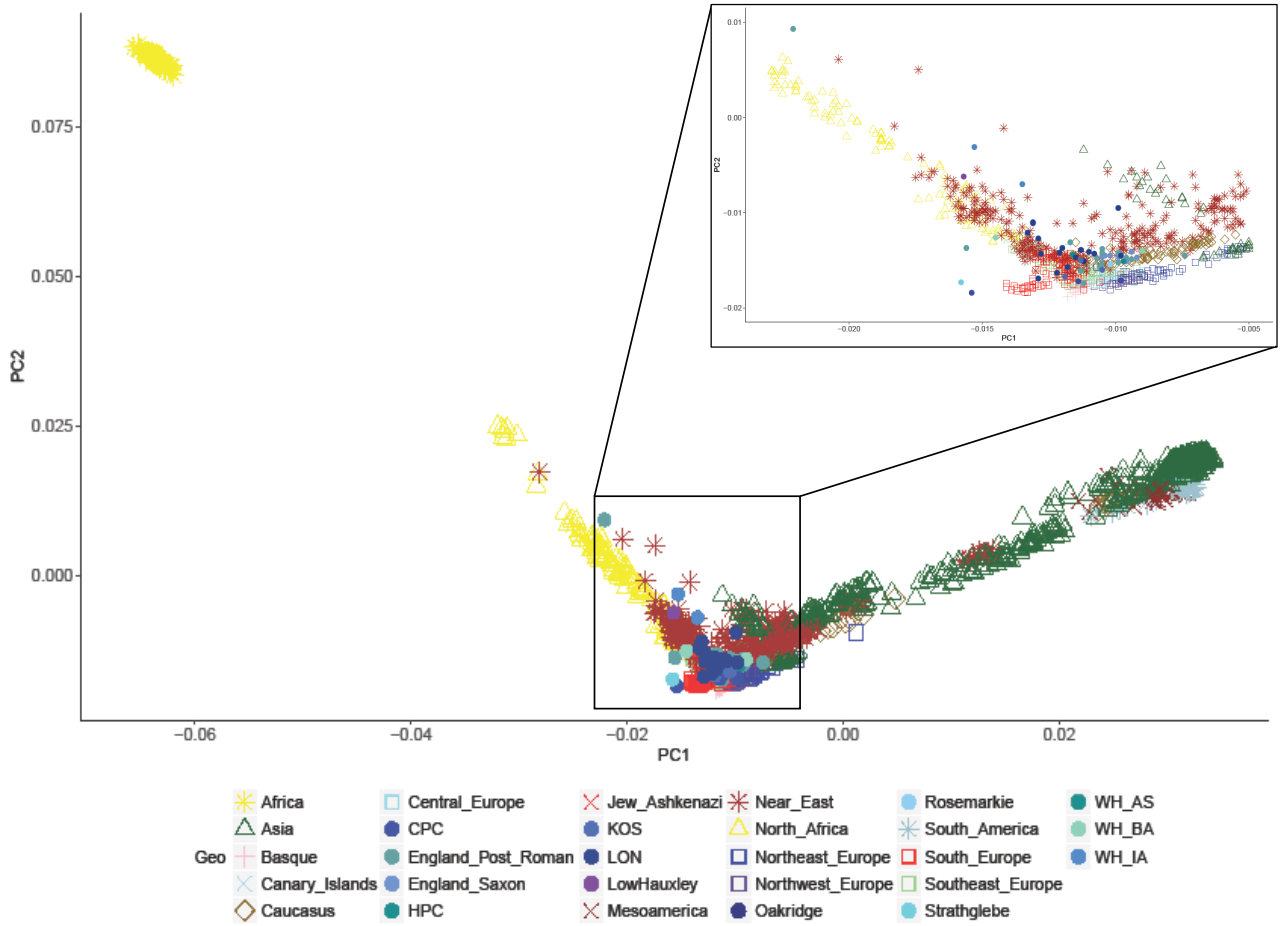

FIGURE 10.4: PCA of all analysed individuals with >10,000 overlapping SNPs projected on top of the published modern data of Lazaridis et al. (2016).

Pictish individual from Rosemarkie Cave (KD001) showed a strong similarity in the proportion of ancestral populations with the Late Neolithic/Bronze Age populations of Europe. The obvious lack of genetic influence through the Roman Empire in Rosemarkie Man can be explained by the general low level of immigration into Scotland at this time. Further studies of the Scottish populations, especially during the Iron Age and adjacent periods, are needed to understand possible differences in genetic admixture between the northern (Scottish, Pictish) and southern (Briton, Anglo-Saxon) populations of the British Isles.

The Anglo-Saxon individuals analysed from the cemetery in West Heslerton, North Yorkshire, were most similar to British Middle Bronze Age individuals and the post-Roman individuals from Worth Matravers, Dorset (Figure 10.2). However, when compared to the English Saxon individuals, the sample from Oakridge, Hampshire (CE005), had similar ancestral proportions as the Iron Age sample from High Pasture Cave and the Pictish Rosemarkie Man, both from Scotland. As archaeologists suspected the individual to be either Viking or Anglo-Saxon, this result requires further investigation and a more detailed STRUCTURE-like analysis, and the addition of samples from Scandinavia may allow a more detailed genetic analysis of whether this young mutilated female was of Saxon or Scandinavian ancestry. The strong similarity between modern north-western and central European populations would make a fineSTRUCTURE (Leslie et al., 2015) approach necessary in order to identify possible differences and give an indication of the origin of the Oakridge individual.

Overall, all ADMIXTURE results presented in this thesis need to be treated with caution in order to avoid over-interpretation of the results, such as, for example, the possible North African component found in the Bell Beaker individual from Low Hauxley. All shot-gun sequenced samples presented have a very low coverage compared to most published ancient data, and the analysis might be skewed. Generally, I found that samples with a minimum of 10,000 SNPs behave as expected when plotted in a PCA or in ADMIXTURE runs. However, repetitive runs and further sequencing are necessary to rule out incorrect clustering and any skewing effects due to the low coverage data (Lawson et al., 2018).

10.2 Mitochondrial DNA haplogroup frequencies

An increase in haplogroup H over time can be observed when comparing the overall mitochondrial haplogroup frequencies across different time periods in the British Isles. Already between the Bronze and Iron Age, the frequency of haplogroup H has increased to similar levels as those seen in the present day population.
Of all published ancient mitochondrial genomes from the British Isles and Ireland, 68 date to the Neolithic. When including the individuals from this study (n = 7), the mitochondrial haplogroup frequency composition consists of 31% of K1, 13% of U5b and J1, 10% of T2, 8% of H1, 6% each of H5 and U5a, and 3% of X2. Basal H, H3, H4, HV, HV0, U2, U8, V, and W1 have each a frequency of 1% (Figure 10.5). Overall, haplogroup H accounts for 17% in this dataset, of which the majority were found in Scotland (n=12), rather than England (n=1).

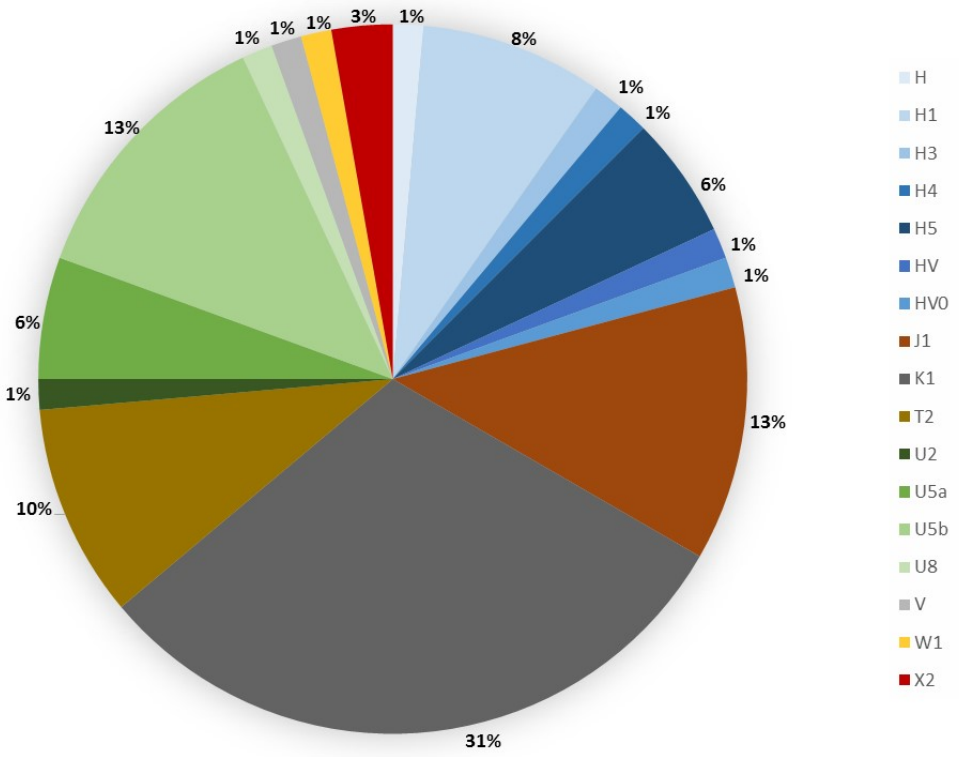

FIGURE 10.5: Mitochondrial DNA haplogroup frequencies of published Neolithic individuals from the British Isles and those reported in this study.

132 individuals date to the Beaker period and Bronze Age, of which 107 are published complete mitogenomes and 25 are from this study. The highest frequency during the Bronze Age in Britain is haplogroup T2 (16%), followed by U5b (13%), K1 (11%), H1 (9%), J1 (8%), U5a (6%), H5 and H39 (each 4%),

H, H2, H3, H6, I2, K2, U4, V10, W1, and X2 (each 2%), and H13, H15, H4, H5'36, H58, HV0, I3, I4, J2, N1, R1, T1, U2, and V (each 1%) (Figure 10.6). The overall frequency of haplogroup H during the Bronze Age is 30%.

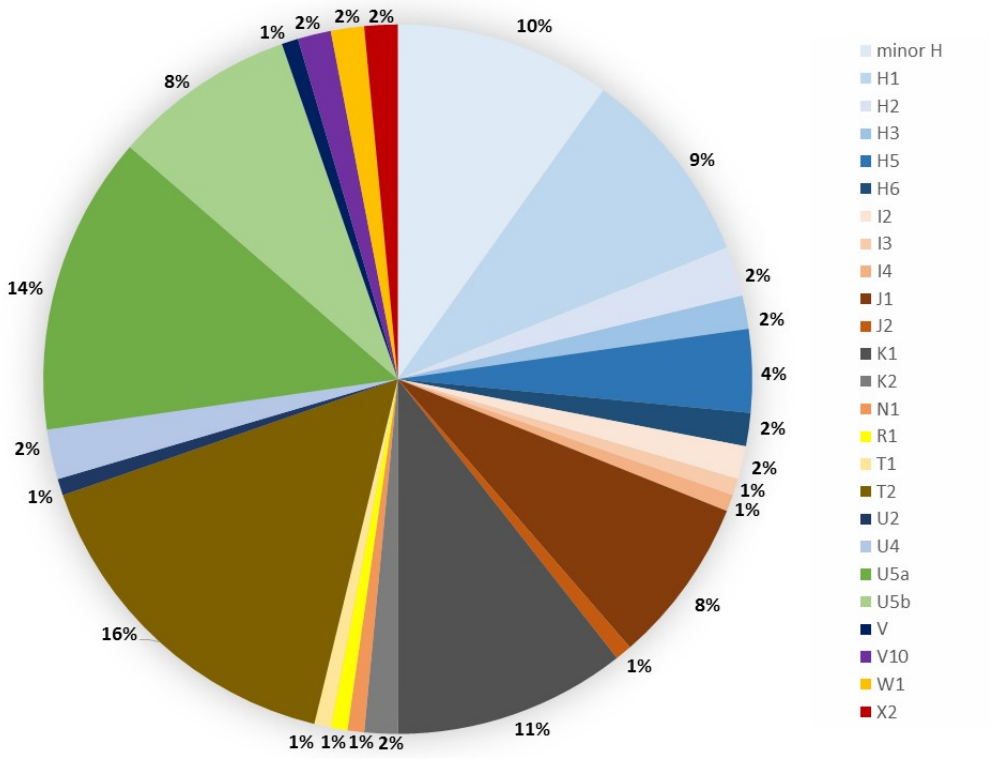

FIGURE 10.6: Mitochondrial DNA haplogroup frequencies of published Bronze Age individuals from the British Isles and those reported in this study.

The published mitochondrial genome dataset consists of four genomes dating to the Iron Age. Including the 12 genomes of this study, the mitochondrial frequencies during the Iron Age in Britain are as follows: H1 38%, J1 19%, H7 13%, and H10, K1, U2, U5 and X2 each 6% (Figure 10.7). The overall frequency of haplogroup H during the Iron Age is 57%, although clearly with a large sampling error.

In the Anglo-Saxon subset, consisting of eight published and 16 newly reported post-Roman and 10 Anglo-Saxon mitogenomes, H1 has a frequency of 26.5%, K1 and T2 of each 15%, HV of 11.8%, H2, U5a and U5b of each 5.9%, and H3, H5, H6, H11 and X2 of each 3% (Figure 10.8). In the Anglo-Saxon data, haplogroup H has an overall frequency of 44%.

Comparing the haplogroup frequencies across the different time periods, the most prominent difference is the increasing frequency of haplogroup H. While K1 was the most frequent haplogroup in the Neolithic, with J1, T2 and U5b also being relatively frequent, mtDNA haplogroups H1 and H5 accounted for only a small fraction of the mitochondrial haplogroups recovered

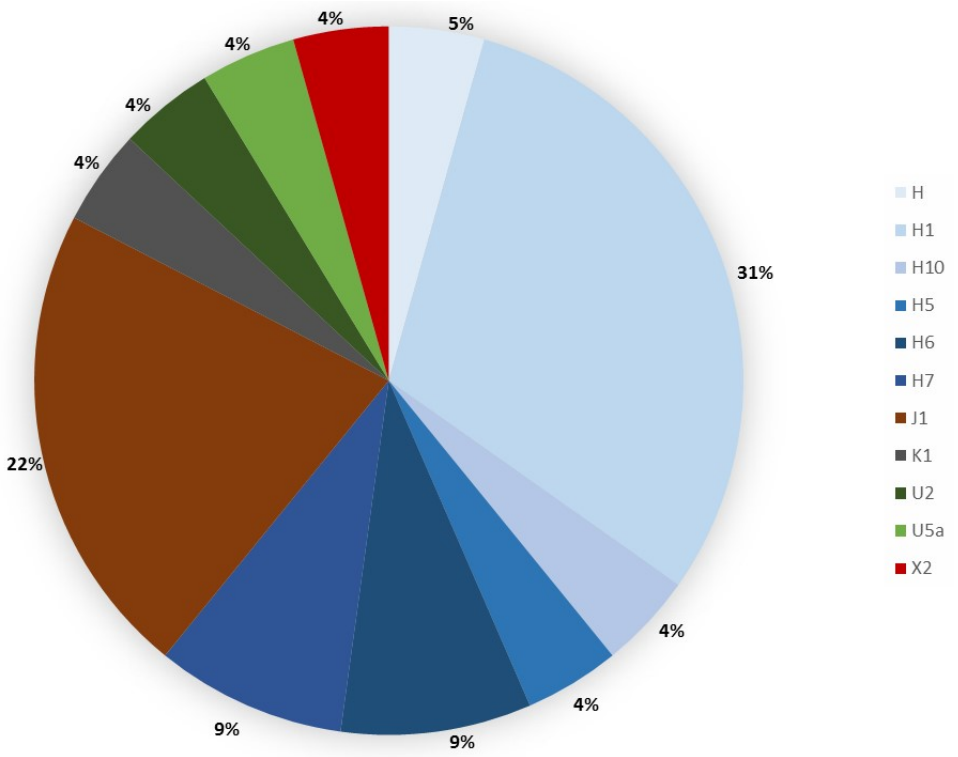

FIGURE 10.7: Mitochondrial DNA haplogroup frequencies of published Iron Age individuals from the British Isles and those reported in this study.

during this time. The arrival of the Bronze Age migrations to the British Isles changed the overall frequency of haplogroup H, with H1 more frequent and the appearance of more subhaplogroups of H. Subhaplogroups of I (I2, I3 and I4), which have not been seen in the Neolithic dataset, were also present and the frequency of K1 drastically decreased from 31% in the British Neolithic to 11% in the British Bronze Age. The frequency of U5b decreased, whilst the presence of U5a increased from 6% to 14%, and the frequency of T2 increased from 10% to 16%.

Haplogroup H1 accounted for almost 40% of the overall haplogroup frequencies in the British Iron Age, although the small sample size should be taken into consideration. The frequency of K1 decreased even more to only 6%, compared to 31% during the Neolithic. Similar to H1, J1 increased in frequency during the Iron Age as well, whereas U5a, which peaked during the Bronze Age, decreased back to its Neolithic frequency.

The mtDNA haplogroup H frequency of the British Anglo-Saxon period was equal to that of the modern British population with 44%. Haplogroup HV* showed an increased frequency of 12%, compared to only 1% during the Neolithic. K1 and T2 together accounted for almost 30%, and U5a and U5b each had a frequency of 6%, which for U5a was similar to the frequency during the

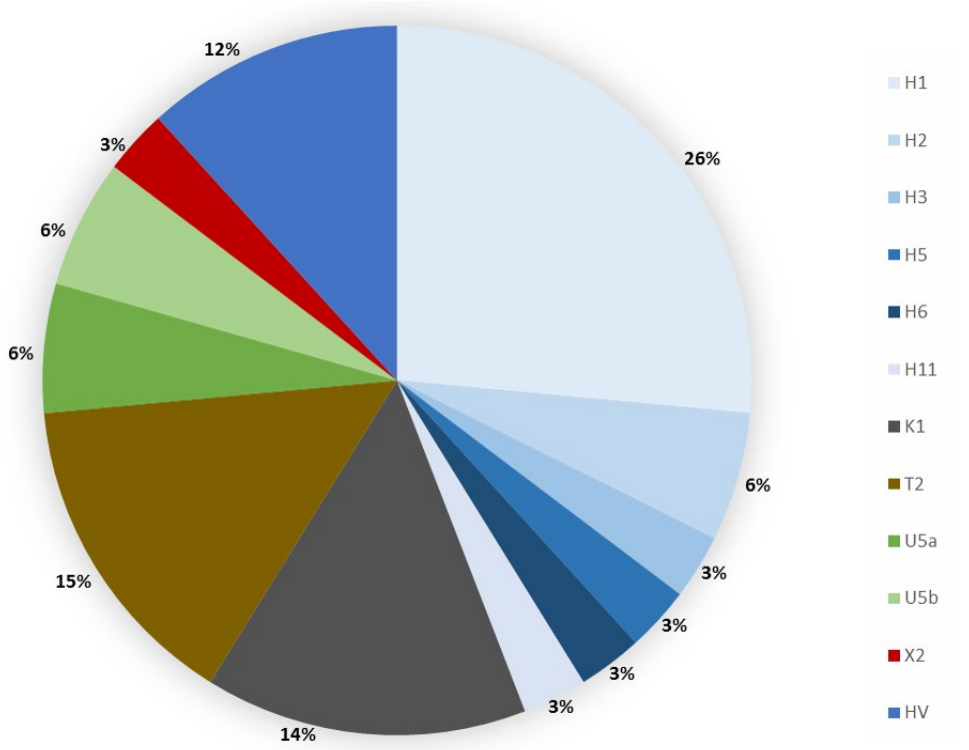

FIGURE 10.8: Mitochondrial DNA haplogroup frequencies of published Anglo-Saxon individuals from the British Isles and those reported in this study, including the post-Roman data.

Iron Age. Due to the rather small Iron Age sample size, it is likely that the presence of U5b in the Iron Age has not yet been detected. Further analysis of Iron Age individuals would be necessary to gain a more complete picture. Generally, U5b is a marker for Late Glacial expansions from the south-west European glacial refuge and very common in the modern Saami population (Silva et al., 2019), and I report here the presence of a subclade of U5b, namely U5b2c, in British Early Neolithic Strathglebe sample.

Mitochondrial haplogroup H, especially H1 and H3 (detailed discussion in Appendix C) from Copper Age Iberia, has been found in high frequency in early Bell Beaker burials in central Europe, alongside Y-chromosome haplogroup R1b. While Y haplogroup R1b spread to Britain during the Bell Beaker and Bronze Age migrations, mtDNA haplogroup H had clearly already arrived with the Neolithic expansions, suggesting a mixed, sex-biased ancestry of female farming lineages from the Mediterranean and male pastoralist lineages of the Steppe (Silva et al., 2019).

10.3 Y-chromosome haplogroup frequencies

As the frequency of mtDNA haplogroup H increased with the Bronze Age migrations in Europe, a similar pattern can be observed for Y-chromosome haplogroup R1b. Overall, the Y-chromosome haplogroup could be assigned to the majority of samples classified genetically as males, with the total number being 33 individuals. All haplogroups presented throughout this study should be understood as predictions only, as all samples had a very low genome coverage and so are likely missing a large number of haplogroup-identifying SNPs (as can be seen in more detail in the individual tables of recovered derived SNPs per male individual in Appendix G). However, the more resolved predicted haplogroups gave a good indication of the overall haplogroup frequencies across the analysed different time periods.

The available Y-chromosome data from the Neolithic consists of 34 published samples (Olalde et al., 2018), and one individual from Strathglebe (KD026) and one from Carsington Pasture Cave (CE001). Over 80% of all individuals belonged to Y-chromosome haplogroup I2 or its subclades (I2a1b accounts for 20%, I2a1b1 for 17.1%, I2a2 for 8.6%, I2a2a for 11.4%, I2a2a1 for 5.7%, I2a2a1a1a for 8.6%, I2a2a1a1a2 for 5.7% and I2a2a1b for 2.9% of the overall Neolithic dataset) (Figure 10.9). 14.3% belonged to unclassified haplogroup I, and one individual was classified as belonging to haplogroup CT (I5374 from Olalde et al. (2018)). However, as the individual predicted to belong to Y-chromosome haplogroup CT had an extremely low coverage, he should be disregarded from conclusions.

The Y-chromosome data from the Bronze Age in Britain contained 59 published (Olalde et al., 2018) and 11 newly reported individuals, nine from the Links of Noltland (KD006, KD045, KD047, KD049, KD057, KD059, KD060, KD061 and KD064) and two from West Heslerton (KD040 and KD041). The vast majority of individuals belonged to subclades of R1b1a1a2 ($\sim$75%). Haplogroups I and I2a1 each have a frequency of $\sim$5.8%, haplogroups I2a2 and R1b1a2a1a each have a frequency of $\sim$4.3%, and haplogroups F, R, and R1 (all from published data) each have a frequency of $\sim$1.4% (Figure 10.10). Again the samples with the predicted haplogroups F, R and R1 had very low Y-chromosome coverage and could thus not be further resolved. These percentages should merely be seen as an indication of presence of a specific haplogroup. The prediction of F is probably incorrect, as haplogroup F for example is basal to both haplogroups I and R.

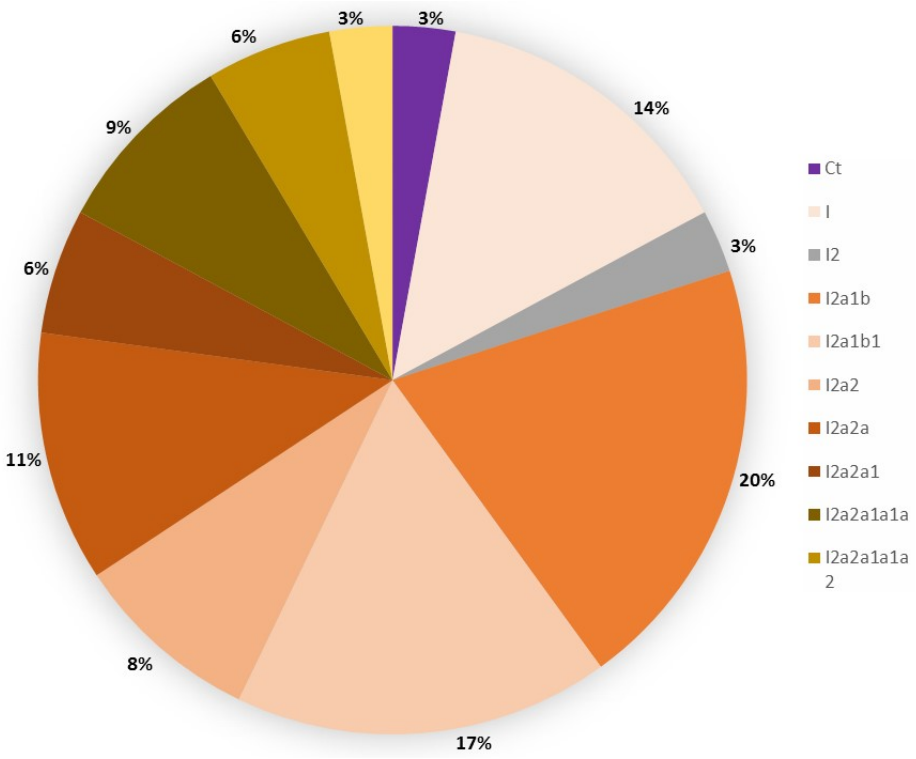

FIGURE 10.9: Y-chromosome haplogroup frequencies of published Neolithic individuals from the British Isles and those reported in this study.

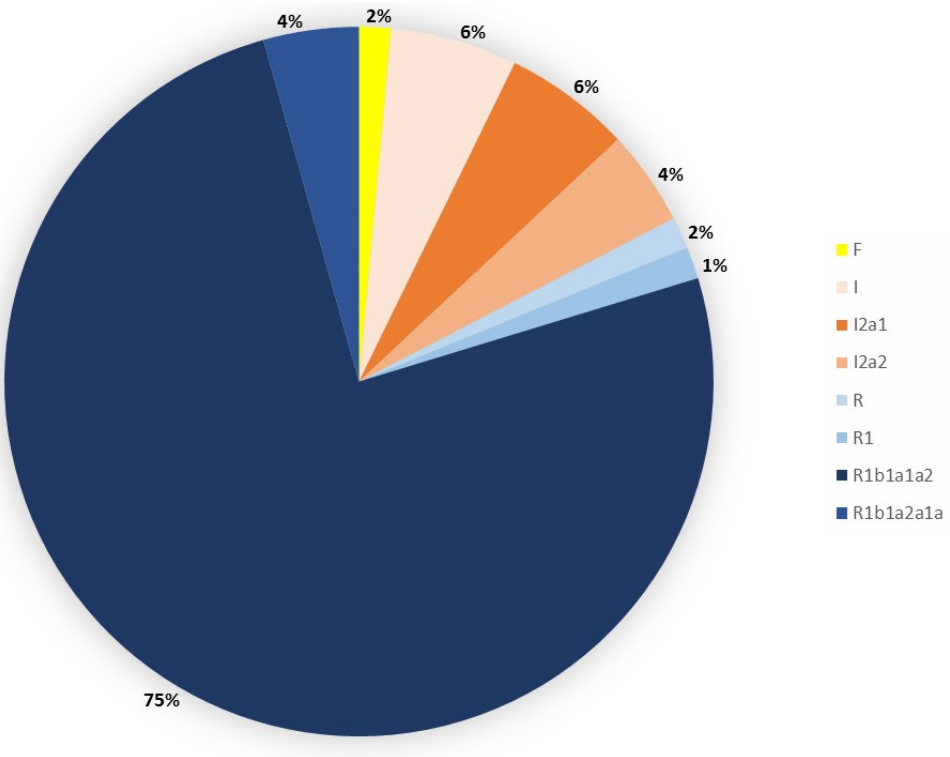

FIGURE 10.10: Y-chromosome haplogroup frequencies of published Bronze Age individuals from the British Isles and those reported in this study.

Overall, only four individuals from the Iron Age in Britain have been identified as male and their Y-chromosome haplogroup has been determined – two published individuals (Schiffels et al., 2016) and three individuals reported in this study, two from the Knowe of Skea (KD004 and KD043), and one from Carsington Pasture Cave (CE004). The two individuals from Hinxton belonged to subclades of R1b1a2a1a (Schiffels et al., 2016), while the two Iron Age individuals from Knowe of Skea and the one from Carsington Pasture Cave all belonged to Y-chromosome haplogroup R1b1a1a2. It is possible that KD004 belonged to a further derived subclade of R1b1a1a2 as one defining SNP each for R1b1a1a2a1a2, R1b1a1a2a1a2c, and R1b1a1a2a1a2c1a2 was also retrieved, and CE004 was typed as R1b1a1a2a1a. In order to not only rely on one single SNP to specify the subhaplogroups, further sequencing is required.

The published Roman and Anglo-Saxon data from Britain only includes two Y-chromosome typed males, who belonged to haplogroups J2 (Martiniano et al., 2016) and I1 (Schiffels et al., 2016), neither of which appears in earlier British samples. Among the 21 Anglo-Saxon and post-Roman males of this study, 18 belonged to subclades of major European Y-chromosome haplogroup R1b1, two belonged to I1a2, and one each belonged to I2a1, I2a2b, and E1b1. The frequencies within R1b1 are shown in Figure 10.11.

The increased frequency of R1b and I2 lineages has been associated with the major expansions of the Late Neolithic and Early Bronze Age across Europe; more specifically, in the case of R1b, to the spread of the Corded Ware culture at the end of the Neolithic, as well as possibly the antecedent Early Bronze Age Yamnaya culture of the Pontic-Caspian Steppe (Silva et al., 2019). Almost all Yamnaya individuals analysed to date fall close to the root of the European R1b1a1a2 cluster, which has been found at high frequencies among the Bronze Age individuals analysed in this study, and also at high frequencies in its subclusters among the analysed post-Roman and Anglo-Saxon samples. This reflects that the impact that the Yamnaya pastoralists and their Corded Ware descendants had on the European gene pool also impacted strongly on the British Isles. Despite being at low frequency in modern-day south-east England, subhaplogroup R1b1a2a1a2c reaches its highest global levels in modern Irish populations ($\sim$80%) and is commonly found in western and northern Britain (Silva et al., 2019).

In the British Neolithic individuals, the observed assimilation of Y-chromosome haplogroup I2 lineages, which are associated with the Mesolithic Atlantic

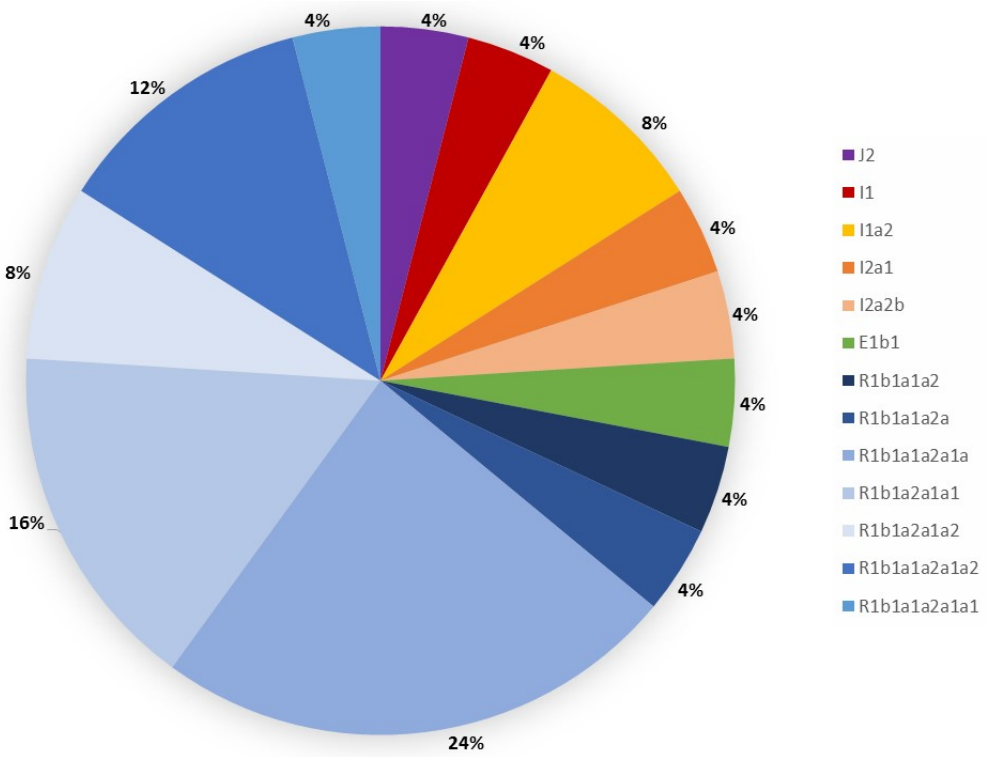

FIGURE 10.11: Y-chromosome haplogroup frequencies of published Anglo-Saxon from the British Isles and (post-)Roman individuals and those reported in this study.

hunter-gatherers, may have taken place (at least in part) during the migration across Europe, rather than once these people arrived into the British Isles (Silva et al., 2019). The levels of hunter-gatherer lineages seen in the Y-chromosome and mitogenomes differ strongly, with 100% assimilation of male lineages but only about a fifth to a quarter of female lineages, suggesting a strong sex bias during the Neolithic migration into north-west Europe (Silva et al., 2019). With the mitochondrial lineages, haplogroup H increased continuously from the Bronze Age until it reached its modern-day frequency during the Anglo-Saxon period. This is very different to the apparent discontinuity seen for Y-chromosome lineages, as R1b almost completely replaced lineages belonging to haplogroup I, and, in the Neolithic specifically, subclusters of I2. Y-chromosome lineages belonging to I1, as well as the R1b-derived cluster R1b1a1a2a1a2, were most likely brought to the British Isles with the Anglo-Saxon migrations, as they have not been seen at earlier dates in the British gene pool (compare Figures 10.9-10.11).

10.4 Sex determination

Osteological and genetic sex determination were not always congruent throughout this study. Where sex had been identified in the ancient human remains, this had either been assumed based on accompanying grave goods (Haak et al., 2008) or osteological analysis. Osteological sex determination is a well established procedure of morphological analysis of ancient human remains, and is widely used in forensic analysis. One of the most reliable skeletal indicators of sex in adults are the pelvic bones (Hrdlička, 1952; Krogman, 1973). Other methods use features of the femur and tibia (Robinson et al., 2011), tooth mineralisation, long bone length, and morphological skull characteristics (Rösing et al., 2007; Chovalopoulou et al., 2013).
Although sex determination of infants and sub-adults has long been considered impossible (Stewart, 1968; Bennett, 1973), refinement of criteria and techniques in the last few decades has supposedly allowed sexing of the infant pelvis using measurements of the ilium (Weaver, 1980). However, a large study comparing several existing methods for sexing, and using new measurements on dimorphic traits of sub-adults, found no effective method to significantly distinguish male and female sub-adult individuals (Satterlee Blake, 2011). Therefore, osteological sex determination, especially of infant and sub-adult remains, still needs further development, and any results have to be considered with caution.
Of the individuals in this study, 36 had an osteological sex identification, 10 of which were uncertain of which two were children. It was not possible to genetically sex one of these individuals but, of the remaining 35, eight were misclassified when compared to genetic sex. Of the 25 individuals with certain morphological sex assignments, 5 were misclassified, which equates to 20%. This leads to the conclusion that, where possible, genetic sex determination should be applied where the aim is to analyse sex biases within sites or even across populations. However, very low coverage genomes might not have enough data to use either karyotyping and/or the Skoglund score (Skoglund et al., 2015) to determine the genetic sex. Based on the dataset analysed in this study, a genome coverage of >0.011x was sufficient to use at least one of the discussed methods for genetic sex determination. In all cases where both the Skoglund script (Skoglund et al., 2013) and the Skoglund score (Skoglund et al., 2015) were applied, the same genetic sex was determined, supporting the robustness of these approaches. When used without

the Skoglund score, the interpretation of karyotype plots of very low coverage samples might be biased towards subjective perception of whether the number of reads of chromosome X are plotting close enough to the regression line to determine an individual as female or not, but it can, in most cases, give an indication where the coverage is too low.

10.5 Conclusions

The use of uniparental markers proved valuable when analysing ancestry of low coverage data. The analysis of the frequencies of maternal and paternal lineages across different time periods from the Neolithic to present-day populations has illuminated the presence of specific haplogroups and refined our understanding of their frequency patterns and their appearance in the British Isles such as, for example, mtDNA lineage H39 in Bronze Age Orkney, mtDNA lineage HV6 in post-Roman Dorset, or Y-chromosome haplogroup outliers such as E1b1.
Genome-wide analysis gave supporting insights into the populations throughout different time periods in the British Isles, as well as an indication of possible North African ancestry in an English Bell Beaker individual (which needs further investigation), as well as African/North African/Near Eastern ancestry in an English post-Roman individual. Kin relationship estimation across cemeteries added an interesting layer to the understanding of social structures during the Bronze Age, Iron Age, and the post-Roman era.
The in-depth analysis of the mitochondrial lineages of an Iron Age woman from the Isle of Skye, and a foetus and a neonate, who had been buried together in what appears to be a sacrificial burial, revealed that intriguingly the female was the mother of only one of the children, which raised questions such as why she was buried with children that were not her own, and why was a possibly pregnant woman apparently killed as a closing ritual to a cave system? In the case of the mutilated young person from Oakridge, genetic data were mainly used to identify the genetic sex, but the fact that the individual turned out to be a young woman opened several questions, such as who was she and why was she punished?
In conclusion, this study has given important insights into the genetic composition of the British Isles, both from ancient and modern data.

Chapter 11

Future work

11.1 Future analysis of modern data

Founder analysis of modern European data is ongoing. This should give an indication of the likely arrival times of mitochondrial lineages into Europe and Britain, which will complement the picture of maternal lineage movements and admixture based on ancient DNA sequencing data. This major founder analysis could also drastically improve our understanding of present day major European mitochondrial haplogroup H, as its dispersal and drastic increase in frequency through time is not yet fully understood.

11.2 Future analysis of ancient data

1. In the recent study investigating a major population replacement in Early Neolithic Britain (Brace et al., 2018), the Outer Hebrides and the Isle of Skye have not been sampled, nor the majority of eastern Scotland. Re-sequencing the samples from Strathglebe, Isle of Skye, to a much higher depth would allow a more in-depth comparison.

2. Further analysis of Bronze Age individuals from Orkney could be of great use to fully understand the extent of population replacement versus continuity on these islands. Although in this study genome-wide analyses support the replacement of populations between the Neolithic and Bronze Age periods, the Y-chromosome and mtDNA lineages also suggest some level of continuity. Future studies would benefit from focusing on islands and remote areas in order to fully understand whether the majority of the Neolithic population of Britain was replaced, or if this was just a phenomenon in distinct geographic areas.

3. The evidence for population continuity from the Bronze Age to Iron Age in Orkney, as shown at the Links of Noltland and the Knowe of

Skea, suggests a very different picture than the postulated major population replacements during the Mesolithic to Neolithic and Neolithic to Bronze Age transitions seen elsewhere in Britain. The generation of more high coverage genomes from across the British Isles during the Bronze to Iron Age transition would allow for a detailed study.

4. The study of genetic continuity versus population admixture during the Iron Age in the British Isles, with a special focus on the Scottish Isles, would certainly be improved by further sequencing to high coverage of the Shetland individual from Milla Skerra.

5. Similarly, more genome-wide data from the Roman, Anglo-Saxon and Viking periods would significantly improve the understanding of population movements and admixture during these times. The study of Worth Matravers presented here already highlights the diversity during the post-Roman period, but, so far, only a few studies have investigated population composition in Roman Britain in any detail. As the Anglo-Saxon cemetery in West Heslerton consists of over 300 burials, a large-scale analysis of this population would be extremely interesting as, not only would it add another layer to our understanding of population genetics during the Anglo-Saxon period, it would also allow for kinship estimation, adding to our understanding of culture and social organisation within burial locations.

6. Generally, the individuals reported here that have high endogenous content should be sequenced to higher coverage, which, as well as allowing a more detailed Y-chromosome analysis of the males, should enable us to investigate selection signals and make more accurate predictions of admixture.

7. The next step in the analysis of ancient populations will be to downscale from analysing entire populations and mass migrations to focus on local groups and investigate their kin relationships. The cemeteries investigated in Bronze Age and Iron Age Orkney, and post-Roman Worth Matravers, as well as the study at Iron Age High Pasture Cave, have already shown that this type of analysis is possible with low coverage data. This type of analysis is adding interesting and important information to our understanding of ancient peoples and their cultures.

References

1000 Genomes Project Consortium (2010). "A map of human genome variation from population-scale sequencing". *Nature* 467 (7319):1061.

Achilli, A. et al. (2004). "The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool". *The American Journal of Human Genetics* 75 (5):910–918.

Adler, C.J. et al. (2011). "Survival and recovery of DNA from ancient teeth and bones". *Journal of Archaeological Science* 38 (5):956–964.

Aldhouse-Green, S. and P. Pettitt (1998). "Paviland Cave: contextualizing the 'Red Lady'". *Antiquity* 72 (278):756–772.

Alexander, D.H., J. Novembre, and K. Lange (2009). "Fast model-based estimation of ancestry in unrelated individuals". *Genome Research*.

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

Álvarez-Iglesias, V. et al. (2009). "New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0". *PloS One* 4 (4):e5112.

Ammerman, A.J. and L.L. Cavalli-Sforza (1984). *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton University Press.

Anderson, S. et al. (1981). "Sequence and organization of the human mitochondrial genome". *Nature* 290 (5806):457.

Anderung, C. et al. (2005). "Prehistoric contacts over the Straits of Gibraltar indicated by genetic analysis of Iberian Bronze Age cattle". *Proceedings of the National Academy of Sciences of the United States of America* 102 (24):8431–8435.

Andrews, R.M. et al. (1999). "Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA". *Nature Genetics* 23 (2):147.

Andrews, S. (2016). "FastQC: a quality control tool for high throughput sequence data. Version 0.11.5". *Babraham Institute, Cambridge, UK http://www. bioinformatics. babraham. ac. uk/projects/fastqc*.

Ansorge, W.J. (2009). "Next-generation DNA sequencing techniques". *New Biotechnology* 25 (4):195–203.

Arbogast, B.S. (2001). "Phylogeography: the history and formation of species". *American Zoologist* 41 (1):134–135.

Armit, I. (in press). "Human remains in Scottish caves: High Pasture Cave in context". In: *High Pasture Cave: Ritual, Memory and Identity in the Iron Age of Skye*. Ed. by S. Birch, G. Cruickshanks, and J. Mckenzie. Oxford: Oxbow Books.

Armit, I. and V. Ginn (2007). "Beyond the grave: human remains from domestic contexts in Iron Age Atlantic Scotland". In: *Proceedings of the Prehistoric Society*. Vol. 73. Cambridge University Press, p. 113–134.

Atkinson, D. (1916). *The Romano-British Site on Lowbury Hill in Berkshire*. University College Reading.

Avise, J.C. et al. (1987). "Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics". *Annual Review of Ecology and Systematics* 18 (1):489–522.

Barker, P. (1981). *Wroxeter Roman City: Excavations 1966-1980*. Department of the Environment.

Barton, N. (1999). "The Lastglacial or Late and Final Upper Palaeolithic colonization of Britain". In: *The Archaeology of Britain: An Introduction from the Upper Palaeolithic to the Industrial Revolution*. Ed. by J. Hunter and I. Ralston. Psychology Press. Chap. 2, p. 13–34.

Barton, R.N.E. et al. (2003). "The Late-glacial reoccupation of the British Isles and the Creswellian". *Journal of Quaternary Science* 18 (7):631–643.

Batini, C. et al. (2017). "Population resequencing of European mitochondrial genomes highlights sex-bias in Bronze Age demographic expansions". *Scientific Reports* 7 (1):12086.

Bayliss, A. and P. Woodman (2009). "A new Bayesian chronology for Mesolithic occupation at Mount Sandel, northern Ireland". In: *Proceedings of the Prehistoric Society*. Vol. 75. Cambridge University Press, p. 101–123.

Behar, D.M. et al. (2008). "Counting the founders: the matrilineal genetic ancestry of the Jewish diaspora". *PloS One* 3 (4):e2062.

Behar, D.M. et al. (2012a). "A "Copernican" reassessment of the human mitochondrial DNA tree from its root". *The American Journal of Human Genetics* 90 (4):675–684.

Behar, D.M. et al. (2012b). "The Basque paradigm: genetic evidence of a maternal continuity in the Franco-Cantabrian region since pre-Neolithic times". *The American Journal of Human Genetics* 90 (3):486–493.

Bennett, K.A. (1973). "On the estimation of some demographic characteristics on a prehistoric population from the American Southwest". *American Journal of Physical Anthropology* 39 (2):223–231.

Bennike, P. (2003). "Ancient trepanations and differential diagnosis: a re-evaluation of skeletal remains from Denmark". *Trepanation: History, Discovery, Theory. Swets & Zeitlinger, Lisse*:95–116.

Bertolin, C. et al. (2011). "Analysis of complete mitochondrial genomes of patients with schizophrenia and bipolar disorder". *Journal of Human Genetics* 56 (12):869–872.

Birch, S. and M. Wildgoose (2013). *Uamh an Ard Achadh (High Pasture Cave). Strath, Isle of Skye 2010/11 (NGR NG 5943 1971). The Preliminary Assessment and Analysis of Late Prehistoric Cultural Deposits from a Limestone Cave and Associated Surface Features. Data Structure Report – HPC007*. National Museums Scotland, University of Durham: West Coast Archaeological Services.

Birch, S. and M. Peteranna (2017). *Archaeological Excavation in Learnie 2B Near Rosemarkie, Ross-shire. Data Structure Report*. Rosemarkie Caves Project.

Blockley, S.P.E. et al. (2006). "The chronology of abrupt climate change and Late Upper Palaeolithic human adaptation in Europe". *Journal of Quaternary Science* 21 (5):575–584.

Bodner, M. et al. (2015). "Helena, the hidden beauty: resolving the most common West Eurasian mtDNA control region haplotype by massively parallel sequencing an Italian population sample". *Forensic Science International: Genetics* 15:21–26.

Bonsall, C. (1984). "Low Hauxley, Northumberland". *Proceedings of the Prehistoric Society* 50:398.

Botigué, L.R. et al. (2013). "Gene flow from North Africa contributes to differential human genetic diversity in southern Europe". *Proceedings of the National Academy of Sciences of the United States of America* 110 (29):11791–11796.

Bownes, J. (2018). "Reassessing the Scottish Mesolithic-Neolithic Transition: Questions of Diet and Chronology". PhD thesis. Glasgow University.

Brace, S. et al. (2018). "Population replacement in Early Neolithic Britain". *bioRxiv*:267443.

Bradley, R. (2007). *The Prehistory of Britain and Ireland*. Cambridge University Press.

Bramanti, B. et al. (2009). "Genetic discontinuity between local hunter-gatherers and central Europe's first farmers". *Science* 326 (5949):137–140.

Brandt, G. et al. (2013). "Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity". *Science* 342 (6155):257–261.

Brandt, G. et al. (2015). "Human paleogenetics of Europe – The known knowns and the known unknowns". *Journal of Human Evolution* 79:73–92.

Breeze, D.J. and B. Dobson (1976). "A view of Roman Scotland in 1975". *Glasgow Archaeological Journal* 4 (4):124–143.

Briggs, A.W. et al. (2007). "Patterns of damage in genomic DNA sequences from a Neandertal". *Proceedings of the National Academy of Sciences of the United States of America* 104 (37):14616–14621.

Brncic, D. (1956). "Chromosomal polymorphism in natural populations of *Drosophila pavani*". *Chromosoma* 8 (1):699–708.

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

Brotherton, P. et al. (2013). "Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans". *Nature Communications* 4:1764.

Broushaki, F. et al. (2016). "Early Neolithic genomes from the eastern Fertile Crescent". *Science*:aaf7943.

Brown, T. (1997). "Clearances and clearings: deforestation in Mesolithic/ Neolithic Britain". *Oxford Journal of Archaeology* 16 (2):133–146.

Byrne, R.P. et al. (2018). "Insular celtic population structure and genomic footprints of migration". *PLoS Genetics* 14 (1):e1007152.

Campos, P.F. et al. (2012). "DNA in ancient bone – Where is it located and how should we extract it?" *Annals of Anatomy–Anatomischer Anzeiger* 194 (1):7–16.

Capelli, C. et al. (2003). "A Y chromosome census of the British Isles". *Current Biology* 13 (11):979–984.

Cardoso, S. et al. (2013). "The expanded mtDNA phylogeny of the Franco-Cantabrian region upholds the pre-Neolithic genetic substrate of Basques". *PLoS One* 8 (7):e67835.

Cassidy, L.M. et al. (2016). "Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome". *Proceedings of the National Academy of Sciences of the United States of America* 113 (2):368–373.

Cavalli-Sforza, L.L. and A.W.F. Edwards (1967). "Phylogenetic analysis: models and estimation procedures". *Evolution* 21 (3):550–570.

Champion, T. (1999). "The Later Bronze Age". In: *The Archaeology of Britain: An Introduction from the Upper Palaeolithic to the Industrial Revolution*. Ed. by J. Hunter and I. Ralston. Psychology Press. Chap. 6, p. 95–112.

Charlesworth, B. and D. Charlesworth (2017). "Population genetics from 1966 to 2016". *Heredity* 118 (1):2–9.

Chenery, C. et al. (2010). "Strontium and stable isotope evidence for diet and mobility in Roman Gloucester, UK". *Journal of Archaeological Science* 37 (1):150–163.

Chovalopoulou, M.-E., E.D. Valakos, and S.K. Manolis (2013). "Sex determination by three-dimensional geometric morphometrics of the palate and cranial base". *Anthropologischer Anzeiger* 70 (4):407–425.

Clark, C.D. et al. (2012). "Pattern and timing of retreat of the last British-Irish ice sheet". *Quaternary Science Reviews* 44:112–146.

Coble, M.D. et al. (2004). "Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians". *International Journal of Legal Medicine* 118 (3):137–146.

Collins, M.J. et al. (2002). "The survival of organic matter in bone: a review". *Archaeometry* 44 (3):383–394.

Cook, J. and R. Jacobi (1994). "A Reindeer antler or 'Lyngby' axe from Northamptonshire and its context in the British Late Glacial". In: *Proceedings of the Prehistoric Society*. Vol. 60. Cambridge University Press, p. 75–84.

Coolen, M.J.L. et al. (2006). "Ancient DNA derived from alkenone-biosynthesizing haptophytes and other algae in Holocene sediments from the Black Sea". *Paleoceanography and Paleoclimatology* 21 (1).

Costa, M.D. et al. (2009). "Data from complete mtDNA sequencing of Tunisian centenarians: testing haplogroup association and the "golden mean" to long-evity". *Mechanisms of Ageing and Development* 130 (4):222–226.

Cowie, T. (1993). "A survey of the Neolithic pottery of eastern and central Scotland". In: *Proceedings of the Society of Antiquaries of Scotland*. Vol. 123. National Museum of Antiquities of Scotland, p. 13–41.

Crawford, M.H. and G. Gmelch (1974). "Human biology of the Irish tinkers: demography, ethnohistory, and genetics". *Social Biology* 21 (4):321–331.

Cui, Y. et al. (2015). "Identification of kinship and occupant status in Mongolian noble burials of the Yuan Dynasty through a multidisciplinary approach". *Philosophical Transactions of the Royal Society B: Biological Sciences* 370 (1660):20130378.

Cunliffe, B. (2013). *Britain Begins*. Oxford University Press.

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

Dark, P. (2000). "Revised 'absolute'dating of the early Mesolithic site of Star Carr, North Yorkshire, in the light of changes in the early Holocene tree-ring chronology". *Antiquity* 74 (284):304–307.

Darling, W.G., A.H. Bath, and J.C. Talbot (2003). "The O and H stable isotope composition of freshwaters in the British Isles. 2, surface waters and groundwater". *Hydrology and Earth System Sciences* 7:183–195.

Davidovic, S. et al. (2015). "Mitochondrial DNA perspective of Serbian genetic diversity". *American Journal of Physical Anthropology* 156 (3):449–465.

Davison, K. et al. (2006). "The role of waterways in the spread of the Neolithic". *Journal of Archaeological Science* 33 (5):641–652.

de Araujo, L.F. et al. (2015). "Mitochondrial genome instability in colorectal adenoma and adenocarcinoma". *Tumor Biology* 36 (11):8869–8879.

de Barros Damgaard, P. et al. (2018). "The first horse herders and the impact of early Bronze Age steppe expansions into Asia". *Science*:eaar7711.

Deguilloux, M.-F. et al. (2012). "European neolithization and ancient DNA: an assessment". *Evolutionary Anthropology: Issues, News, and Reviews* 21 (1):24–37.

Deguilloux, M.F. et al. (2014). "Ancient DNA and kinship analysis of human remains deposited in Merovingian necropolis sarcophagi (Jau Dignac et Loirac, France, 7th–8th century AD)". *Journal of Archaeological Science* 41:399–405.

Dent, J.S. (1983). "The impact of Roman rule on native society in the territory of the Parisi". *Britannia* 14:35–44.

Derenko, M. et al. (2013). "Complete mitochondrial DNA diversity in Iranians". *PloS One* 8 (11):e80673.

Derenko, M. et al. (2014). "Western Eurasian ancestry in modern Siberians based on mitogenomic data". *BMC Evolutionary Biology* 14 (1):217.

Di Bernardo, G. et al. (2009). "Ancient DNA and family relationships in a Pompeian house". *Annals of Human Genetics* 73 (4):429–437.

Dinnis, R. (2012). "The timing of Aurignacian occupation of the British Peninsula". *Quartär* 59:67–83.

Doherty, C. (1980). "Exchange and trade in Early Medieval Ireland". *The Journal of the Royal Society of Antiquaries of Ireland* 110:67–89.

Dowd, M. and R.F. Carden (2016). "First evidence of a Late Upper Palaeolithic human presence in Ireland". *Quaternary Science Reviews* 139:158–163.

Drews, C. (2005). *Refuse or Ritual: The Mammal Bones from High Pasture Cave, Skye*.

Duggan, A.T. et al. (2013). "Investigating the prehistory of Tungusic peoples of Siberia and the Amur-Ussuri region with complete mtDNA genome sequences and Y-chromosomal markers". *PloS One* 8 (12):e83570.

Dulias, K. (2019). "Archaeogenetics and Palaeogenetics of the British Isles". In: University of Huddersfield.

Edwards, C.J. (2012). "Ancient DNA and analyses of mammalian remains from Ireland". *The Irish Naturalists' Journal*:55–63.

Edwards, R. and A. Brooks (2008). "The island of Ireland: drowning the myth of an Irish land-bridge?" *The Irish Naturalists' Journal*:19–34.

Eltsov, N. and N. Volodko (2011). *MtPhyl: software tool for human mtDNA analysis and phylogeny reconstruction*.

Epp, L.S. et al. (2015). "Lake sediment multi-taxon DNA from North Greenland records early post-glacial appearance of vascular plants and accurately tracks environmental changes". *Quaternary Science Reviews* 117:152–163.

Evans, J.A., C.A. Chenery, and A.P. Fitzpatrick (2006). "Bronze Age childhood migration of individuals near Stonehenge, revealed by strontium and oxygen isotope tooth enamel analysis". *Archaeometry* 48 (2):309–321.

Faerman, M. et al. (1998). "Determining the sex of infanticide victims from the Late Roman era through ancient DNA analysis". *Journal of Archaeological Science* 25 (9):861–865.

Fendt, L. et al. (2011). "Accumulation of mutations over the entire mitochondrial genome of breast cancer cells obtained by tissue microdissection". *Breast Cancer Research and Treatment* 128 (2):327–336.

Fernandes, D.M. et al. (2018). "A genomic Neolithic time transect of hunter-farmer admixture in central Poland". *Scientific Reports* 8 (1):14879.

Finlay, N. (2000). "Outside of life: traditions of infant burial in Ireland from cillin to cist". *World Archaeology* 31 (3):407–422.

Finnilä, S., M.S. Lehtonen, and K. Majamaa (2001). "Phylogenetic network for European mtDNA". *The American Journal of Human Genetics* 68 (6):1475–1484.

Forster, P. et al. (1996). "Origin and evolution of Native American mtDNA variation: a reappraisal". *American Journal of Human Genetics* 59 (4):935.

Fraumene, C. et al. (2006). "High resolution analysis and phylogenetic network construction using complete mtDNA sequences in Sardinian genetic isolates". *Molecular Biology and Evolution* 23 (11):2101–2111.

Fregel, R. et al. (2009). "The maternal aborigine colonization of La Palma (Canary Islands)". *European Journal of Human Genetics* 17 (10):1314.

Fu, Q. et al. (2012). "Complete mitochondrial genomes reveal Neolithic expansion into Europe". *PLoS One* 7 (3):e32473.

Fu, Q. et al. (2016). "The genetic history of Ice Age Europe". *Nature* 534 (7606):200–205.

Gamba, C. et al. (2014). "Genome flux and stasis in a five millennium transect of European prehistory". *Nature Communications* 5:5257.

Gandini, F. et al. (2016). "Mapping human dispersals into the Horn of Africa from Arabian Ice Age refugia using mitogenomes". *Scientific Reports* 6:25472.

Gasparre, G. et al. (2007). "Disruptive mitochondrial DNA mutations in complex I subunits are markers of oncocytic phenotype in thyroid tumors". *Proceedings of the National Academy of Sciences of the United States of America* 104 (21):9001–9006.

Gianoarli, L. et al. (2014). "Mitogenomes of polar bodies and corresponding oocytes". *PloS One* 9 (7):e102182.

Gilbert, E. et al. (2017a). "Genomic insights into the population structure and history of the Irish Travellers". *Scientific Reports* 7:42187.

Gilbert, E. et al. (2017b). "The Irish DNA atlas: revealing fine-scale population structure and history within Ireland". *Scientific Reports* 7 (1):17199.

Gkiasta, M. et al. (2003). "Neolithic transition in Europe: the radiocarbon record revisited". *Antiquity* 77 (295):45–62.

Gómez-Carballa, A. et al. (2011). "Evolutionary analyses of entire genomes do not support the association of mtDNA mutations with Ras/MAPK pathway syndromes". *PLoS One* 6 (4):e18348.

Goodacre, S. et al. (2005). "Genetic evidence for a family-based Scandinavian settlement of Shetland and Orkney during the Viking periods". *Heredity* 95 (2):129–135.

Greenspan, B. *Direct Submission - Family Tree DNA - Genealogy by Genetics*.

Günther, T. et al. (2015). "Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques". *Proceedings of the National Academy of Sciences of the United States of America* 112 (38):11917–11922.

Haak, W. et al. (2005). "Ancient DNA from the first European farmers in 7500-Year-Old Neolithic sites". *Science* 310 (5750):1016–1018.

Haak, W. et al. (2008). "Ancient DNA, Strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age". *Proceedings of the National Academy of Sciences of the United States of America*:pnas–0807592105.

Haak, W. et al. (2010). "Ancient DNA from European Early Neolithic farmers reveals their Near Eastern affinities". *PLoS Biology* 8 (11):e1000536.

Haak, W. et al. (2015). "Massive migration from the steppe was a source for Indo-European languages in Europe". *Nature* 522 (7555):207–211.

Hansen, H.B. et al. (2017). "Comparing ancient DNA preservation in petrous bone and tooth cementum". *PloS One* 12 (1):e0170940.

Harding, J. (1997). "Interpreting the Neolithic: the monuments of North Yorkshire". *Oxford Journal of Archaeology* 16 (3):279–295.

Haselgrove, C. (1999). "The Iron Age". In: *The Archaeology of Britain: An Introduction from the Upper Palaeolithic to the Industrial Revolution*. Ed. by J. Hunter and I. Ralston. Psychology Press. Chap. 7, p. 113–134.

Haughton, C. and D. Powlesland (1999a). "West Heslerton The Anglian Cemetery. Catalogue of the Anglian Graves and Associated Assemblages". In: vol. 2. English Heritage.

— (1999b). "West Heslerton The Anglian Cemetery. The Excavation and Discussion of the Evidence". In: vol. 1. English Heritage.

Haynes, S. et al. (2002). "Bone preservation and ancient DNA: the application of screening methods for predicting DNA survival". *Journal of Archaeological Science* 29 (6):585–592.

Helgason, A. et al. (2001). "mtDNA and the islands of the north Atlantic: estimating the proportions of Norse and Gaelic ancestry". *The American Journal of Human Genetics* 68 (3):723–737.

Herrnstadt, C. et al. (2002). "Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups". *The American Journal of Human Genetics* 70 (5):1152–1171.

Hervella, M. et al. (2012). "Ancient DNA from hunter-gatherer and farmer groups from northern Spain supports a random dispersion model for the Neolithic expansion into Europe". *PloS One* 7 (4):e34417.

Hewitt, G.M. (2001). "Speciation, hybrid zones and phylogeography – or seeing genes in space and time". *Molecular Ecology* 10 (3):537–549.

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

Hill, J.D. (1995). "The pre-Roman Iron Age in Britain and Ireland (ca. 800 B.C. to A.D. 100): an overview". *Journal of World Prehistory* 9 (1):47–98.

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

Hofreiter, M. et al. (2001). "Ancient DNA". *Nature Reviews Genetics* 2 (5):353–359.

Housley, R.A. et al. (1997). "Radiocarbon evidence for the Lateglacial human recolonisation of northern Europe". In: *Proceedings of the Prehistoric Society*. Vol. 63. Cambridge University Press, p. 25–54.

Hrdlička, A. (1952). *Practical anthropometry*. Wistar Institute of Anatomy and Biology.

Hull, B.D. (2008). "Social differentiation and diet in Early Anglo-Saxon England: stable isotope analysis of archaeological human and animal remains". PhD thesis. University of Oxford, p. 364.

Imanishi, H. et al. (2011). "Mitochondrial DNA mutations regulate metastasis of human breast cancer cells". *PloS One* 6 (8):e23401.

Ingman, M. and U. Gyllensten (2007). "Rate variation between mitochondrial domains and adaptive evolution in humans". *Human Molecular Genetics* 16 (19):2281–2287.

International Human Genome Sequencing Consortium (2001). "Initial sequencing and analysis of the human genome". *Nature* 409 (6822):860–921.

Irish, J.D. et al. (2008). "A late Magdalenian perinatal human skeleton from Wilczyce, Poland". *Journal of Human Evolution* 55 (4):736.

Jacobi, R.M. and T.F.G. Higham (2009). "The early Lateglacial re-colonization of Britain: new radiocarbon evidence from Gough's Cave, southwest England". *Quaternary Science Reviews* 28 (19):1895–1913.

Jay, M. et al. (2013). "British Iron Age chariot burials of the Arras culture: a multi-isotope approach to investigating mobility levels and subsistence practices". *World Archaeology* 45 (3):473–491.

Jobling, M.A. and C. Tyler-Smith (2003). "The human Y chromosome: an evolutionary marker comes of age". *Nature Reviews Genetics* 4 (8):598–612.

Jones, E.R. et al. (2015). "Upper Palaeolithic genomes reveal deep roots of modern Eurasians". *Nature Communications* 6:8912.

Jones, E.R. et al. (2017). "The Neolithic transition in the Baltic was not driven by admixture with Early European Farmers". *Current Biology* 27 (4):576–582.

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

Juras, A. et al. (2017). "Diverse origin of mitochondrial lineages in Iron Age Black Sea Scythians". *Scientific Reports* 7:43950.

Juras, A. et al. (2018). "Mitochondrial genomes reveal an east to west cline of steppe ancestry in Corded Ware populations". *Scientific Reports* 8 (1):11603.

Just, R.S. et al. (2014). "Full mtGenome reference data: development and characterization of 588 forensic-quality haplotypes representing three U.S. populations". *Forensic Science International: Genetics* 14:141–155.

Keller, A. et al. (2012). "New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing". *Nature Communications* 3:698.

King, J. and G. Cole (2016). "The radiocarbon dating of human remains from seven sites in north Hampshire, with special reference to the results from Old Basing". *Hampshire Studies* 71 (1):76–88.

King, T.E. et al. (2014). "Identification of the remains of King Richard III". *Nature Communications* 5:5631.

Kircher, M., S. Sawyer, and M. Meyer (2012). "Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform". *Nucleic Acids Research* 40 (1):e3–e3.

Kloss-Brandstätter, A. et al. (2010). "Somatic mutations throughout the entire mitochondrial genome are associated with elevated PSA levels in prostate cancer patients". *The American Journal of Human Genetics* 87 (6):802–812.

Kloss-Brandstätter, A. et al. (2015). "Validation of next-generation sequencing of entire mitochondrial genomes and the diversity of mitochondrial DNA mutations in oral squamous cell carcinoma". *PLoS One* 10 (8):e0135643.

Knipper, C. et al. (2017). "Female exogamy and gene pool diversification at the transition from the Final Neolithic to the Early Bronze Age in central Europe". *Proceedings of the National Academy of Sciences of the United States of America*:201706355.

Krings, M. et al. (1997). "Neandertal DNA sequences and the origin of modern humans". *Cell* 90 (1):19–30.

Kristiansen, K. (2014). *The decline of the Neolithic and the rise of Bronze Age society*. Ed. by C. Fowler, J. Harding, and D. Hofmann. Oxford University Press.

Krogman, W.M. (1973). "The human skeleton in forensic medicine". *C.C. Thomas, Springfield*.

Kruz, A. (2018). "Radiocarbon dating and bayesian modelling of the post-Roman cemetery". In: *Multi-Period Occupation at Football Field, Worth Matravers, Dorset 2006-2011, Bar Monograph (British Series)*. Ed. by L. Ladle. Chap. 12, p. 216–220.

Kuhn, J.M.M., M. Jakobsson, and T. Günther (2018). "Estimating genetic kin relationships in prehistoric populations". *PloS One* 13 (4):e0195491.

Lacan, M. et al. (2011a). "Ancient DNA reveals male diffusion through the Neolithic Mediterranean route". *Proceedings of the National Academy of Sciences of the United States of America* 108 (24):9788–9791.

Lacan, M. et al. (2011b). "Ancient DNA suggests the leading role played by men in the Neolithic dissemination". *Proceedings of the National Academy of Sciences of the United States of America*:201113061.

Lacan, M. et al. (2013). "Ancestry of modern Europeans: contributions of ancient DNA". *Cellular and Molecular Life Sciences* 70 (14):2473–2487.

Ladle, L. (2018a). "Interpretation and discussion". In: *Multi-Period Occupation at Football Field, Worth Matravers, Dorset 2006-2011, Bar Monograph (British Series)*. Ed. by L. Ladle. Chap. 16.

— (2018b). *Multi-Period Occupation at Football Field, Worth Matravers, Dorset 2006-2011, Bar Monograph (British Series)*.

Lawson, D.J., L. Van Dorp, and D. Falush (2018). "A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots". *Nature Communications* 9 (1):3258.

Lazaridis, I. et al. (2014). "Ancient human genomes suggest three ancestral populations for present-day Europeans". *Nature* 513 (7518):409.

Lazaridis, I. et al. (2016). "Genomic insights into the origin of farming in the ancient Near East". *Nature* 536 (7617):419–424.

Lazaridis, I. et al. (2017). "Genetic origins of the Minoans and Mycenaeans". *Nature* 548 (7666):214.

Leach, S. et al. (2010). "A Lady of York: migration, ethnicity and identity in Roman Britain". *Antiquity* 84 (323):131–145.

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

Li, H. (2011). "A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data". *Bioinformatics* 27 (21):2987–2993.

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

Li, H. et al. (2009b). "The sequence alignment/map format and SAMtools". *Bioinformatics* 25 (16):2078–2079.

Li, S. et al. (2014). "Variation and association to diabetes in 2000 full mtDNA sequences mined from an exome study in a Danish population". *European Journal of Human Genetics* 22 (8):1040–1045.

Lippold, S. et al. (2014). "Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences". *Investigative Genetics* 5 (1):13.

Lipson, M. et al. (2017). "Parallel palaeogenomic transects reveal complex genetic history of early European farmers". *Nature* 551 (7680):368–372.

Livarda, A. and R. Madgwick (2017). "Ritual and religion: bioarchaeological perspectives". In: *The Bioarchaeology of Ritual and Religion*. Ed. by A. Livarda, R. Madgwick, and S. Riera-Mora. Oxford: Oxbow Books, p. 1–13.

Lloyd, R.E. et al. (2015). "Identification and functional prediction of mitochondrial complex III and IV mutations associated with glioblastoma". *Neuro-Oncology* 17 (7):942–952.

Loogväli, E.-L. et al. (2004). "Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia". *Molecular Biology and Evolution* 21 (11):2012–2021.

Macaulay, V. and M.B. Richards (2013). "Mitochondrial genome sequences and their phylogeographic interpretation". *Encyclopedia of Life Sciences*.

MacHugh, D.E. et al. (2000). "The extraction and analysis of ancient DNA from bone and teeth: a survey of current methodologies". *Ancient Biomolecules* 3:81–102.

Malaspinas, A.-S. et al. (2014). "*bammds*: a tool for assessing the ancestry of low-depth whole-genome data using multidimensional scaling (MDS)". *Bioinformatics* 30 (20):2962–2964.

Mallory, J.P. (2013). *The Origins of the Irish*. Thames & Hudson.

Malmström, H. et al. (2009). "Ancient DNA reveals lack of continuity between Neolithic hunter-gatherers and contemporary Scandinavians". *Current Biology* 19 (20):1758–1762.

Malyarchuk, B. et al. (2010a). "Mitogenomic diversity in Tatars from the Volga-Ural region of Russia". *Molecular Biology and Evolution* 27 (10):2220–2226.

Malyarchuk, B. et al. (2010b). "The peopling of Europe from the mitochondrial haplogroup U5 perspective". *PLoS One* 5 (4):e10285.

Margaryan, A. et al. (2017). "Eight millennia of matrilineal genetic continuity in the south Caucasus". *Current Biology* 27 (13):2023–2028.

Martin, M. (2011). "Cutadapt removes adapter sequences from high-throughput sequencing reads". *EMBnet. Journal Bioinformatics in Action* 17 (1):pp–10.

Martiniano, R. et al. (2014). "Genetic evidence of African slavery at the beginning of the trans-Atlantic slave trade". *Scientific Reports* 4:5994.

Martiniano, R. et al. (2016). "Genomic signals of migration and continuity in Britain before the Anglo-Saxons". *Nature Communications* 7:10326.

Mathieson, I. et al. (2015). "Genome-wide patterns of selection in 230 ancient Eurasians". *Nature* 528 (7583):499.

Mathieson, I. et al. (2018). "The genomic history of southeastern Europe". *Nature* 555 (7695):197–203.

Mays, S. and J. Steele (1996). "A mutilated human skull from Roman St Albans, Hertfordshire, England". *Antiquity* 70 (267):155–161.

McEvoy, B., K. Simms, and D.G. Bradley (2008). "Genetic investigation of the patrilineal kinship structure of early medieval Ireland". *American Journal of Physical Anthropology* 136 (4):415–422.

Meyer, M. and M. Kircher (2010). "Illumina sequencing library preparation for highly multiplexed target capture and sequencing". *Cold Spring Harbor Protocols* 2010 (6):pdb–prot5448.

Mielnik-Sikorska, M. et al. (2013). "The history of slavs inferred from complete mitochondrial genome sequences". *PloS One* 8 (1):e54360.

Millett, M. and R. Gowland (2015). "Infant and child burial rites in Roman Britain: a study from East Yorkshire". *Britannia* 46:171–189.

Mishmar, D. et al. (2003). "Natural selection shaped regional mtDNA variation in humans". *Proceedings of the National Academy of Sciences of the United States of America* 100 (1):171–176.

Mithen, S. (1999). "Hunter-gatherers of the Mesolithic". In: *The Archaeology of Britain: An Introduction from the Upper Palaeolithic to the Industrial Revolution*. Ed. by J. Hunter and I. Ralston. Psychology Press. Chap. 3, p. 35–57.

Mittnik, A. et al. (2018). "The genetic prehistory of the Baltic Sea region". *Nature communications* 9 (1):442.

Moore, L.T. et al. (2006). "A Y-chromosome signature of hegemony in Gaelic Ireland". *The American Journal of Human Genetics* 78 (2):334–338.

Moorjani, P. et al. (2011). "The history of African gene flow into southern Europeans, Levantines, and Jews". *PLoS Genetics* 7 (4):e1001373.

Mrevlishvili, G.M. and D.V. Svintradze (2005). "DNA as a matrix of collagen fibrils". *International Journal of Biological Macromolecules* 36 (5):324–326.

Müldner, G. and M.P. Richards (2007). "Stable isotope evidence for 1500 years of human diet at the city of York, UK". *American Journal of Physical Anthropology* 133 (1):682–697.

Mullis, K. et al. (1986). "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction". In: *Cold Spring Harbor Symposia on Quantitative Biology*. Vol. 51. Cold Spring Harbor Laboratory Press, p. 263–273.

Myres, N.M. et al. (2011). "A major Y-chromosome haplogroup R1b Holocene era founder effect in central and western Europe". *European Journal of Human Genetics* 19 (1):95.

Neparáczki, E. et al. (2017). "Revising mtDNA haplotypes of the ancient Hungarian conquerors with next generation sequencing". *PloS One* 12 (4):e0174886.

Neparaczki, E. et al. (2018). "Mitogenomic data indicate admixture components of Asian Hun and Srubnaya origin in the Hungarian Conquerors". *bioRxiv*:250688.

Newman, C. (1998). "Reflections on the making of a 'royal site' in early Ireland". *World Archaeology* 30 (1):127–141.

Nikitin, A.G. et al. (2017). "Subdivisions of haplogroups U and C encompass mitochondrial DNA lineages of Eneolithic – Early Bronze Age Kurgan populations of western north Pontic steppe". *Journal of Human Genetics* 62 (6):605.

Noonan, J.P. et al. (2006). "Sequencing and analysis of Neanderthal genomic DNA". *Science* 314 (5802):(17), 1113–1118.

North, K.E., L.J. Martin, and M.H. Crawford (2000). "The origins of the Irish travellers and the genetic structure of Ireland". *Annals of Human Biology* 27 (5):453–465.

Ó Corráin, D. (1998). "The Vikings in Scotland and Ireland in the ninth Century". *Peritia* 12:296–339.

Okasha, S. (2016). "Population genetics". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E.N. Zalta. Metaphysics Research Lab, Stanford University.

Okonechnikov, K., A. Conesa, and F. Garcıa-Alcalde (2016). "Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data". *Bioinformatics* 32 (2):292–294.

Olalde, I. et al. (2015). "A common genetic origin for early Farmers from Mediterranean Cardial and central European LBK cultures". *Molecular Biology and Evolution* 32 (12):3132–3142.

Olalde, I. et al. (2018). "The Beaker phenomenon and the genomic transformation of northwest Europe". *Nature* 555 (7695):190–196.

Oliver, M. (1992). "Excavation of an Iron Age and Romano-British settlement site at Oakridge, Basingstoke, Hampshire, 1965-6". *Proceedings of the Hampshire Field Club Archaeological Society* 48:55–94.

Olivieri, A. et al. (2017). "Mitogenome diversity in Sardinians: a genetic window onto an Island's past". *Molecular Biology and Evolution* 34 (5):1230–1239.

Orlando, L. et al. (2006). "Revisiting Neandertal diversity with a 100,000 year old mtDNA sequence". *Current Biology* 16 (11):R400–R402.

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

O'Sullivan, N. et al. (2018). "Ancient genome-wide analyses infer kinship structure in an Early Medieval Alemannic graveyard". *Science Advances* 4 (9):eaao1262.

Pääbo, S. et al. (2004). "Genetic analyses from ancient DNA". *Annual Review of Genetics* 38:645–679.

Pacheu-Grau, D. et al. (2012). "Mitochondrial antibiograms in personalized medicine". *Human Molecular Genetics* 22 (6):1132–1139.

Pala, M. et al. (2012). "Mitochondrial DNA signals of Late Glacial recolonization of Europe from Near Eastern refugia". *The American Journal of Human Genetics* 90 (5):915–924.

Palanichamy, M.G. et al. (2015). "West Eurasian mtDNA lineages in India: an insight into the spread of the Dravidian language and the origins of the caste system". *Human Genetics* 134 (6):637–647.

Palomo-Díez, S. et al. (2018). "Kinship analysis and allelic dropout: a forensic approach on an archaeological case". *Annals of Human Biology* 45 (4):1–4.

Parker Pearson, M. (1999). "The Earlier Bronze Age". In: *The Archaeology of Britain: An Introduction from the Upper Palaeolithic to the Industrial Revolution*. Ed. by J. Hunter and I. Ralston. Psychology Press. Chap. 5, p. 77–94.

Patterson, N., A.L. Price, and D. Reich (2006). "Population structure and eigenanalysis". *PLoS Genetics* 2 (12):e190.

Pearson, A. (2006). "Piracy in late Roman Britain: a perspective from the Viking Age". *Britannia* 37:337–353.

Pearson, M.P. et al. (2016). "Beaker people in Britain: migration, mobility and diet". *Antiquity* 90 (351):620–637.

Peltzer, A. et al. (2016). "EAGER: efficient ancient genome reconstruction". *Genome Biology* 17 (1):60.

Pereira, J.B. et al. (2017). "Reconciling evidence from ancient and contemporary genomes: a major source for the European Neolithic within Mediterranean Europe". *Proceedings of the Royal Society B: Biological Sciences* 284 (1851):20161976.

Pereira, L. et al. (2007). "No evidence for an mtDNA role in sperm motility: data from complete sequencing of asthenozoospermic males". *Molecular Biology and Evolution* 24 (3):868–874.

Pereira, L. et al. (2009). "The diversity present in 5140 human mitochondrial genomes". *The American Journal of Human Genetics* 84 (5):628–640.

Pichler, I. et al. (2010). "Drawing the history of the Hutterite population on a genetic landscape: inference from Y-chromosome and mtDNA genotypes". *European Journal of Human Genetics* 18 (4):463.

Pinhasi, R. et al. (2015). "Optimal ancient DNA yields from the inner ear part of the human petrous bone". *PloS One* 10 (6):e0129102.

Pollard, A.M., M. Pellegrini, and J.A. Lee-Thorp (2011). "Technical note: some observations on the conversion of dental enamel $\delta^{18}O_p$ values to $\delta^{18}O_w$ to determine human mobility". *American Journal of Physical Anthropology* 145 (3):499–504.

Pope, A.M. et al. (2011). "Mitogenomic and microsatellite variation in descendants of the founder population of Newfoundland: high genetic diversity in an historically isolated population". *Genome* 54 (2):110–119.

Posth, C. et al. (2016). "Pleistocene mitochondrial genomes suggest a single major dispersal of non-Africans and a Late Glacial population turnover in Europe". *Current Biology* 26 (6):827–833.

Powlesland, D., C. Haughton, and J. Hanson (1986). "Excavations at Heslerton, North Yorkshire 1978–82". *Archaeological Journal* 143 (1):53–173.

Price, A.L. et al. (2006). "Principal components analysis corrects for stratification in genome-wide association studies". *Nature Genetics* 38 (8):904.

Pryor, F. (2004). *Britain AD: A Quest for Arthur, England, and the Anglo-Saxons*. HarperCollins UK.

Quintáns, B. et al. (2004). "Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing". *Forensic Science International* 140 (2-3):251–257.

R Core Team (2018). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: `https://www.R-project.org/`.

Ralf, A. et al. (2018). "Yleaf: software for human Y-chromosomal haplogroup inference from next-generation sequencing data". *Molecular Biology and Evolution* 35 (5):1291–1294.

Ramsey, C.B. (2017). "Methods for summarizing radiocarbon datasets". *Radiocarbon* 59 (6):1809–1833.

Ramsey, C.B. et al. (2002). "Radiocarbon dates from the Oxford AMS system: archaeometry datelist 31". *Archaeometry* 44:1–150.

Ramsey, C.B. et al. (2004a). "Improvements to the pretreatment of bone at Oxford". *Radiocarbon* 46 (1):155–163.

Ramsey, C.B., T. Higham, and P. Leach (2004b). "Towards high-precision AMS: progress and limitations". *Radiocarbon* 46 (1):17–24.

Ramsey, C.B. et al. (2010). "Developments in the calibration and modeling of radiocarbon dates". *Radiocarbon* 52 (3):953–961.

Raule, N. et al. (2014). "The co-occurrence of mtDNA mutations on different oxidative phosphorylation subunits, not detected by haplogroup analysis, affects human longevity and is population specific". *Aging Cell* 13 (3):401–407.

Redfern, R.C., C. Hamlin, and N.B. Athfield (2010). "Temporal changes in diet: a stable isotope analysis of late Iron Age and Roman Dorset, Britain". *Journal of Archaeological Science* 37 (6):1149–1160.

Redfern, R.C. and S.N. DeWitte (2011). "A new approach to the study of Romanization in Britain: a regional perspective of cultural change in late Iron Age and Roman Dorset using the Siler and Gompertz–Makeham models of mortality". *American Journal of Physical Anthropology* 144 (2):269–285.

Reidla, M. et al. (2003). "Origin and diffusion of mtDNA haplogroup X". *The American Journal of Human Genetics* 73 (5):1178–1190.

Reimer, P.J. et al. (2013). "IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP". *Radiocarbon* 55 (4):1869–1887.

Relethford, J.H. and M.H. Crawford (2013). "Genetic drift and the population history of the Irish travellers". *American Journal of Physical Anthropology* 150 (2):184–189.

Renaud, G. et al. (2015). "Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA". *Genome Biology* 16 (1):224.

Renfrew, C. (2000). "Archaeogenetics: towards a population prehistory of Europe". *Archaeogenetics: DNA and the population prehistory of Europe*:3–11.

Richards, J. D. (1999). "The Scandinavian presence". In: *The Archaeology of Britain: An Introduction from the Upper Palaeolithic to the Industrial Revolution*. Ed. by J. Hunter and I. Ralston. Psychology Press. Chap. 11, p. 194–209.

Richards, M.B. et al. (2000). "Tracing European founder lineages in the Near Eastern mtDNA pool". *The American Journal of Human Genetics* 67 (5):1251–1276.

Richards, M.B., C. Capelli, and J.F. Wilson (2008). "Genetics and the origins of the British population". *Encyclopedia of Life Sciences*.

Rizzi, E. et al. (2012). "Ancient DNA studies: new perspectives on old samples". *Genetics Selection Evolution* 44 (1):21.

Robinson, M.S. and M.A. Bidmos (2011). "An assessment of the accuracy of discriminant function equations for sex determination of the femur and tibia from a South African population". *Forensic Science International* 206 (1-3):212–e1.

Roostalu, U. et al. (2007). "Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in west Eurasia: the Near Eastern and Caucasian perspective". *Molecular Biology and Evolution* 24 (2):436–448.

Rootsi, S. et al. (2004). "Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe". *The American Journal of Human Genetics* 75 (1):128–137.

Rösing, F.W. et al. (2007). "Recommendations for the forensic diagnosis of sex and age from skeletons". *HOMO - Journal of Comparative Human Biology* 58 (1):75–89.

Rowley-Conwy, P. (2004). "How the West was lost: a reconsideration of agricultural origins in Britain, Ireland, and aouthern Scandinavia". *Current Anthropology* 45 (S4):S83–S113.

Saag, L. et al. (2017). "Extensive farming in Estonia started through a sex-biased migration from the steppe". *Current Biology* 27 (14):2185–2193.

Saillard, J. et al. (2000). "mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion". *The American Journal of Human Genetics* 67 (3):718–726.

Salemi, M., A.-M. Vandamme, and P. Lemey (2009). *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*. Cambridge University Press.

Sampietro, M.L. et al. (2007). "Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe". *Proceedings of the Royal Society of London B: Biological Sciences* 274 (1622):2161–2167.

Sánchez-Quinto, F. et al. (2012). "Genomic affinities of two 7,000-year-old Iberian hunter-gatherers". *Current Biology* 22 (16):1494–1499.

Sanger, F., S. Nicklen, and A.R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors". *Proceedings of the National Academy of Sciences of the United States of America* 74 (12):5463–5467.

Satterlee Blake, K.A. (2011). "An investigation of sex determination from the subadult pelvis: a morphometric analysis". PhD thesis. University of Pittsburgh.

Saville, A. and Y. Hallén (1994). "The 'Obanian Iron Age': human remains from the Oban cave sites, Argyll, Scotland". *Antiquity* 68 (261):715–723.

Sawicki, M.P. et al. (1993). "Human Genome Project". *The American Journal of Surgery* 165 (2):258–264.

Schiffels, S. et al. (2016). "Iron Age and Anglo-Saxon genomes from east England reveal British migration history". *Nature communications* 7:10408.

Scholz, M.B., C.-C. Lo, and P.S.G. Chain (2012). "Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis". *Current Opinion in Biotechnology* 23 (1):9–15.

Schönberg, A. et al. (2011). "High-throughput sequencing of complete human mtDNA genomes from the Caucasus and west Asia: high diversity and demographic inferences". *European Journal of Human Genetics* 19 (9):988.

Schubert, M. et al. (2012). "Improving ancient DNA read mapping against modern reference genomes". *BMC Genomics* 13 (1):178.

Schubert, M., S. Lindgreen, and L. Orlando (2016). "AdapterRemoval v2: rapid adapter trimming, identification, and read merging". *BMC Research Notes* 9 (1):88.

Schuenemann, V.J. et al. (2017). "Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods". *Nature Communications* 8:15694.

Schulting, R.J. and M.P. Richards (2002a). "Finding the coastal Mesolithic in southwest Britain: AMS dates and stable isotope results on human remains from Caldey Island, South Wales". *Antiquity* 76 (294):1011–1025.

— (2002b). "The wet, the wild and the domesticated: the Mesolithic–Neolithic transition on the west coast of Scotland". *European Journal of Archaeology* 5 (2):147–189.

Semino, O. et al. (2000). "The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective". *Science* 290 (5494):1155–1159.

Seoane, M. et al. (2011). "The mitochondrial genome is a "genetic sanctuary" during the oncogenic process". *PloS One* 6 (8):e23327.

Sequeira, A. et al. (2015). "Mitochondrial mutations in subjects with psychiatric disorders". *PloS One* 10 (5):e0127280.

Shapiro, áB and Michael Hofreiter (2014). "A paleogenomic perspective on evolution and gene function: new insights from ancient DNA". *Science* 343 (6169):1236573.

Sheehan, J. (2008). "The *longphort* in Viking Age Ireland". *Acta Archaeologica* 79 (1):282–295.

Sheridan, J.A. (2010). "The Neolithisation of Britain and Ireland: the 'Big Picture'". In: *Landscapes in Transition*. Ed. by B. Finlayson and G. Warren. Oxbow, p. 89–105.

Shlush, L.I. et al. (2008). "The Druze: a population genetic refugium of the Near East". *PLoS One* 3 (5):e2105.

Sikora, M. et al. (2014). "Population genomic analysis of ancient and modern genomes yields new insights into the genetic ancestry of the Tyrolean Iceman and the genetic structure of Europe". *PLoS Genetics* 10 (5):e1004353.

Silva, M. et al. (2017). "A genetic chronology for the Indian Subcontinent points to heavily sex-biased dispersals". *BMC Evolutionary Biology* 17 (1):88.

Silva, M. et al. (2019). "Once upon a time in the West: The archaeogenetics of Celtic origins". In: *Exploring Celtic Origins: New Ways Forward in Archaeology, Genetics and Linguistics*. Ed. by J. Koch and B. Cunliffe. Oxford: Oxbow Books. Chap. 6, p. 153–191.

Siva, N. (2008). "1000 Genomes Project". *Nature Biotechnology* 26 (256).

Skaletsky, H. et al. (2003). "The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes". *Nature* 423 (6942):825–837.

Skoglund, P. et al. (2013). "Accurate sex identification of ancient human remains using DNA shotgun sequencing". *Journal of Archaeological Science* 40 (12):4477–4482.

Skoglund, P. et al. (2014). "Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers". *Science* 344 (6185):747–750.

Skoglund, P. et al. (2015). "Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds". *Current Biology* 25 (11):1515–1519.

Skonieczna, K. et al. (2015). "Heteroplasmic substitutions in the entire mitochondrial genomes of human colon cells detected by ultra-deep 454 sequencing". *Forensic Science International: Genetics* 15:16–20.

Smith, L.M. et al. (1986). "Fluorescence detection in automated DNA sequence analysis". *Nature* 321 (6071):674.

Soares, P. et al. (2009). "Correcting for purifying selection: an improved human mitochondrial molecular clock". *The American Journal of Human Genetics* 84 (6):740–759.

Soares, P. et al. (2010). "The archaeogenetics of Europe". *Current Biology* 20 (4):R174–R183.

Soares, P. et al. (2012). "The expansion of mtDNA haplogroup L3 within and out of Africa". *Molecular Biology and Evolution* 29 (3):915–927.

Soares, P.A. (2009). "Phylogenetics and phylogeography of human mitochondrial DNA in island southeast Asia". PhD dissertation. University of Leeds.

Soares, P.A. et al. (2016). "Resolving the ancestry of Austronesian-speaking populations". *Human Genetics* 135 (3):309–326.

Soini, H.K. et al. (2012). "Mitochondrial DNA sequence variation in Finnish patients with matrilineal diabetes mellitus". *BMC Research Notes* 5 (1):350.

Soini, H.K. et al. (2013). "Mitochondrial DNA variant m.15218A>G in Finnish epilepsy patients who have maternal relatives with epilepsy, sensorineural hearing impairment or diabetes mellitus". *BMC Medical Genetics* 14 (1):73.

Stewart, T.D. (1968). "Identification by the skeletal structures". *Gradwohl's Legal Medicine*:123–154.

Sukernik, R.I. et al. (2012). "Mitochondrial genome diversity in the tubalar, even, and ulchi: Contribution to prehistory of native Siberians and their affinities to Native Americans". *American Journal of Physical Anthropology* 148 (1):123–138.

Taanman, J.W. (1999). "The mitochondrial genome: structure, transcription, translation and replication". *Biochimica et Biophysica Acta* 1410 (2):103–123.

The UK10K Consortium (2015). "The UK10K project identifies rare variants in health and disease". *Nature* 526 (7571):82.

Thomas, J. (1988). "Neolithic explanations revisited: the Mesolithic-Neolithic transition in Britain and south Scandinavia". 54:59–66.

Thomas, M.G., M.P.H. Stumpf, and H. Härke (2006). "Evidence for an apartheid-like social structure in early Anglo-Saxon England". *Proceedings of the Royal Society of London B: Biological Sciences* 273 (1601):2651–2657.

Torroni, A. et al. (1998). "mtDNA analysis reveals a major Late Paleolithic population expansion from southwestern to northeastern Europe". *The American Journal of Human Genetics* 62 (5):1137–1152.

Torroni, A. et al. (2006). "Harvesting the fruit of the human mtDNA tree". *Trends in Genetics* 22 (6):339–345.

Tuross, N. (1994). "The biochemistry of ancient DNA in bone". *Experientia* 50 (6):530–535.

Valdiosera, C. et al. (2018). "Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia". *Proceedings of the National Academy of Sciences of the United States of America*:201717762.

Van Der Walt, E.M. et al. (2012). "Characterization of mtDNA variation in a cohort of South African paediatric patients with mitochondrial disease". *European Journal of Human Genetics* 20 (6):650.

Van Klinken, G.J. (1999). "Bone collagen quality indicators for palaeodietary and radiocarbon measurements". *Journal of Archaeological Science* 26 (6):687–695.

Venter, J.C. et al. (2001). "The sequence of the human genome". *Science* 291 (5507):1304–1351.

Verano, J.W., L.S. Anderson, and R. Franco (2000). "Foot amputation by the Moche of ancient Peru: osteological evidence and archaeological context". *International Journal of Osteoarchaeology* 10 (3):177–188.

Vianello, D. et al. (2013). "HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment". *Human Mutation* 34 (9):1189–1194.

Waddington, C. and C. Bonsall (2016). "Archaeology and Environment on the North Sea Littoral. A Case Study from Low Hauxley". In: Archaeological Research Services.

Waldron, T., G.M. Taylor, and D. Rudling (1999). "Sexing of Romano-British baby burials from the Beddingham and Bignor villas". *Sussex Archaeological Collections* 137:71–79.

Weale, M.E. et al. (2002). "Y chromosome evidence for Anglo-Saxon mass migration". *Molecular Biology and Evolution* 19 (7):1008–1021.

Weaver, D.S. (1980). "Sex differences in the ilia of a known sex and age sample of fetal and infant skeletons". *American Journal of Physical Anthropology* 52 (2):191–195.

Weiss, K.M., K.K. Kidd, and J.R. Kidd (1992). "Human genome diversity project". *Evolutionary Anthropology: Issues, News, and Reviews* 1 (3):80–82.

Weissensteiner, H. et al. (2016). "HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing". *Nucleic Acids Research* 44 (W1):W58–W63.

White, M.J. and D.C. Schreve (2000). "Island Britain – Peninsula Britain: palaeogeography, colonisation, and the Lower Palaeolithic settlement of the British Isles". *Proceedings of the Prehistoric Society* 66:1–28.

Whitehouse, N.J. et al. (2014). "Neolithic agriculture on the European western frontier: the boom and bust of early farming in Ireland". *Journal of Archaeological Science* 51:181–205.

Whittle, A.W.R., F.M.A. Healy, and A. Bayliss (2011). *Gathering time: dating the early Neolithic enclosures of southern Britain and Ireland*. Oxbow Books.

Willerslev, E. and A. Cooper (2005). "Ancient DNA". *Proceedings of the Royal Society of London B: Biological Sciences* 272 (1558):3–16.

Wilson, J.F. et al. (2001). "Genetic evidence for different male and female roles during cultural transitions in the British Isles". *Proceedings of the National Academy of Sciences* 98 (9):5078–5083.

Winther Pedersen, M. et al. (2015). "Ancient and modern environmental DNA". *Philosophical Transactions of the Royal Society Biology* 370 (1660):20130383.

Woodbridge, J. et al. (2014). "The impact of the Neolithic agricultural transition in Britain: a comparison of pollen-based land-cover and archaeological ^{14}C date-inferred population change". *Journal of Archaeological Science* 51:216–224.

Woodham-Smith, C. et al. (1962). "The Great Hunger: Ireland 1845-9." *The Great Hunger: Ireland 1845-9.*

Woodman, P.C. (1986). "Problems in the colonisation of Ireland". *Ulster Journal of Archaeology* 49:7–17.

Wright, S. (1949). "The genetical structure of populations". *Annals of Human Genetics* 15 (1):323–354.

Wygal, B.T. and S.M. Heidenreich (2014). "Deglaciation and human colonization of northern Europe". *Journal of World Prehistory* 27 (2):111–144.

Yang, D.Y. et al. (1998). "Technical note: improved DNA extraction from ancient bones using silica-based spin columns". *American Journal of Physical Anthropology* 105 (4):539–543.

Zaragoza, M.V. et al. (2010). "Mitochondrial DNA variant discovery and evaluation in human cardiomyopathies through next-generation sequencing". *PloS One* 5 (8):e12295.

Zsurka, G. et al. (2007). "Inheritance of mitochondrial DNA recombinants in double-heteroplasmic families: potential implications for phylogenetic analysis". *The American Journal of Human Genetics* 80 (2):298–305.

Appendix A

Publications arising from this thesis

A.1 Published book chapters

- **Dulias K.** and Edwards C.J. (2019) Genetic analysis. In: *Excavations at Milla Skerra, Sandwick, Unst. Rhythms of Life in Iron Age Shetland* (ed. O. Lelong). Oxbow Books, Oxford; pp. 114-115.

- Silva M., **Dulias K.**, Oteo-Garcia G., Gandini F., Edwards C., Pala M., Soares P., Wilson J.F. and Richards M.B. (2019) Once upon a time in the West: The archaeogenetics of Celtic origins. In: *Exploring Celtic Origins: New Ways Forward in Archaeology, Genetics and Linguistics* (eds. J. Koch and B. Cunliffe). Oxbow Books, Oxford; pp. 153-191

A.2 Published article

- Olalde I., Mallick S., Patterson N., Rohland N., Villalba-Mouco V., Silva M., **Dulias K.**, Edwards C.J., Gandini F., Pala M., Soares P., Ferrando-Bernal M., Adamski N., Broomandkhoshbacht N., Cheronet O., Culleton B.J., Fernandes D., Lawson A.M., Mah M., Oppenheimer J., Stewardson K., Zhang Z., Jiménez Arenas J.M., Toro Moyano I.J., Salazar García D.C., Castanyer P., Santos M., Tremoleda J., Lozano M., Fernández-Eraso J., Mujika-Alustiza J.A., Barroso C., Bermúdez F.J., Burch J., Coromina N., Viguera Mínguez E., Vivó D., Cebrià A., Fullola J.M., García-Puchol O., Morales J.I., Oms X., Vergès J.M., Díaz-Carvajal A., Ollich-Castanyer I., Garcia Borja P., Silva A.M., Alonso-Fernández C., Jiménez Echevarría J., Moreno Márquez A., Ramos-García P., Muñoz J.R., Vijande Vila E., Lillios K.T., Mack J., Waterman A., Delibes G., Agustí B.,

Codina F., Esparza A., de Prado G., Estalrrich A., Finlayson C., Finlayson G., Finlayson S., Giles-Guzmán F., Majó T., Rosas A., Velasco Vázquez J., Aguilella Arzo G., Barciela G.V., Benítez de Lugo Enrich L., Benito Sánchez M., Garcia Atienzar G., Hernández Pérez M.S., Llanos A., Carrión Marco Y., Collado Beneyto I., Fernández Flores A., López-Serrano D., Sanz Tormo M., Valera A.C., Blasco C., Daura J., De Pedro Michó M.J., Diez-Castillo A.A., Flores-Fernández R., Francès Farré J., Gonçalves V.S., Guerra-Doce E., Herrero-Corral A.M., Juan-Cabanilles J., Liesau C., López-Cachero F.J., López-Reyes D., McClure S.B., Merino Pérez M., Foix A.O., Pascual Berlanga G., Ríos P., Sanz Borràs M., Sousa A.C., Vidal Encinas J.M., Kennett D.J., Richards M.B., Alt K.W., Haak W., Pinhasi R., Lalueza-Fox C. and Reich D. (2019) The genomic history of the Iberian Peninsula over the past 8000 years. *Science*.

- Sheridan A., Armit I., Reich D., Booth T., Bernardos R., Barnes I., Thomas M., Charlton S., Craig O., Lawson J., **Dulias K.**, Edwards C.J., Pala M., Richards M.B., Margaryan A., Kristiansen K., Willerslev E., Allentoft M., Britton K., Noble G., Girdland Flink L., Talamo S., Curtis N., Cooper A., Cole S. and Brown L. (2018) A summary round-up list of Scottish archaeological human remains that have been sampled/analysed for DNA as of January 2019. *Discovery and Excavation in Scotland*, 19.

A.3 Submitted article

- **Dulias K.**, Birch S., Wilson J.F., Justeau P., Gandini F., Flaquer A., Soares P., Pala M., Richards M.B. and Edwards C.J. (submitted) Maternal relationships within an Iron Age burial at the High Pasture Cave, Isle of Skye, Scotland. *Journal of Archaeological Science*.

- Silva M., Justeau P., Rodrigues S., Oteo-Garcia G., **Dulias K.**, Foody G., Fichera A., Yau B., Rito T., Wilson J.F., Gandini F., Edwards C.J., Pala M., Soares P.A. and Richards M.B. (submitted) Untangling Neolithic and Bronze Age mitochondrial lineages in South Asia. *Annals of Human Biology*.

A.4 In preparation

- Cole G., Ditchfield P.W., **Dulias K.**, Edwards C.J., Reynolds A. and Waldron T. (in prep.) Crime and Punishment? Archaeological evidence

for facial mutilation described in 9th century Anglo-Saxon English law codes.

- **Dulias K.** and Edwards C.J. (in prep.) Ancient mitochondrial DNA analysis of human remains from High Pasture Cave. In: *High Pasture Cave: Ritual, Memory and Identity in the Iron Age of Skye* (eds. S. Birch, G. Cruickshanks and J. Mckenzie). Oxbow Books, Oxford.

Appendix B

Phylogeographic analysis of the British Isles

B.1 Introduction

The geographic distribution of genetic lineages, their phylogeny and their time depth are the basis of phylogeographic analysis. Non-recombining DNA (the mitochondrial genome and the non-recombining region of the Y-chromosome) is widely used to trace genetic lines of descent. The analysis of non-recombining DNA sequences, with the aim of identifying and dating migrations into new geographic areas, is called founder analysis. Founder sequences are used to date lineages that have arisen autochthonously within a specific population or region after immigration from an external source (Richards et al., 2000; Macaulay et al., 2013). This approach uses the variation that occurs in sequences that migrated from the source to the settled population as means of dating them, and being able to distinguish the diversity which has accumulated following the settlement event, from the diversity that existed before, is key to this analysis (Macaulay et al., 2013).
Due to back migration and a generally non-unidirectional flow of migration complicating the identification of founder types, modification of the method and certain assumptions are required. Archaeological information or phylogeographic analysis of the lineages in the two (source and sink) regions of interest is used to assume a migration model. This assumed archaeological model allows for testing the extent of dispersal in each time window within the model (Macaulay et al., 2013). The procedure was formalised by Richards et al. (2000) by developing criteria that take the effects of both gene flow and recurrent mutation into account. These criteria, $f1$ and $f2$, were designed to minimise their effects requiring either one or two derived branches within the source, corresponding to $f1$ and $f2$, respectively (Richards et al., 2000). High-resolution studies can facilitate the identification of founder types that

have moved multiple times (Macaulay et al., 2013). The use of complete mitochondrial sequences for founder analysis, rather than short fragments or SNP capture data, allows for a more nuanced picture, as it accounts for the entire variation of the mitochondrial genome. Therefore, complete mitochondrial genomes were generated to build a large dataset of modern sequences to identify founder types in Britain.

B.2 Materials and Methods

New and published modern whole mtDNA genome sequences were used for the phylogeographic analysis.

B.2.1 New modern samples

For the new sequences sampled as part of this study, consented buccal swabs and/or extracted DNA sent by collaborators, originating from the British Isles and Ireland, France and Germany, were analysed. A subset of 8,359 individuals, provided by BritainsDNA (now incorporated into Source BioScience) in collaboration with Prof. Jim Wilson, was chosen for whole mitogenome sequencing after analysis of frequency patterns. Out of the total dataset, 5,337 were British or Irish. After several months of negotiations, Source BioScience randomly selected 1,618 DNA extracts from the overall available dataset. All available Welsh and Irish samples were sequenced, as well as a randomised number of samples from Scotland and England. The final subset for complete mitogenome sequencing consisted of 677 samples.
A total of 1,381 samples were processed for NGS sequencing of the complete mitochondrial genome. This included modern samples from Britain, Ireland, France and Germany (see Table B.1 for total numbers). 60 Irish and British samples were collected by Dr. Ceiridwen J. Edwards, M. George B. Foody and Peter Harrop. Peter Harrop extracted at the University of Huddersfield under my supervision, while I did the whole genome amplification and further processing. 677 DNA extracts from Britain and Ireland came from Source BioScience, 308 DNA extracts from Germany and 336 DNA extracts from France, were amplified and resulted in a total of 1,381 samples being processed for NGS sequencing of the complete mitochondrial genome. Furthermore, newly sequenced unpublished samples from Belgium (n=16), Greece (n=70), Italy (n=42), Portugal (n=12) and Spain (n=88), were added to the overall batch of sequences for analysis, as well as a total of 3,872 new

sequences from Germany, the Netherlands and Shetland (see Table B.1 for details).

B.2.2 Published data

To facilitate founder analysis all published and online available modern complete mitochondrial genomes were used. Additionally, the whole mitochondrial genome data from Byrne et al. (2018) was used to facilitate the analysis of the generated Irish data, as the sample number of Irish individuals in this study was comparably low. Due to the immense size of the dataset and the shared use of the data for different projects the published data was split into different haplogroups and within the timeframe of this project I generated the phylogenetic trees in xml format for haplogroups H3, H6, H7, H8 and H31. These were combined with all other remaining trees generated by members of the group to run a complete founder analysis on Europe, including the British Isles. Drs. Pedro Soares, Teresa Rito, and Francesca Gandini were involved in the compiling of published sequence data and the generation of phylogenetic trees, as well as Alessandro Fichera, M. George B. Foody, Gonzalo Oteo-Garcia and Marina Silva. Details about the phylogenetic reconstruction can be found in Appendix C.

TABLE B.1: Overview of all new and published samples used in this study.

Country	Sample number	Source	Data type	Collaborators	Published
Belgium	16	DNA extracts	whole-genome sequences	A. Fichera	No
Britain	5,337	Mitotype	SNP data	Source BioScience/Prof. Jim Wilson	No
	677	DNA extracts	whole-genome sequences		No
	60	buccal swabs	whole-genome sequences	Dr. C.J. Edwards, M.G.B. Foody, P. Harrop	No
French	336	DNA extracts	whole-genome sequences		No
Germany	2,881	Sequences	Sequence data	KORA project, Dr. Antonia Flaquer	No
	308	DNA extracts	whole-genome sequences		No
Greece	70	DNA extracts	whole-genome sequences	M.G.B. Foody	No
Ireland	408	Sequences	Sequence data	Byrne et al. (2018)	Yes
Italy	42	DNA extracts	whole-genome sequences	A. Fichera	No
Netherlands	491	Sequences	Sequence data	GONL project, Prof. Walther Parson	No
Portugal	12	DNA extracts	whole-genome sequences	M. Silva	No
Shetland	500	Sequences	Sequence data	Prof. Jim Wilson	No
Spain	88	DNA extracts	whole-genome sequences	G. Oteo-Garcia, M. Silva	No

B.2.3 Modern sample processing

B.2.3.1 DNA extraction of buccal swabs

DNA from buccal swabs was extracted using the PureLink Genomic DNA Kit (ThermoFisher Scientific). The buccal swabs were placed in a sterile 2 ml microcentrifuge tube using tweezers, and the complete cotton swab was pulled off the wooden stick. 400 µl PBS and 20 µl Proteinase K were added to the sample directly, and mixed by pipetting. An equal volume of PureLink Genomic Lysis/Binding Buffer was added to the lysate, and mixed by brief vortexing. After this step the manufacturer's protocol was followed from the incubation step without any further changes. The elution volume was 100 µl and the samples were stored in 1.5 ml centrifuge tubes at -20 °C. The quality of the DNA extracts was assessed by measuring their concentrations using the Qubit 3.0 Fluorometer with the High Sensitivity assay for dsDNA.

B.2.3.2 DNA amplification of complete mitochondrial genomes

The Polymerase Chain Reaction (PCR) was used to increase the targeted DNA fragments in comparison to other genomic DNA in the DNA extract. The set-up for modern samples was undertaken in the modern genetics lab at Huddersfield. To amplify the complete mitochondrial genome (16,568 bp), two fragments were used for each sample. The amplification was set up the same way for both fragments, only changing the primers for each fragment. Nuclease-free water (11 µl), GoTaq®Long PCR Master Mix (PROMEGA) 2X (12.5 µl), 5871for (GCTTCACTCAGCCATTTTACCT) (0.5 µl, 10 pmol), 13829rev (AGTCCTAGGAAAGTGACAGCGA) (0.5 µl, 10 pmol) as primers for the first fragment and 13477for (GCAGGAATACCTTTCCTCACAG) (0.5 µl, 10 pmol) and 6345rev (AGATGGTTAGGTCTACGGAGGC) (0.5 µl, 10 pmol) for the second fragment, and 2 µl of sample DNA. Amplification took place under the following thermal cycling conditions: 2 minutes at 94 °C; 30 cycles at 30 seconds at 94 °C, 30 seconds at 55 °C, 9 minutes at 65 °C, and a final extension of 10 minutes at 72 °C. Amplification products were visually checked to be the correct size using 1% agarose gels running at 80 V for 45 minutes.

B.2.3.3 Purification of PCR products

The PCR products were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, USA) to remove inhibiting salts and dNTPs, and to concentrate the DNA content. The elution volume used was 50 µl.

B.2.3.4 Measurement of DNA quantity

A measurement of DNA quantity was needed to calculate the necessary volume of each sample, in order to pool the products of the two fragments together equimolarly for high-throughput sequencing. DNA concentration was assessed using 1 µl of the purified PCR product using the Qubit® dsDNA High Sensitivity assay kit on the Qubit® 3.0 Fluorometer. Each sample was measured twice, and the mean value of each sample was used for further calculations in order to minimise pipetting or machine errors.

B.2.3.5 Modern sample preparation for NGS sequencing of mitogenomes

After quantification both fragments were pooled equimolarly with a final volume of 40 µl and a concentration of 1 ng per µl. The pooled samples were then sequenced on a MiSeq250 (PE-Nano x1) by the Earlham Institute (Norwich, UK).

B.2.3.6 Data analysis of modern mitogenome sequences

B.2.3.6.1 Newly sequenced genomes

EAGER (Peltzer et al., 2016) was used for the quality control, removal of duplicates, mapping and SNP calling, using the following settings: AdapterRemoval for the removal of adapters and read merging; bwa-mem to align sequences to the rCRS (Andrews et al., 1999); duplicates were removed using DeDup; HaplotypeCaller was used for SNP calling, with 'Downsampling' set to 250, 'Ploidy' set to 100 and 'Standard Call Confidence' set to 30, and for SNP filtering the 'Minimum Coverage Filter' was set to 2.
In order to filter the mutations and heteroplasmies bcftools v.1.3 was used (Li et al., 2009b; Li, 2011). Mutations with an allele frequency >0.7 were considered variants, whilst mutations with an allele frequency >0.3 and <0.7 were considered heteroplasmies. All heteroplasmies were manually checked in Geneious® (version 6.1.8). Mutations in non-considered positions (around 302-315 including 309.1C(C) and 315.1C, AC insertions and deletions at 515-522, 16182C, 16183C, and 16193.1C(C)) were excluded from analyses. The final haplotypes were used to determine the haplogroup designations with HaploGrep 2 (Weissensteiner et al., 2016) and converted into FASTA files with HaploFasta. Haplogroups of the SNP data of the 5,337 British and Irish individuals from Source BioScience were as well determined using HaploGrep 2 (Weissensteiner et al., 2016).

B.2.3.6.2 Published genomes

Published sequences were aligned to the rCRS using Sequencher™ version 5.1 (`http://www.genecodes.com`), and converted to variants with mtDNAGeneSyn (Pereira et al., 2009). Gaps were cleaned and haplogroup assignment verified with Haplogrep 2 (Weissensteiner et al., 2016).
The 408 mitochondrial genomes published by Byrne et al. (2018) were realigned to the rCRS using bwa aln with -q 15 and -l 1024. Samtools was used to filter, sort and remove duplicates from the alignments (Li et al., 2009b). Alignments were filtered for quality 30 and F4. Variants were called using GATK HaplotypeCaller with a standard call confidence of 30 and sample ploidy 100, in order to identify heteroplasmies (see section B.2.3.6.1). Overall, 349 sequences of the Byrne et al. (2018) dataset passed all quality filters and had accompanying geographical information of the maternal grandmother, and were used for further analysis.

B.2.3.7 Founder Analysis

For the founder analysis a software tool developed in the group of Dr. Pedro Soares, University of Minho, Portugal, was used. To estimate arrival times, founder types were identified and, on the basis of their coalescence times partitioned into clusters that derived from the founder types in the sink population (Soares et al., 2016).
The overall founder analysis of the British Isles, with Britain as a sink population and various mainland European countries as possible source populations, is still ongoing. Results presented in this chapter are founder analyses of haplogroup H3, which is among the most frequent haplotypes of major haplogroup H in Britain, with around 12% of all haplogroup H from Britain being H3.
In order to identify founder types of haplogroup H3 that derived from a source population in central Europe, the following countries were considered to represent central Europe: Austria, Belgium, Czech Republic, Denmark, France, Germany, Luxembourg, the Netherlands and Poland. The total number of samples from these countries in the dataset analysed was 260, whilst the total number of samples in the sink population (the British Isles, *i.e.* England, Scotland, Wales and Ireland) accounted for 108.
For the founder analysis with founder types originating in Iberia, Portugal and Spain (including the Basque country) were considered as the source population. The total number of samples included in the source was 154, and the

samples in the sink were the same as described above.
Founder analyses were carried out including an initial 200 year scan, which is a very fine-scale partitioning of the data, to evaluate arrivals over time more objectively than choosing only archaeological dates. The approach by Saillard et al. (2000) was used to estimate errors in order to allow for non-star-like clustering of founders. The number of samples in the ρ estimation was replaced by an effective number of samples, which was based on the number that would be present in a completely star-like network (Soares et al., 2012). ρ was defined as the average distance in mutations from each sample to the root or most recent common ancestor of the tree/network (Forster et al., 1996). A mutation rate of one mutation every 2,600 years was employed, because the founder analysis software does not use a time-dependent mutation rate, therefore, the average of the time-dependent age estimates was used (Soares, 2009). The calculation of the time-dependent mutation rate accounts for the effect of purifying selection on more recent nodes of the complete human mitochondrial tree (Soares et al., 2009). Different criteria of varying stringency levels for the identification of founder types were employed, *i.e.* $f1$ and $f2$, which are threshold levels minimising the effects of recurrent mutations (Richards et al., 2000), and specifying that matches in the source should have one ($f1$ or two ($f2$) derived branches, but no matches on the tips of the source phylogenies (Soares, 2009). The molecular clock was used to convert the values to time depth as a proxy for minimum arrival times of founder clusters in the sink population (Richards et al., 2000).

B.3 Results

B.3.1 mtDNA SNP data of the British Isles

The mtDNA SNP data of Britain consisted of 2,303 samples from England. The English data included 104 haplogroups, of which 44.90% belonged to 47 subhaplogroups of H and nine subhaplogroups of HV (Figure B.1A). Other diverse haplogroups presented in this dataset were U (eight subhaplogroups; Figure B.2A), haplogroup V (12 subhaplogroups) and haplogroup W (six subhaplogroups). The most frequent (>5.00%) haplogroups among the English samples were: H1 (17.24%), U5 (11.12%), J1 (8.55%), T2 (7.69%) and K1 (7.21%)(Table B.3). For Scotland, 1,926 samples were present in this dataset, including 88 haplogroups. The 43 subclades belonging to haplogroup H

made up around 40.39% of the data (details in Figure B.1B), while other diverse haplogroups were V, U, HV, W and I. The most frequent (>5.00%) subhaplogroups among the Scottish were: H1 (13.24%), J1 (12.98%), U5 (8.10%), K1 (7.48%) and T2 (6.65%)(Table B.3). Within macro-haplogroup U the subhaplogroups with the highest frequency in Scotland were K1 (35%), U5a (24%) and U5b (14%)(Figure B.2). The combined Irish dataset (newly generated sequences n=374, and Byrne et al. (2018) data n=349) consisted of 723 sequences, of which 39.56% belonged to haplogroup H and its subclades (Figure B.1C). Haplogroup U had an overall frequency of 25.45%, with K1 accounting for 8.30%, U5a for 6.22%, U5b for 3.87%, U4 for 2.35%, and K2,U1,U3, U6 and U8 each accounting for less than one percent frequency (Figure B.2C). Haplogroup J had a frequency of 12.03%, with J1 accounting for the majority of it (Figure B.3). The most frequent subhaplogroups in Ireland were H1 with 13.97%, J1 with 11.20%, K1 with 8.30%, T2 with 7.88%, H3 with 7.61 %, and U5a with 6.22% (Table B.3). The Welsh data consisted of 733 samples, including 60 subhaplogroups with 43.93% belonging to 32 subclades of H (details of the subhaplogroups of H are given in Figure B.1). Haplogroups HV, U, V and W were other diverse haplogroups of the Welsh subset. H1 (12.96%), J1 (12.14%), K1 (8.59%), U5 (6.82%), H3 (5.59%) and T2 (5.59%) were the most frequent subhaplogroups (>5.00%) of the Welsh samples (Table B.2). The different contributions of the subhaplogroups of U and JT are given in Figures B.2 and B.3, respectively. An overview of the major haplogroups and their frequency and distribution across the British Isles based on this SNP dataset is given in Table L.1 and Figure L.1 in the Appendix.

Considering the different haplogroup frequencies of the regions within the British Isles, more details appear. The frequency of H1 ranged between 6.13-19.53%, with the minimum in North Wales and the maximum in the Ireland Ulster. Next to haplogroup H1, J1 (0-15.63%) and U5 (4.69-14.29%) were the overall most frequent haplogroups. For more details on the haplogroup frequency and composition per region please refer to Tables L.2-L.3 and Figures B.1-B.3, as well as Figures L.2-L.5 in the appendix. Since haplogroup H had the highest frequency in all countries of the British Isles, the frequencies of its main subhaplogroups was disseminated (Figure B.1). Subhaplogroups H1 and H3 were the most frequent in all four countries ranging from 28-38% frequency and 10-17%, respectively. In England and Wales subhaplogroup H5 reached a frequency of 9% and 8%, respectively, whereas in Ireland and Scotland it had a frequency of 6% and 7%, respectively. In Wales subhaplogroup H6 reached a frequency of 11%, while in Ireland, Scotland and England it

ranged between 4-6% (Figure B.1). The frequency of subhaplogroup H2 differed from 5% in England to 9% in Wales, with 7% and 8% in Scotland and Ireland. H11 was most frequent in Ireland with 4%, its frequency in Scotland and Wales was 3%, while in England it occured with less than 2% and was included in the minor H clades in Figure B.1. The frequency of HV also differed quite strongly between the countries of the British Isles. Its highest frequency was in Ireland with 8%, its frequency in England and Wales was 4% and 3% in Scotland.

TABLE B.2: Haplogroup frequencies per country of the British Isles. Frequencies are given in percent.

sample number	Haplogroups															
	Countries	H	U2	U4	U5a	U5b	U6	U other	K	T1	T2	J1	J2	I	V	HV	minor
2303	England	43.25	0.61	2.39	6.12	4.99	0.09	1.86	8.81	1.82	7.69	8.55	1.48	3	2.91	1.65	4.76
723	Ireland	39.56	2.63	2.35	6.22	3.87	0.28	0.83	9.27	1.94	7.88	11.20	0.83	3.18	3.32	2.77	3.17
1926	Scotland	40.39	1.14	2.28	5.14	2.96	0.05	1.03	8.88	1.92	6.65	12.98	2.44	4.62	3.69	1.3	4.51
733	Wales	43.93	1.09	0.68	3.41	3.41	0	1.37	11.73	2.32	5.59	12.14	0.95	5.18	3.68	1.91	2.59

TABLE B.3: Haplogroup frequencies per region within the British Isles in percent. Additionally, sample numbers per region are given.

sample number	Haplogroups																
	Regions	H	U	J	K	T	I	V	HV	W	X	L	N	R	M	A	B
615	England SE	41.5	15.6	11.5	8.9	10.4	2.4	3.7	1.0	1.5	1.5	1.1	0.2	0.3	0.3		
215	England E	46.5	14.4	12.1	5.6	7.0	3.3	1.4	2.3	2.3	3.3	0.5		0.9		0.5	
417	England C	44.1	13.4	9.4	10.3	11.3	3.1	1.9	2.2	1.2	2.4	0.2		0.2	0.2		
292	England Yorks	41.8	20.9	7.9	10.3	8.2	3.1	2.1	1.4	1.0	0.7	0.7	0.7	0.7		0.3	0.3
472	England N	43.6	16.9	8.7	7.0	9.3	3.8	3.4	2.1	2.1	1.7	0.4	0.4	0.4			
193	England SW	43.5	15.5	9.3	11.9	9.3	2.6	4.1	1.6		1.0			0.5		0.5	
128	Ireland Ulster	36.7	11.7	17.2	7.8	8.6	4.7	4.7	3.1	3.1	2.3						
86	Ireland Leinster	45.3	18.6	10.5	9.3	9.3	2.3	1.2	2.3	1.2							
75	Ireland Munster	33.3	13.3	9.3	14.7	13.3	2.7	6.7	2.7	2.7			1.3				
44	Ireland Connacht	38.6	22.7	4.5	6.8	18.2			4.5	2.3	2.3						
278	Scotland SE	38.5	11.9	18.3	5.0	9.0	5.0	4.3	1.8	2.2	2.5	0.7	0.7				
742	Scotland SW	40.2	12.8	13.7	11.2	9.3	3.5	4.9	0.8	2.0	1.1	0.1	0.7				
346	Scotland C	42.2	13.3	13.3	8.7	9.5	5.5	2.3	1.4	2.3	0.9	0.3	0.3				
270	Scotland NE	44.8	11.5	16.7	8.1	5.2	4.4	2.2	2.2	1.1	2.6	0.7	0.4				
177	Scotland N&W	38.4	15.8	14.1	7.9	9.6	4.0	2.3	1.1	2.8	2.8	1.1					
499	Wales S	47.5	10.4	9.4	13.2	8.4	4.6	2.0	1.4	1.4	1.6						
212	Wales N	34.9	8.5	21.7	8.5	6.6	7.1	7.5	3.3	0.5	1.4						
5075	Total	42.1	14.0	12.2	9.4	9.1	3.8	3.3	1.7	1.7	1.7	0.4	0.3	0.2	0.1	0.1	0.0

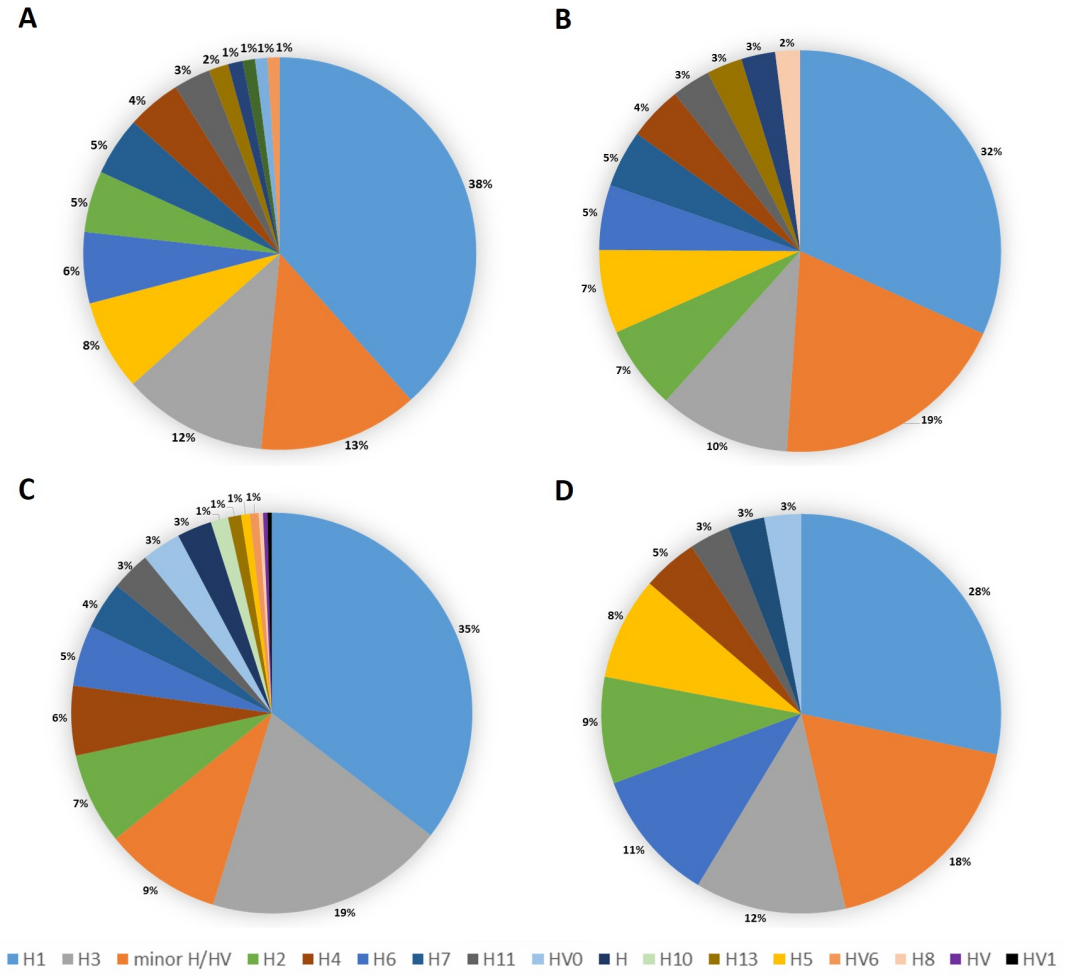

FIGURE B.1: Haplogroup H compositions of the Source BioScience SNP data and the Irish sequence data from Byrne et al. (2018) in the four countries of the British Isles. A – England, minor H/HV consists of 32 subhaplogroups with each less than 1%; B – Scotland, minor H/HV consists of 37 subhaplogroups with each less than 1%; C – Ireland, minor H/HV consists of 13 subhaplogroups with less than 0.35%, which is the frequency of H8, HV and HV1; D – Wales, minor H/HV consists of 27 subhaplogroups with each less than 1%.

B.3.2 Phylogeographic analysis of complete mtDNA genomes

A random subset of 677 samples from the British Isles Source BioScience dataset were used for NGS sequencing, along with additional British and Irish samples (60 in total), samples from Germany (308 in total) and France (336 in total). Out of the 1,381 processed samples 106 failed at the sequencing stage or did not pass quality filters, leaving a final dataset of 1,275 sequences (736 British and Irish samples, 328 French and 211 German).

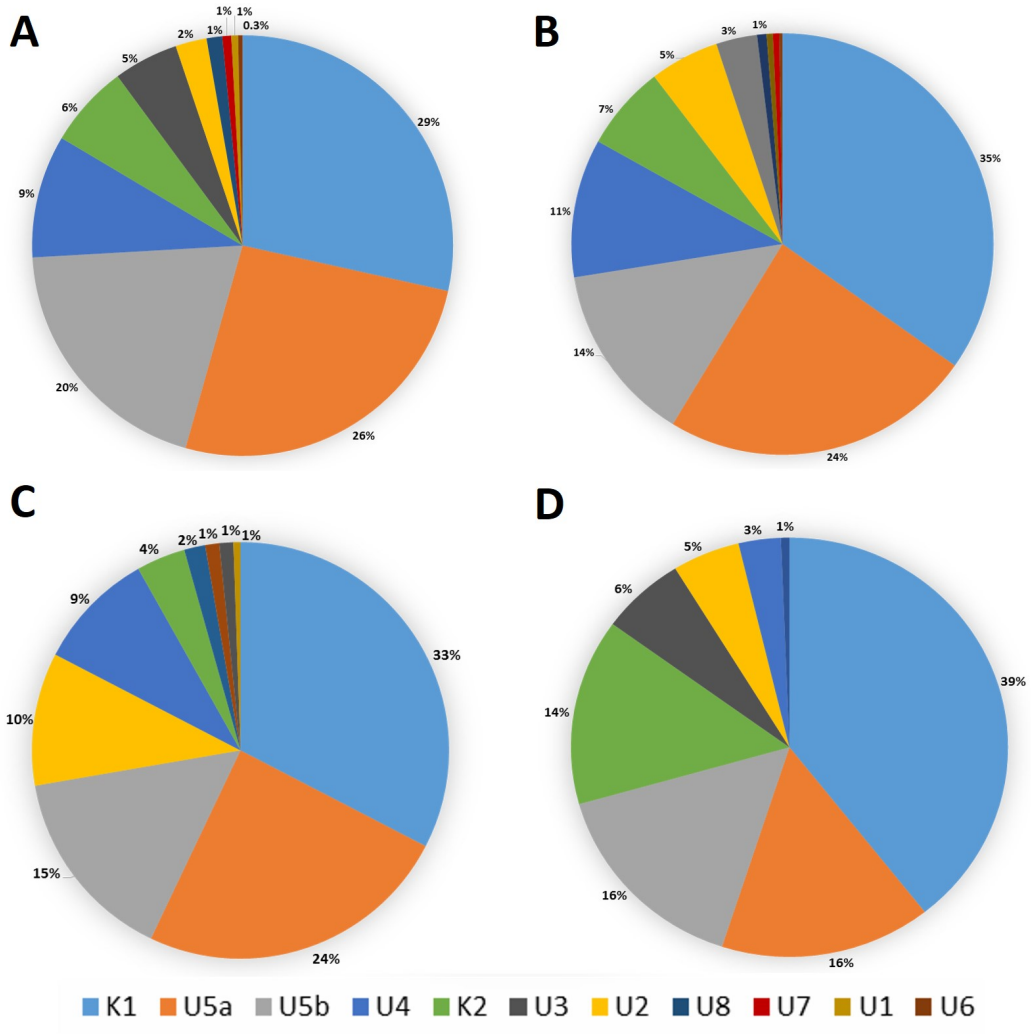

FIGURE B.2: Haplogroup U (including K) composition of the Source BioScience SNP data and the Irish sequence data from Byrne et al. (2018) in the four countries of the British Isles. A – England; B – Scotland; C – Ireland; D – Wales.

B.3.2.1 Haplogroup composition of the British Isles and Ireland

The dataset from the British Isles and Ireland consisted of 736 samples. Of these 192 were from England, 142 from Ireland, 147 from Scotland, 235 from Wales and 20 of unknown geographic origin within the British Isles. As the haplogroup frequencies of the generic samples with unknown geographic origin would not add much detail to the frequency overview, they were excluded from the frequency analysis due to their low number.

In this population dataset, the haplogroups with the highest frequencies were H (36-43%), J (9-20%), U (9-15%), K (7-12%) and T (7-10%). The dataset included 418 different haplotypes, of which 183 belonged to haplogroup H and eight to HV. Other diverse haplogroups were U (with 60 haplotypes),

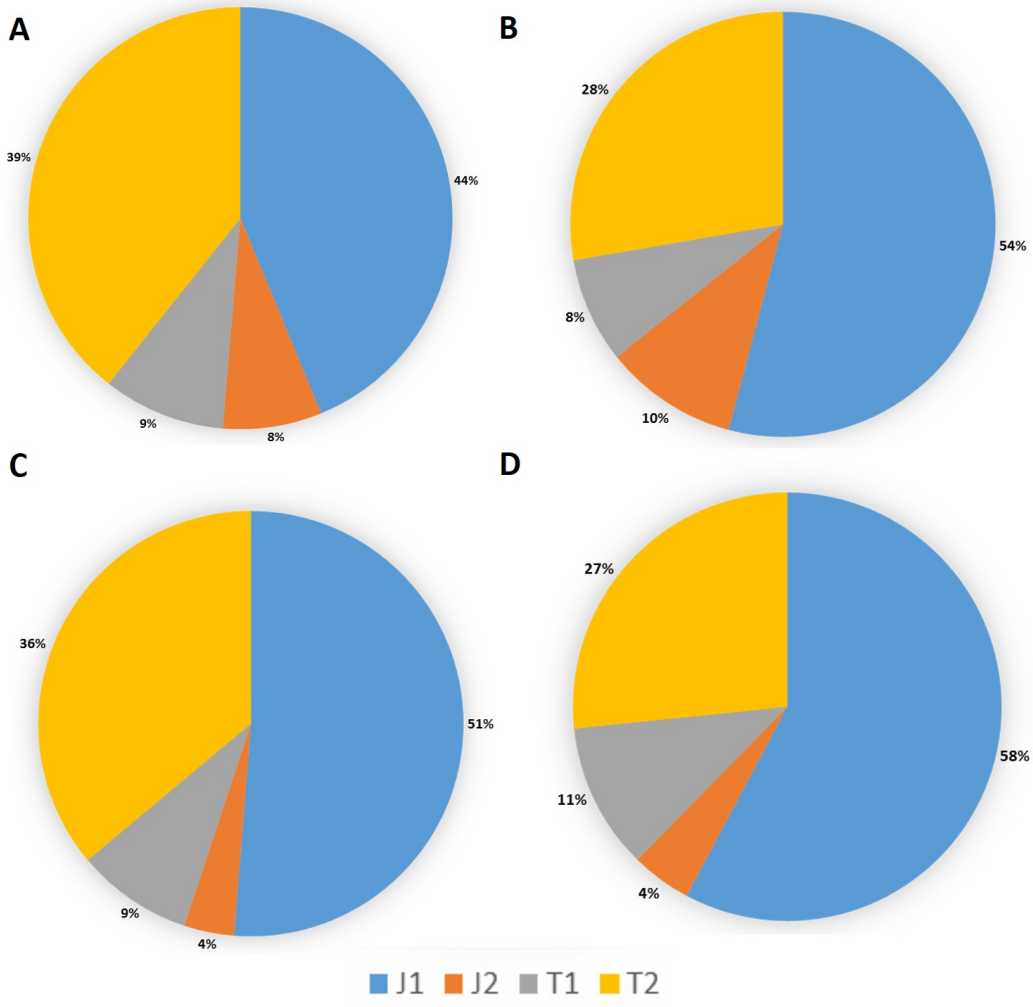

FIGURE B.3: Haplogroup JT composition of the Source BioScience SNP data and the Irish sequence data from Byrne et al. (2018) in the four countries of the British Isles. A – England; B – Scotland; C – Ireland; D – Wales.

K (with 45 haplotypes), J (with 36) and T (with 35)(Table B.4). The most frequent haplotypes among the English samples were: H1 (12.01%), U5 (9.95%), T2 (7.85%), K1 (7.33%), H3 (6.81%), J1 (5.76%), H5 (4.71%). The most frequent haplotypes among the Irish samples were: J1 (19.59%), H1 (14.43%), K1 (7.22%), T2 (6.19%), U5 (6.19%) and H3 (4.12%). The most frequent haplotypes among the Scottish samples were: U5 (10.88%), J1 (10.20%), T2 (7.48%), H1 (6.8%), H3 (5.44%), H2 (4.76%), K1 (4.76%), J2 (4.08%) and X2 (4.08%). Among the Welsh samples the most frequent haplotypes were: J1 (9.36%), H1 (8.94%), K1 (8.09%), H3 (7.66%), U5 (5.11%), T2 (4.68%), H6 (4.26%) and K2 (4.26%).

TABLE B.4: Haplogroup frequencies of the countries within the British Isles based on complete mitochondrial sequences. Frequencies are given in percent. Total percent calculated, excluding the 20 United Kingdom samples with unspecified geographic origin.

Country	mtDNA Haplogroups													
	H	J	U	K	T	I	V	X	W	HV	N	M	R	R0
England	43.46	9.42	14.14	9.42	9.95	2.09	3.66	2.09	3.14	1.57			0.52	0.52
Ireland	36.08	20.62	12.37	7.22	7.22	7.22	3.09	2.06	1.03	1.03	1.03	1.03		
Scotland	40.14	14.29	14.97	7.48	8.84	2.04	2.72	4.08	3.40	1.36	0.68			
Wales	43.40	11.06	9.36	12.34	6.81	6.38	2.55	2.55	2.13	2.98		0.43		
United Kingdom	35.00	5.00	5.00	30.00	20.00	5.00								
Total	40.77	13.84	12.71	9.11	8.20	4.43	3.00	2.69	2.42	1.53	0.42	0.36	0.13	0.13

B.3.2.2 Haplogroup composition of France

The French data consisted of 328 samples, belonging to 15 different haplogroups. Haplogroups with the highest frequencies were H (44%), U (15%), T (9%), J (8%), and K (8%) (Table B.5). The most diverse haplogroups were H (with 96 haplotypes), U (33), J (22), T (19) and K (17). Amongst these, the most frequent haplotypes were: H1 (15.85%), U5 (9.76%), T2 (8.23%), J1 (7.32%) and H3 (4.88%).

TABLE B.5: Haplogroup frequencies of the French dataset. Frequencies are given in percent.

Country	mtDNA Haplogroup														
	B	H	HV	I	J	K	L	M	N	R0	T	U	V	W	X
France	0.30	44.21	2.44	3.66	7.93	7.62	0.91	0.61	0.61	0.30	9.45	14.63	2.74	2.74	1.83

B.3.2.3 Haplogroup composition of Germany

The German subset consisted of 211 samples, belonging to 13 different haplogroups of which three showed the highest frequencies: H (45%), U (16%) and T (13%) (Table B.6). Most diverse haplogroups were H (with 70 haplotypes), U (28), T (14), J (13) and K (10). The most frequent haplotypes in this subset were H (16.11%), T2 (10.43%), U5 (8.53%), J1 (8.53%) and H5 (5.21%).

TABLE B.6: Haplogroup frequencies of the German dataset. Frequencies are given in percent.

Country	mtDNA Haplogroups												
	C	D	H	HV	I	J	K	N	T	U	V	W	X
Germany	0.47	0.47	45.50	1.90	2.37	9.48	5.69	0.47	12.80	15.64	1.42	2.84	0.95

B.3.3 Founder Analysis

Founder analyses were carried out with complete mitochondrial genomes to investigate the genetic input of haplogroup H3 into Britain through time.

B.3.3.1 Central Europe to Britain

The 200-year scan of founder-lineages dispersing into Britain identified two major coalescence peaks under the two employed criteria, $f1$ and $f2$, at 4 kya and 7.0-7.2 kya (Figure B.4).

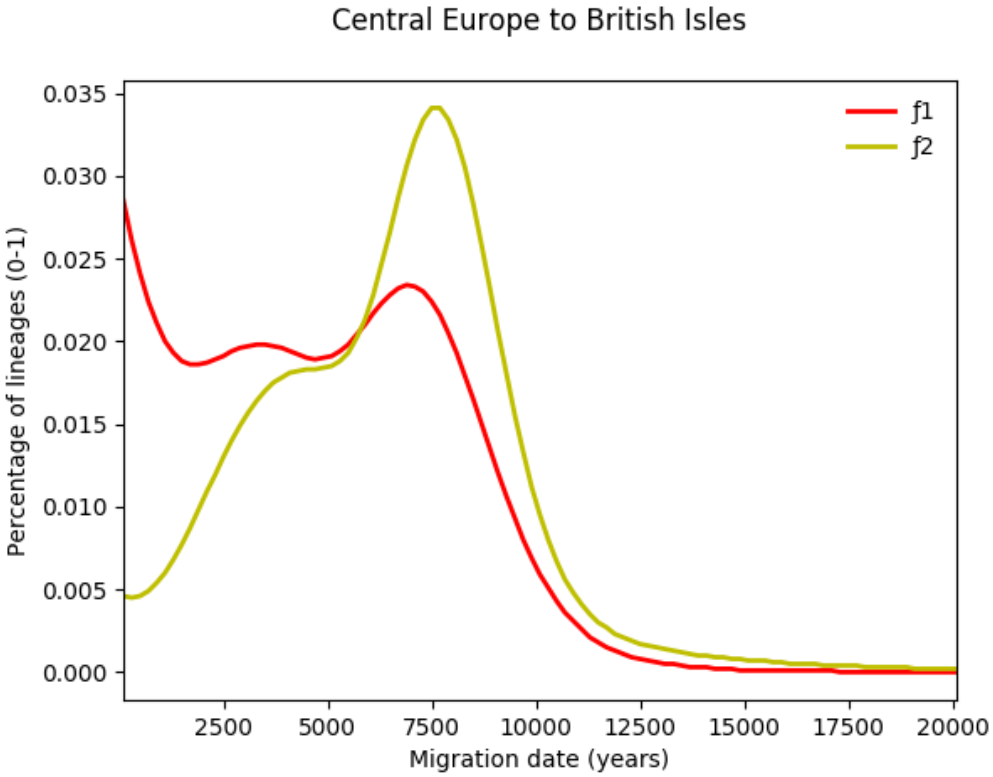

FIGURE B.4: Founder analysis results for Britain, assuming Central Europe as source for mtDNA variation. Probabilistic distribution of mtDNA founder clusters across migration times scanned at 200-year intervals from 0-20 kya, using two criteria for founder identification, $f1$ and $f2$.

The founder types in Britain were then partitioned using a migration model informed not only by the results of the scan, but also archaeological evidence, to quantify the contribution of the immigration events to the extant mtDNA gene pool in Britain. The model using Central Europe as source assumed migrations at 4 kya and 7.1 kya, corresponding to Neolithic/Bronze Age and pre-Neolithic migrations (Figure B.5). An additional dispersal at 0.1 kya was assumed to be due to historical/recent gene flow.

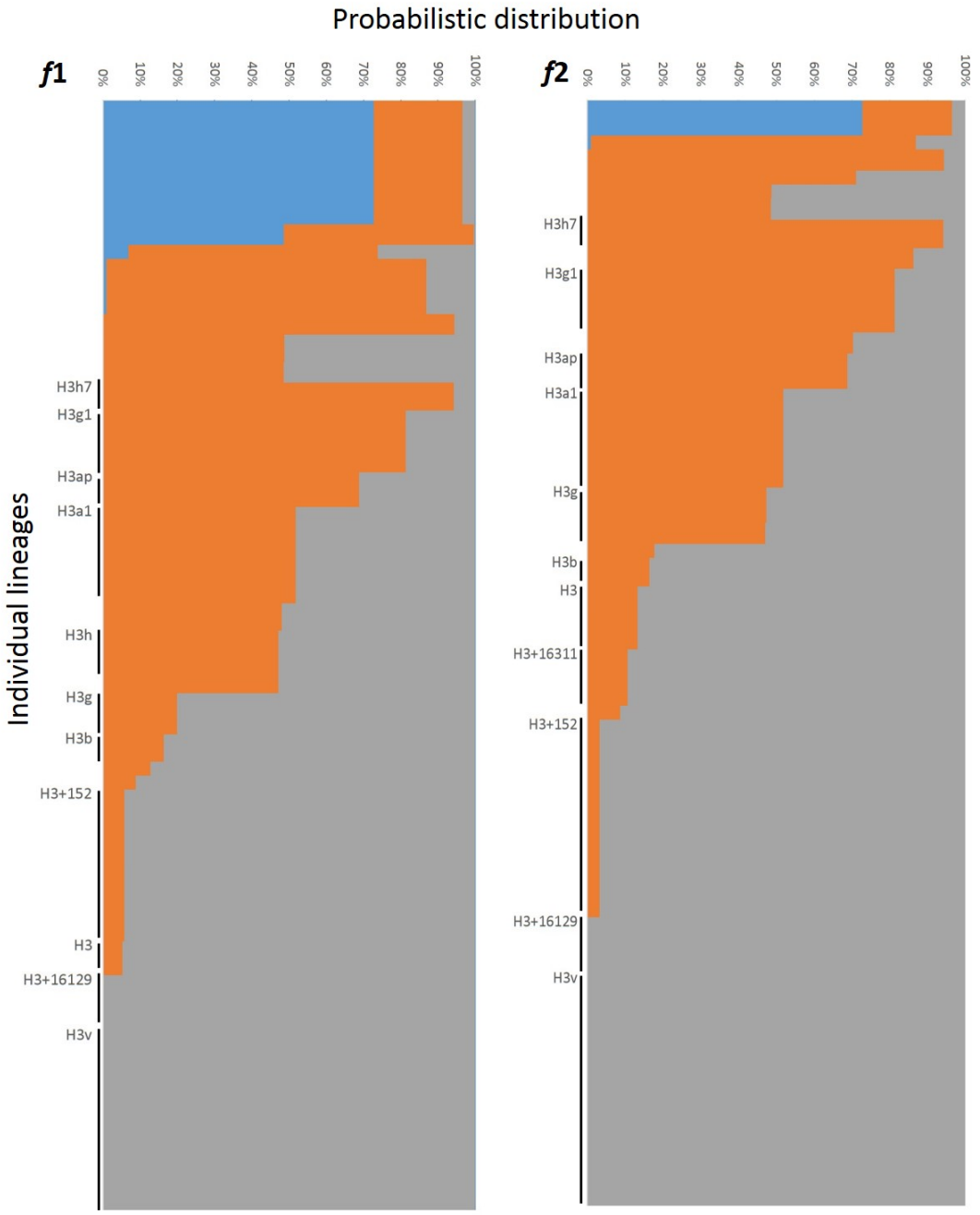

FIGURE B.5: Probabilistic distribution of each individual lineage in mtDNA variation in a three-migration model using two criteria, $f1$ and $f2$, for founder identification. Individual founder clusters with more than 2% frequency in overall Britain (sink population) are indicated at the left-hand side of each plot. Blue = migration at 0.1 kya; Orange = migration at 4 kya; Grey = migration at 7.1 kya.

B.3.3.2 Iberia to Britain

The 200-year scan of founder-lineages dispersing into Britain identified two major coalescence peaks under the two employed criteria, $f1$ and $f2$, at 3.8-4.1 kya and 7.0-7.2 kya (Figure B.6).

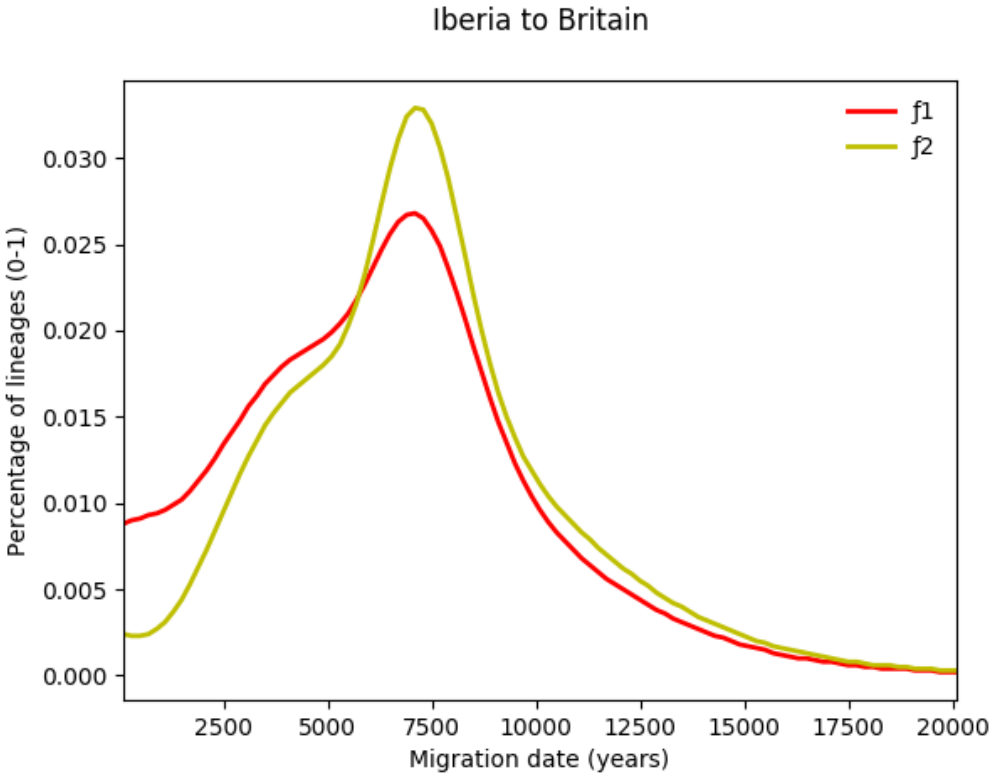

FIGURE B.6: Founder analysis results for Britain, assuming Iberia as source for mtDNA variation. Probabilistic distribution of mtDNA founder clusters across migration times scanned at 200-year intervals from 0-20 kya, using two criteria for founder identification, $f1$ and $f2$.

As before for central Europe, the founder types in Britain were partitioned using a migration model informed by archaeological evidence. The model using Iberia as the source assumed migrations at 4 kya and 7.1 kya, corresponding to Neolithic/Bronze Age and pre-Neolithic immigrations. An additional dispersal at 0.1 kya was assumed to account for any historical/recent gene flow (Figure B.7).

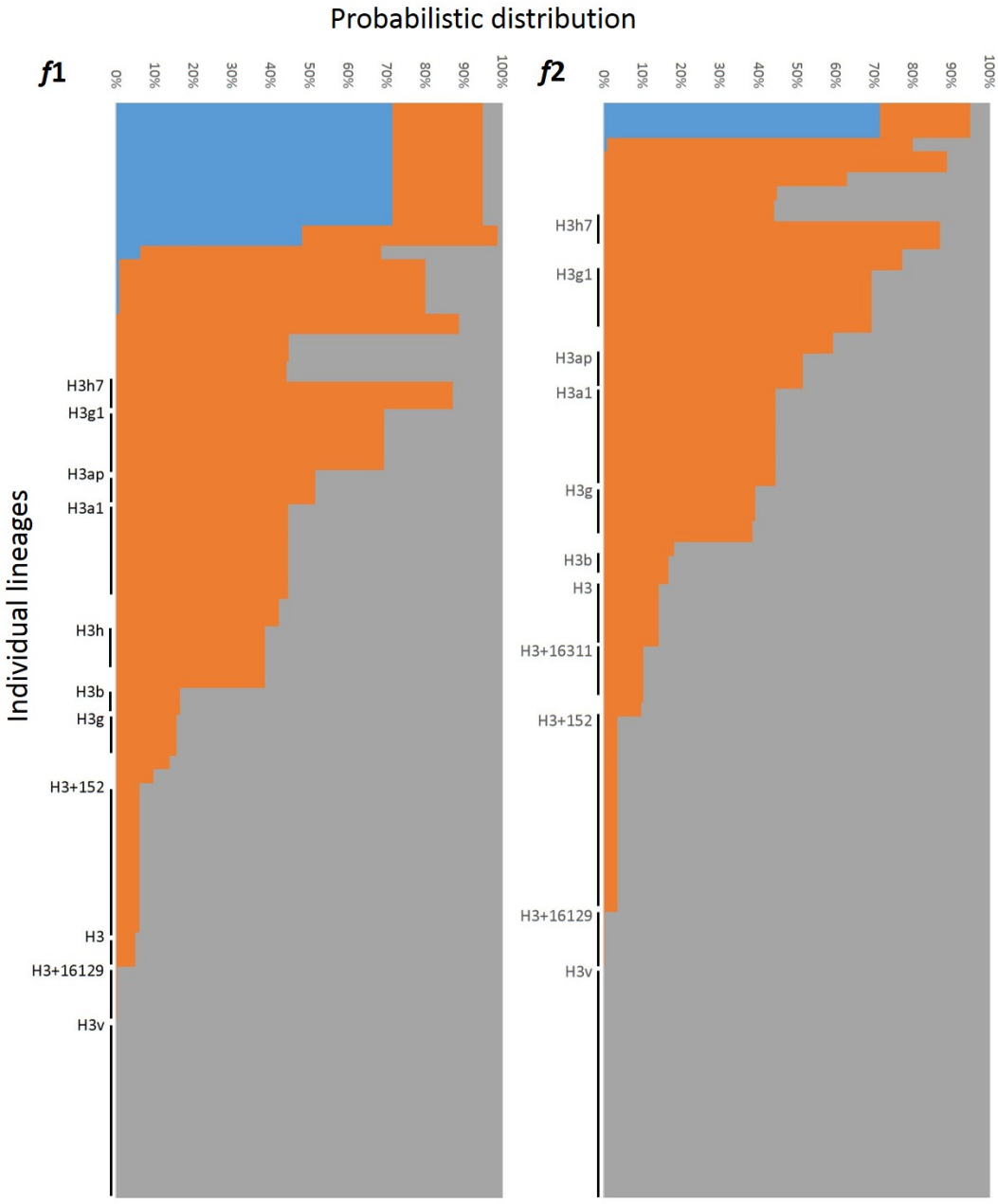

FIGURE B.7: Probabilistic distribution of each individual lineage in mtDNA variation in a three-migration model using two criteria, $f1$ and $f2$, for founder identification. Individual founder clusters with more than 2% frequency in overall Britain (sink population) are indicated at the left-hand side of each plot. Blue = migration at 0.1 kya; Orange = migration at 4 kya; Grey = migration at 7.1 kya.

B.4 Discussion

The haplogroup composition of H differs only slightly among the four countries in the modern SNP data of the British Isles (Figure B.1). H1 varied between 38% in England and 28% in Wales, while it comprised about 30% both in Scotland and Ireland. The largest percentage of H3 was found in Ireland. The higher frequency of H3 in Ireland compared to the rest of the British Isles might be due to its likely origin in Iberia and its northward spread along the Atlantic fringe. As discussed to some extent in Appendix C mtDNA haplogroup H3 has its frequency peak in Iberia and was found in Neolithic Portugal (Olalde et al., 2018).
Haplogroups H*, H1-H7 and HV were the majority haplogroups in all countries, whereas haplogroup H11 was absent in England, but contributed with 3-4% to the genetic landscape of Scotland, Ireland and Wales. Haplogroup H11 is mainly found in modern Denmark (21%), Britain and Ireland (95%) and Finland (8.2%), and has so far been found in Neolithic Lithuania (Mittnik et al., 2018), Early Bronze Age Germany (Brotherton et al., 2013), and Middle Bronze Age Hungary (Allentoft et al., 2015).
The country with the most subgroups contributing to the majority of the data was Scotland. Two of the haplogroups, H8 and H13 respectively, were present to an obviously larger degree in the Scottish population than in the other populations of the British Isles. Haplogroup H13 has been associated with the Yamnaya culture of the Caucasus (Haak et al., 2015).
North and South Wales (NW and SW respectively) differed in several aspects from each other (for an overview, see Figure L.5). The frequency of H1 differed about 10% between NW (6.1%) and SW (16.2%). The frequency of H3 was slightly higher in the North, but the frequencies did not differ significantly. Most of the minor H subhaplogroups were virtually absent in NW, but present in low frequencies in SW. J1 had its highest frequency in NW, with 20.3%, but was much lower in SW with 8.6%. Also the frequency of haplogroup V differed greatly between the North and South of Wales, with 7.1% and 1.4% respectively (Table B.3). Comparing the distribution of the most common haplogroups across the British Isles, it was obvious that H1 was the overall most frequent haplogroup. The average haplogroup frequency of H3 was slightly higher in England and Ireland than in Scotland and Wales, as were the average frequencies for HV, I3, I4 and X2. Instead, the average frequency of haplogroup H8 was higher in Scotland and Wales than in England, while it was absent in Ireland. A similar pattern was observable for

haplogroups I1, I2, J1, T2 and U5, except they also appeared in Ireland (Tables B.2 and B.3).

The comparison of the haplogroup frequencies of the SNP data with the haplogroup frequencies of the complete mitogenome data, showed that the overall frequency patterns do not differ very much. Haplogroups H (40.7%), J (13.8%), U (12.7%), K (9.1%), and T (8.2%) were the most frequent haplogroups, differing only slightly in the frequencies of the two datasets. The biggest differences were frequencies within Ireland, as H only accounted for ∼36% and J for ∼20%, rather than ∼40% and ∼11%, respectively.

Founder analysis was carried out to investigate the genetic input into Britain through time, with complete mitochondrial genomes. The 200-year scan of founder lineages dispersing into Britain identified two major coalescence peaks under the two employed criteria, *f1* and *f2*, for both assumed source populations, Central Europe (Figure B.4) and Iberia (Figure B.6), at 4 kya and 7.1 kya, and 4 kya and 7.1 kya, respectively. The coalescing at the time of the first settlements (∼8 kya) accounted for ∼43 to 59% of modern H3 lineages in Britain, assuming a central European source, and ∼46 to 59% assuming Iberia as the source (Figures B.5 and B.7).

Using central Europe as source, the oldest arrivals date to ∼8 kya, largely lineages basal to H3, lineages of H3v, H3+16129, and H3+152. The lineages with the biggest impact during the Neolithic/Bronze Age migrations were H3h7 and H3g1, as well as some minor clades with less than 2% frequency in the sink population, such as for example minor H3v subclades, H3b1b, H3y, H3aa, and H3z+152.

Using Iberia as source population, the oldest arrivals date to ∼7 kya, largely lineages basal to H3, H3v, H3+152, and H3+16129. Lineages with the biggest impact during the Neolithic/Bronze Age migrations were H3h7 and H3g1, as well as some minor subclades with less than 2% frequency in the sink population, such as for example minor lineages of H3v and H3z+152.

Generally, no major differences can be seen between the *f1* and *f2* criteria in both analyses. The likely overestimation of recent migrations of the *f1* criterion are expected, due to it being less conservative, as it allows only one derived branch.

When matching the identified founder types of haplogroup H3, with the biggest proportion of genetic input into Britain, to known ancient samples from the sink and source populations, only a few individuals carrying those haplogroups were found. The oldest individual belonging to basal H3 dates

to the Early Neolithic in Portugal (Olalde et al., 2018), other individuals belonging to this haplogroup were found in France and Scotland, however, this haplogroup was still most prevalent in Iberia during the Middle Neolithic. The presence of individuals along the western/northwestern extent of Europe, namely the Atlantic fringe, belonging to this lineage suggest a spread with the Atlantic Neolithic, across and into central Europe, supported by the finds from Germany (Brotherton et al., 2013). By the Bronze Age H3 spread west along the Mediterranean to Sardinia and into northeastern Europe (Brotherton et al., 2013; Günther et al., 2015; Mathieson et al., 2015; Olalde et al., 2015; Lipson et al., 2017; Olivieri et al., 2017; Olalde et al., 2018). To date the Scottish Neolithic sample belonging to basal H3, is the only ancient sample found to belong to one of the identified founder types. The founder analysis using central Europe as source differs quite significantly from published data. The source population for the Neolithic arrivals to Britain was mostly Iberia, while central European lineages mainly arrived with the Bell Beaker migrations into Britain (Olalde et al., 2018). The discrepancy between the here presented results and the published data is possibly due to low sample sizes in the source populations of central Europe (n=244) and the sink population of the British Isles (n=108). Founder analysis with a larger dataset would hopefully resolve this issue.

Appendix C

The European haplogroup H enigma

C.1 Introduction

During the Last Glacial Maximum (LGM), the majority of European populations either became extinct or retreated to southern refugia, and so a large fraction of the surviving lineages today derive from a founder effect or bottleneck associated with this event (Richards et al., 2000). Most European mitochondrial haplogroups have a Near Eastern origin and arrived in the continent during the Palaeolithic, or with the Neolithic transition and migrations from eastern populations (Torroni et al., 1998; Achilli et al., 2004; Sampietro et al., 2007; Brandt et al., 2013), and, for many haplogroups, multiple arrivals into Europe at different time periods are very likely (Achilli et al., 2004; Brotherton et al., 2013).

Among modern mitogenomes, a peculiar observation is the dominance of haplogroup H in present-day Europeans. The complex evolutionary history of this haplogroup makes it uncertain when and how it became the dominant European haplogroup. H is thought to have evolved in the Near East and Caucasus region and reached Europe during the Palaeolithic, although a considerable increase in frequency occurred during the Neolithic period (Achilli et al., 2004; Soares et al., 2010; Brotherton et al., 2013). Therefore, the evolution of H is still questionable. Haplogroup H has been argued to have originated in the Near East between 35,000 and 25,000 years ago (Torroni et al., 1998; Álvarez-Iglesias et al., 2009), which is a much older age estimate than the $\sim$18 kya reported by Soares et al. (2010). It clusters within the branch of macro-haplogroup R0, which is primarily Eurasian, with H contributing between 40 and 60% of all European haplogroups seen today (Achilli et al., 2004; Roostalu et al., 2007; Brotherton et al., 2013).

H has a seemingly uniform spread across Europeans, but it contains a complex substructure of phylogeographically informative subclades (Loogväli et al., 2004). It has an extreme star-like phylogeny, with the vast majority of subclades directly branching from the root, and a great majority of the branches indicated by only one defining mutation. The most common and widespread subhaplogroups are H1, H2 and H3, but H4 to H19 are also rather common. Conversely, the majority of the 102 remaining named subclades are minor, represented by only a few individuals. Within Britain the frequencies of most common subhaplogroups are different to the most common subhaplogroups within Europe, e.g. H5 has a higher frequency than H2, and H10 and H12 are very rare. The unique structure and distribution of H, and the peculiar frequency patterns of its subclades, has been widely discussed (Richards et al., 2000; Achilli et al., 2004; Loogväli et al., 2004; Roostalu et al., 2007; Torroni et al., 2006; Brotherton et al., 2013). The diversity seen in H is distinctive and makes it difficult to understand the origin of the different clades. The dominating star-like structure of H and its major subhaplogroups H1 and H3 have been thought to suggest dramatic late-glacial expansions (Torroni et al., 2006).

In Mesolithic hunter-gatherers from Central and Northern Europe, haplogroup H is virtually absent. This suggests that H was introduced from the Near East and Southeast Europe before the Linearbandkeramik (LBK) period, as H is seen after this time with an increasing frequency (Brotherton et al., 2013). In contrast, short fragment sequencing information suggested that haplogroup H was already present during the Mesolithic in the Iberian peninsula (Hervella et al., 2012). Complete genome sequences of ancient samples from the pre-Neolithic in Iberia could so far not support this.

Most of the subclades seen in the present-day Near East and Caucasus expanded into Europe after the LGM, but probably before the Holocene (Roostalu et al., 2007) and, during the Neolithic, there was a consistent and strong exponential growth, especially in Iberia from where it spread into Central and Northwestern Europe (Achilli et al., 2004; Brotherton et al., 2013). Neolithic processes, such as demographic changes during the Middle Neolithic period and the Bell Beaker phenomenon during the Bronze Age, are likely key factors for the massive expansion and increase of haplogroup H in Western Europe (Brotherton et al., 2013; Brandt et al., 2015). The subhaplogroup diversity of H that was established during the Middle Neolithic can still be observed in the present-day populations of Central Europe (Brotherton et al., 2013). The "Atlantic zone" expansion that followed the LGM, 15,000-10,000

years ago, might have carried H lineages into Central and Northern Europe (Torroni et al., 1998; Brotherton et al., 2013). This also supports the hypothesis that one of the earliest populations of the British Isles, which would have included haplogroup H, may have spread up the Atlantic littoral, with sailors reaching Ireland and Wales before spreading into Central Europe (Cunliffe, 2013).

The classification of haplogroup H lineages by first hypervariable segment (HVS-I) is hindered by mutations at fast-evolving nucleotide sites that frequently occur. Additionally, sequence information of just the HVS region leaves a substantial fraction of the haplogroup H phylogeny unresolved. The analysis of haplogroup H diversity and the phylogeography of European maternal lineages thus depends upon full mitochondrial genome information from representative population samples and geographic distribution (Loogväli et al., 2004). The analysis of distinct subhaplogroups of H should reveal spatial frequency patterns and migration events (Achilli et al., 2004), such as the origin and spread of haplogroup H3, which is associated with the spread of the Neolithic and the Bell Beaker culture from Iberia into Central and Northern Europe (Álvarez-Iglesias et al., 2009). Furthermore, haplogroup H7 has no HVS motif and has hitherto been completely invisible in previous studies not using complete mitochondrial genomes.

In the chapter, subhaplogroups H3, H6, H7, H8 and H31 have been analysed in parallel for their spatial frequency patterns. H3, along with H1, are the most common subclades in Western Europe, with the highest frequencies seen in Iberia. Between them, they are the most common subhaplogroups within H in Europe (Brotherton et al., 2013). They are found in similar geographic areas, but H3 represents a much smaller fraction of H than H1 (Achilli et al., 2004). Subhaplogroups H3, H6 and H8 show distinct phylogeographic patterns (Loogväli et al., 2004). Together with H7, H3 and H6 cover the majority of Finnish, US/UK and Galician haplogroup H sequences (Herrnstadt et al., 2002; Loogväli et al., 2004; Quintáns et al., 2004). The overall frequency of these three haplogroups in Britain is 9.14% (Dulias, 2019). Furthermore, H6 shares a state at position 16362 with the clade of H8 (Loogväli et al., 2004) and H31, while the clade of H8 and H31 is defined by a mutation at position 146. Therefore, subhaplogroups H3, H6, H7, H8 and H31 have been analysed in parallel for their spatial frequency patterns in this chapter.

C.2 Materials and Methods

C.2.1 Materials

Along with newly generated data, all available published data of complete modern and ancient mitochondrial genomes was used to reconstruct phylogenetic trees of haplogroups H3, H6, H7, H8 and H31. Samples with unknown geographic origin, as well as unspecified "European" samples and those from outside Europe that have European/Caucasian ancestry (e.g. USA), were included in the phylogenetic analysis, as more samples allow for a more precise phylogenetic tree reconstruction. These were removed from the detailed phylogeographic analysis, as they do not contribute any geographic information.

C.2.2 Phylogenetic analysis

Most commonly, discrete-character methods, such as parsimony and maximum likelihood, are employed to infer evolutionary trees (Salemi et al., 2009). These approachs are based on the concept of using the minimum number of mutations to account for the branching within the tree. The phylogenies of complete European mitochondrial lineages were reconstructed using mtPhyl V4.015 (Eltsov et al., 2011) (`https://sites.google.com/site/mtphyl/home`), applying the discrete-character method maximum parsimony. Recurrent mutations can cause mtPhyl to provide a putative tree, which might differ considerably from the most parsimonious tree, especially when dealing with a large dataset. Therefore, all phylogenetic trees reconstructed with mtPhyl were manually reviewed based on mtDNA PhyloTree build 17 (`http://www.phylotree.org/tree/index.htm`). Heteroplasmies and "hotspot" mutations were not considered for the phylogenetic trees (see subsection B.2.3.6, exceptions are position 152 and position 16311), and the molecular clock developed by Soares et al. (2009) was used for the mutation rate. The reviewed parsimonious phylogenetic trees were then converted into xml trees as input files for Founder Analysis used in Chapter B.

C.2.3 Phylogenetic age estimation

The phylogenetic framework assumes that the evolution of observed sequences follows a tree structure (Salemi et al., 2009). Maximum likelihood infers the probability of point mutations along a branch of a phylogenetic tree, thereby

determining the tree topology, branch lengths and parameters of the evolutionary model (Salemi et al., 2009). The average number of mutations between a cluster of sequences and their most recent common ancestor within a parsimonious phylogenetic reconstruction can be calculated as a rho statistic (ρ) (Forster et al., 1996). Originally, this formula assumed that all mutations were independent, which is a very rare occurrence, and therefore the formula was revised to take both the tree shape and the type of counted mutations into account for the variance calculation (Saillard et al., 2000). Rho (ρ) and sigma (σ^2) were obtained from the coalescent time estimates exported from the Founder Analysis software, which furthermore outputs the age estimates using Soares' clock, which accounts for purifying selection (Soares et al., 2009).

C.3 Results and Discussion

Haplogroup H has been proposed to have originated in the Near East ∼18 kya (Soares et al., 2010), and then expanded with a Paleolithic wave into Europe and was further involved in the repopulation of Central and Northern Europe from ice-age refugia after the LGM (Achilli et al., 2004). As the complete molecular dissection of subhaplogroups of haplogroup H and their spatial distribution and founding events is still under analysis, only the subhaplogroups that I analysed (H3, H6, H7, H8 and H31)are presented here.

C.3.1 Phylogeny of H3

For the phylogenetic reconstruction of mitochondrial haplogroup H3, 1,292 published and unpublished sequences were used. The references and numbers of published sequences are given in Table C.1. 29 published ancient sequences were also included in the analysis, and these derived from the Czech Republic (n=2), Denmark (n=2), France (n=1), Germany (n=9), Portugal (n=1), Russia (n=2), Scotland (n=3), Spain (n=10) and Switzerland (n=1) (Brotherton et al., 2013; Allentoft et al., 2015; Günther et al., 2015; Haak et al., 2015; Mathieson et al., 2015; Lipson et al., 2017; Olalde et al., 2018). 275 newly generated sequences were included; 92 from Germany, 65 from Spain, 57 from the British Isles and Ireland, 23 from Italy, 16 from the Netherlands, 14 from France, six from Belgium, one from Cyprus and one from Luxembourg.

TABLE C.1: All published samples used for the phylogenetic reconstruction of haplogroup H3. Bold numbers indicate ancient samples.

Country	Sample Number	Reference
Italy	438	1000 Genomes Project Consortium (2010), Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), Achilli et al. (2004), Fraumene et al. (2006), Gasparre et al. (2007), Behar et al. (2012a), Lippold et al. (2014), Bodner et al. (2015), and Olivieri et al. (2017)
	1	Olivieri et al. (2017)
Basque Country	82	Behar et al. (2012a), Cardoso et al. (2013), Lippold et al. (2014), and Batini et al. (2017)
Denmark	75	Li et al. (2014)
	2	Allentoft et al. (2015)
British Isles & Ireland	51	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), 1000 Genomes Project Consortium (2010), Behar et al. (2012a), and Batini et al. (2017)
	3	Olalde et al. (2018)
Caucasian USA	33	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), 1000 Genomes Project Consortium (2010), Just et al. (2014), and Sequeira et al. (2015)
Finland	27	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), Finnilä et al. (2001), 1000 Genomes Project Consortium (2010), Behar et al. (2012a), Soini et al. (2012), Soini et al. (2013), and Raule et al. (2014)

Continued on next page

Table C.1 – continued from previous page

Country	Sample Number	Reference
Spain	19	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), Achilli et al. (2004), Fregel et al. (2009), 1000 Genomes Project Consortium (2010), Gómez-Carballa et al. (2011), and Batini et al. (2017)
	10	Günther et al. (2015), Mathieson et al. (2015), Lipson et al. (2017), and Olalde et al. (2018)
Germany	14	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*) and Behar et al. (2012a)
	9	Brotherton et al. (2013), Haak et al. (2015), and Mathieson et al. (2015)
Undefined European	14	Coble et al. (2004) and Pope et al. (2011)
France	10	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), Behar et al. (2012a), and Lippold et al. (2014)
	1	Olalde et al. (2018)
Netherlands	6	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*) and Behar et al. (2012a)
Russia	6	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), Behar et al. (2012a), Derenko et al. (2014), and Lippold et al. (2014)
	2	Mathieson et al. (2015)
South Tyrol	4	Pichler et al. (2010)
Poland	4	Behar et al. (2012a) and Skonieczna et al. (2015)
Tunisia	3	Costa et al. (2009)

Continued on next page

Table C.1 – continued from previous page

Country	Sample Number	Reference
Georgia	2	Achilli et al. (2004)
India	2	Behar et al. (2012a) and Palanichamy et al. (2015)
Iran	2	Derenko et al. (2013)
Portugal	2	Behar et al. (2012a)
	1	Olalde et al. (2018)
Sweden	2	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*) and Behar et al. (2012a)
Armenia	1	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*)
Austria	1	Behar et al. (2012a)
Belgium	1	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*)
Bulgaria	1	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*)
Croatia	1	Fu et al. (2012)
Czech Republic	1	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*)
	2	Olalde et al. (2018)
Faroe Islands	1	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*)
Japan	1	Imanishi et al. (2011)
Lithuania	1	Behar et al. (2012a)

Continued on next page

Table C.1 – continued from previous page

Country	Sample Number	Reference
Malta	1	Behar et al. (2012a)
Norway	1	Behar et al. (2012a)
South Africa	1	Van Der Walt et al. (2012)
Switzerland	**1**	Olalde et al. (2018)
Unknown	187	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), Mishmar et al. (2003), Kloss-Brandstätter et al. (2010), Zaragoza et al. (2010), Bertolin et al. (2011), Fendt et al. (2011), Gómez-Carballa et al. (2011), Seoane et al. (2011), Behar et al. (2012a), Pacheu-Grau et al. (2012), Gianoarli et al. (2014), Li et al. (2014), and Kloss-Brandstätter et al. (2015)

H3 is one of the major subclades of haplogroup H, and is widespread and common among Europeans. Similarly to haplogroup H as a whole, H3 also has a star-like phylogeny, although with fewer nested clades. In general, most subclades of H3 are defined by one single mutation and branch directly from the root, which is similar to the overall structure of haplogroup H. H3 accounts for up to 18% of haplogroup H. It has a ρ age estimate of $\sim$9.4 thousand years ago (kya), which corresponds to the late Mesolithic/Early Neolithic (Table C.2). It is defined by position 6776 and is divided into five major and several minor subclades. The major subclades are H3b ($\sim$9.5 kya) accounting for $\sim$6% of H3, H3c (9 kya) accounting for $\sim$5%, a 99% Sardinian clade defined by position 93 (6.6 kya) accounting for $\sim$17%, a clade defined by position 16311 (7.6 kya) accounting for $\sim$8%, and a cluster defined by position 152 ($\sim$8.2 kya) accounting for $\sim$20% of H3 that includes the rather small branch of H3a ($\sim$8.3 kya). The age estimates (ρ) of the different branches within H3 are controversial; for example, H3b is older than H3, but this is

due to the very large confidence intervals around the dates. The ρ estimate of the Sardinian specific clade that accounts for ∼17% of the reconstructed H3 in this analysis also is problematic, as its estimated age is much younger compared to the Sardinian specific subclades of H3 reported in Olivieri et al. (2017), although the confidence interval covers the reported age of 9-10.5 kya. These highlight that ρ calculations are rather imprecise compared to maximum likelihood (ML) age estimates. Calculating ML age estimates would support the discussion and give more precise dates of the branches.

TABLE C.2: Rho age estimates and 95% confidence intervals of main nodes of the H3 phylogenetic tree.

Node	Rho	Standard error	Age	95% confidence interval
> H3	3.983	0.546	10,564	7656 - 13,521
>> H3+152	3.122	0.525	8222	5464 - 11,026
>>> H3+152+16311	1.600	0.693	4159	622 - 7778
>>>> H3j	1.000	0.707	2585	0 - 6244
>>> H3a	3.167	1.163	8342	2293 - 14,614
>>>> H3a1	2.167	0.593	5659	2596 - 8782
>>>>> H3a1a	1.056	0.455	2730	422 - 5075
>>> H3g	2.901	0.966	7626	2606 - 12,801
>>>> H3g1	1.404	0.276	3644	2230 - 5070
>>>>> H3g1a	1.333	0.667	3457	68 - 6924
>>>>> H3g1b	0.600	0.447	1546	0 - 3834
>>>> H3g2	1.00	0.745	2585	0 - 6444
>>>> H3g3	1.857	0.655	4838	1478 - 8270
>>>> H3g4	2.500	0.791	6549	2456 - 10,746
>>> H3i	1.333	0.648	3457	163 - 6825

Continued on next page

Table C.2 – continued from previous page

Node	Rho	Standard error	Age	95% confidence interval
>>>> H3i1	0.800	0.49	2065	0 - 4581
>>> H3k	2.481	0.686	6499	2943 - 10,134
>>>> H3k1	2.231	0.9	5830	1199 - 10,598
>>>>> H3k1a	1.100	0.539	2846	114 - 5629
>> H3+16189	2.857	0.728	7507	3710 - 11,392
>> H3b	3.566	1.022	9426	4064 - 14,958
>>>> H3b1	3.450	1.028	9110	3723 - 14,670
>>>>> H3b1a	1.429	0.535	3707	979 - 6485
>>>>> H3b1b	3.000	1.063	7892	2367 - 13,603
>>>>>> H3b1b1	1.889	0.556	4922	2065 - 7832
>>>>>>> H3b1b1a	0.500	0.500	1287	0 - 3843
>>>> H3b2	0.333	0.333	857	0 - 2551
>>>> H3b3	2.000	0.938	5216	414 - 10,169
>>>> H3b4	1.667	1.000	4334	0 - 9590
>>>>> H3b4a	0.500	0.500	1287	0 - 3843
>>>> H3b5	2.000	1.414	5216	0 - 12,737
>>>> H3b6	2.125	0.857	5548	1146 - 10,076
>>>>> H3b6a	0.333	0.333	857	0 - 2551
>>>> H3b7	0.429	0.247	1102	0 - 2360
>> H3c	3.397	0.874	8965	4378 - 13,678
>>> H3c1	3.500	1.287	9246	2527 - 16,235
>>> H3c2	2.030	0.730	5297	1542 - 9142

Continued on next page

Table C.2 – continued from previous page

Node	Rho	Standard error	Age	95% confidence interval
>>>> H3c2a	0.864	0.284	2230	789 - 3685
>>>>> H3c2a1	1.000	0.577	2585	0 - 5566
>>>> H3c2b	2.000	0.791	5216	1159 - 9380
>>>>> H3c2b1	2.000	1.000	5216	103 - 10,501
>>>> H3c2c	0.250	0.250	642	0 - 1909
>>> H3c3	1.500	0.866	3895	0 - 8425
>> H3+73	3.250	0.642	8567	5193 - 12,009
>>>> H3ak	3.800	1.183	10,063	3845 - 16,508
>>> H3d	0.667	0.471	1719	0 - 4134
>> H3e	0.500	0.289	1287	0 - 2757
>> H3+93	2.548	0.827	6677	2393 - 11,076
>>> H3+93+15315	1.603	0.501	4165	1600 - 6774
>>>> H3+93+15315+73	1.300	0.700	3370	0 - 7009
>> H3t	0.857	0.350	2213	440 - 4008
>> H3r	2.500	0.791	6549	2456 - 10,746
>>> H3r1	0.500	0.500	1287	0 - 3843
>> H3q	2.444	0.831	6400	2103 - 10,813
>>> H3q1	1.500	0.935	3895	0 - 8793
>> H3v	3.363	0.322	8874	7170 - 10,594
>>> H3v1	0.333	0.333	857	0 - 2551
>>> H3v2	1.500	1.118	3895	0 - 9767
>> H3x	5.250	2.035	14,067	3270 - 25,509

Continued on next page

Table C.2 – continued from previous page

Node	Rho	Standard error	Age	95% confidence interval
>>> H3x1	1.857	1.182	4838	0 - 11,089
>> H3as	0.667	0.471	1719	0 - 4134
>> H3aa	2.200	0.721	5748	2030 - 9554
>> H3ab	2.000	0.639	5216	1929 - 8573
>> H3y	0.667	0.333	1719	34 - 3422
>> H3ai	0.667	0.471	1719	0 - 4134
>> H3an	3.667	1.453	9699	2114 - 17,628
>> H3+16311	2.900	0.858	7622	3158 - 12,208
>>> H3h	1.937	0.380	5049	3089 - 7032
>>>> H3h1	0.647	0.428	1668	0 - 3860
>>>> H3h2	1.375	0.718	3567	0 - 7306
>>>>> H3h2a	1.000	0.707	2585	0 - 6244
>>>> H3h3	2.000	0.721	5216	1511 - 9010
>>>>> H3h3a	0.600	0.346	1546	0 - 3315
>>>>> H3h3b	0.667	0.471	1719	0 - 4134
>>>> H3h4	2.333	1.202	6103	0 - 12,507
>>>> H3h5	0.429	0.247	1102	0 - 2360
>>>> H3h6	1.400	0.600	3632	575 - 6752
>>>> H3h7	1.077	0.288	2786	1320 - 4267
>>> H3n	1.200	0.490	3108	616 - 5642
>>> H3m	1.286	0.553	3333	517 - 6202
>> H3p	3.500	1.803	9246	0 - 19,108

Continued on next page

Table C.2 – continued from previous page

Node	Rho	Standard error	Age	95% confidence interval
>> H3+6953	3.857	1.623	10,219	1745 - 19,119
>>> H3s	2.500	1.258	6549	86 - 13,277
>> H3u	3.136	1.125	8260	2404 - 14,324
>>> H3u1	0.667	0.471	1719	0 - 4134
>> H3+16248	3.143	1.37	8277	1176 - 15,688
>>> H3w	1.333	0.707	3457	0 - 7136
>> H3z	3.133	0.819	8252	3968 - 12,646
>>> H3z1	2.667	0.816	6995	2759 - 11,343
>>> H3z2	0.500	0.500	1287	0 - 3843
>> H3ad	1	0.707	2585	0 - 6244
>> H3ac	0.333	0.333	857	0 - 2551
>> H3ae	1.333	0.667	3457	68 - 6924
>> H3af	2.214	1.062	5786	341 - 11,422
>> H3ag	2.333	0.707	6103	2448 - 9843
>>> H3ag1	1.000	0.707	2585	0 - 6244
>> H3ah	1.125	0.415	2912	803 - 5051
>> H3aj	0.500	0.500	1287	0 - 3843
>> H3am	1.833	0.833	4775	513 - 9155
>> H3ao	2.188	0.640	5715	2409 - 9090
>>> H3ao1	2.167	1.143	5659	0 - 11,726
>> H3ap	1.714	0.391	4460	2448 - 6498
>> H3ar	1.600	0.632	4159	926 - 7460

Continued on next page

Table C.2 – continued from previous page

Node	Rho	Standard error	Age	95% confidence interval
>> H3aq	0.500	0.500	1287	0 - 3843
>> H3at	2.125	1.023	5548	307 - 10,968
>>> H3at1	1.286	0.589	3333	337 - 6389
>> H3au	1.750	0.901	4554	0 - 9291
>> H3+7849	2.625	1.281	6883	294 - 13,748
>>> H3av	0.571	0.286	1472	29 - 2929

Less than 1% of the samples within H3 are from outside Europe (including geographically European parts of Russia). Those samples are from the Caucasus, India, Iran, Tunisia and Japan. Modern samples at the root of H3 are found mainly in Italy (including Sardinia), Spain (including the Basque country), plus single instances in Denmark and England. Fifteen of the 29 ancient H3 sequences are basal to H3; Late Neolithic Germany and France (n=3) (Brotherton et al., 2013; Haak et al., 2015; Olalde et al., 2018), Neolithic Portugal and Scotland (n=2) (Olalde et al., 2015; Olalde et al., 2018), Early Bronze Age Germany (n=1) (Brotherton et al., 2013), and Chalcolithic Spain and Czech Republic (n=9) (Günther et al., 2015; Mathieson et al., 2015; Lipson et al., 2017; Olalde et al., 2018). This fits largely with the argument of Iberia being the source of H3 (Achilli et al., 2004; Torroni et al., 2006; Roostalu et al., 2007; Soares et al., 2010), as H3, together with H1, has a frequency peak in Iberia. The Franco-Cantabrian refugal area in Iberia appears to be the source of late-glacial expansions of hunter-gatherers who repopulated Central and Northern Europe after the LGM (Achilli et al., 2004). However, recent ancient DNA analysis of sequences from Iberia showed that most of Iberian mitochondrial haplogroups in that period were U5b (Sánchez-Quinto et al., 2012). Based on the sequences found at the root of H3, an Iberian origin is a possibility, but an origin in Italy cannot be excluded. The high frequency of H3 in Sardinia and the presence of several Sardinian specific lineages within H3

(Olivieri et al., 2017), as well as the large Sardinian cluster seen in this analysis and the presence of basal H3 sequences originating in Sardinia, all support Sardinia as a putative origin of haplogroup H3. Recent ancient DNA analyses of Bronze Age Iberia reported several basal H3 sequences from Chalcolithic Iberia (Günther et al., 2015; Mathieson et al., 2015; Lipson et al., 2017; Olalde et al., 2018), but one individual from the Nuragic period (3008 ± 25, MAMS-26895) in Sardinia has been reported as basal H3 (Olivieri et al., 2017).

Haplogroup H3 is more frequent in the West (80% of samples) than the East (remaining 20%) of Europe and almost absent in the Near East and Anatolia as already reported (Loogväli et al. (2004)) and as described above. Olivieri et al. (2017) suggested that haplogroup H3 may have arrived prior to the Neolithic in Sardinia, either from Spain or elsewhere in the Western Mediterranean. In order to identify the likely source population of H3, Founder Analysis was undertaken as described in chapter B. The analysis is currently on-going and only preliminary results are presented in chapter B with a focus on the British Isles. The overall analysis of haplogroup H is very complex and does not fit within the major aim of this thesis.

H3a, H3g, H3i, H3j and H3k all cluster within the branch defined by position 152. Two modern German samples are basal to 152, as well as one Middle Neolithic German sample from Salzmünde and one sample from the Unetice culture in Germany (Brotherton et al., 2013). This suggests that H3 already arrived in Central Europe by the time of the Middle Neolithic ($\sim$6 kya). Within H3a (8.3 kya), all samples cluster inside H3a1 (5.7 kya), which has a basal Danish sample and is 25% Scottish, with a further 14% from Ireland and Britain, H3a1a is exclusively British. It is otherwise found in Italy, the Netherlands and Poland. This distribution and the origin of the basal samples supports a spread of H3a from Germany or Central Europe mainly towards the Northwest and, with lower frequencies, towards the South and East.

H3g is the biggest clade of this branch. It has an age estimate of 7.5 kya, two Late Bronze Age Srubnaya samples from Russia(Mathieson et al., 2015) at the base, as well as a sample from Armenia and two samples from Denmark. The Srubnaya culture has been associated with Indo-Aryan dispersals from Central Asia into South Asia (India) and West Eurasia (Silva et al., 2017). The modern Indian sample which is basal to H3g is potentially a direct connection to the Srubnaya dispersals into South Asia. H3g1 (3.6 kya) has six basal Danish samples and is overall $\sim$36% Danish. H3g1b (1.5 kya) is entirely British (mainly found in Scotland), and newly named H3g1c,

defined by positions 13711 and 16399 (1.5 kya), is only found in Wales and Shetland. Other samples within H3g1 are mainly from Northwestern Europe with basal Danish samples and some sharing of lineages from Italy and Spain with the North. H3g2 (2.6 kya) is completely Irish and British. H3g3 (4.8 kya) and H3g4 (6.5 kya) on the other hand are found in Britain, France, the Netherlands and Russia, and in Belgium, Germany and Ireland, respectively. Samples directly branching off H3k (6.5 kya) are mainly from Italy and Spain (apart from one German sample). H3k1 (5.8 kya) has Spanish samples branching off, while H3k1a (2.8 kya) is only found in Britain. The British samples being nested within the Spanish ones, show a direct Atlantic connection, which is much more recent than the Early Neolithic, indicating that links along the Atlantic fringe continue throughout prehistory, although the age estimates are very rough. Further clusters branching off H3k are Spanish and Basque, German and Spanish. Other minor branches within 152 are Basque, Belgian, Danish, Dutch, German, Italian, Irish, Sardinian and Spanish.

Two Late Neolithic samples from Denmark and Germany (Allentoft et al., 2015; Haak et al., 2015) and one Early Bronze Age sample from Germany (Brotherton et al., 2013) can be found basal to H3b (9.5 kya). Again, this suggests that H3 was already well established and widespread during the Late Neolithic in Central Europe. Samples from England, Italy and Sardinia are basal lineages to H3b, as well as a small clade defined by position 11014, which is found in Britain, Ireland, France and Sardinia. All subclades of H3b belong to a cluster defined by position 16129. Basal to position 16129 are French, Danish and Italian samples. Directly branching from 16129 are samples from Denmark, England, Finland, Germany, Italy, South Tyrol and Spain. H3b1 (9.1 kya) is divided into H3b1a (3.7 kya) and H3b1b (7.9 kya). H3b1a is found in Belgium, Denmark, Ireland and Wales, and is similar to H3b1b, which comprises of Northwestern European samples from Denmark, France, Scotland and Wales, the latter two of which fall within H3b1b1 (4.9 kya). The overall distribution of H3b, and the partly very specific geographic distribution among Northwestern/Northern European countries, suggests a rapid movement throughout Europe and local evolution into the different clades of H3b, with Northeastern samples appearing to have arrived through Northwestern and Central Europe.

H3c (9 kya) contains mainly Spanish samples (including the Basque country) of which one dates to the Chalcolithic and a few lineages in Italy and

Northwestern Europe. two basal samples from Denmark and one basal ancient Chalcolithic sample from Spain (Günther et al., 2015). H3c1 (9.2 kya) includes one sample from Italy, while more recently diverged samples are from Britain and Germany. H3c2 (5.3 kya) is the biggest subclade and can be divided into H3c2a (2.2 kya), H3c2b (5.2 kya) and H3c2c (600 a). H3c2a is mainly found in the Basque country, with some exceptions found in Spain. H3c2b is only found in Britain and H3c2c is completely Basque. The distribution of these clusters suggest a migration $\sim$5.2 kya from Iberia along the Atlantic fringe towards the British Isles. One further cluster within H3c2 dates to 7.9 kya and contains both a German and a Spanish branch. H3c3 contains a basal ancient Spanish sample from the Chalcolithic (Mathieson et al., 2015) and a modern German sample, supporting a migration from the Southwest to Central and Northwestern Europe. Other clades or samples basal to H3c are from the Basque country, Denmark, France, Germany, Italy, the Netherlands, Portugal and Spain.

The cluster defined by a transition at position 16311 (dating to 7.6 kya, if a genuine clade) is divided into H3h (5 kya), H3m (3.3 kya) and H3n (3.1 kya). Basal to this clade are two Scottish samples from the Middle and Early Bronze Age (Olalde et al., 2018), suggesting that the origin of this clade is in Scotland, and that the initial development of the clade may have been in this geographic region. Modern samples directly branching off the root of 16311 are from Denmark, Finland, Italy and Scotland. H3h contains basal samples from Belgium and Italy. Samples directly branching off H3h are from Belgium, Britain, Denmark, France, Germany and Italy. H3h1 (1.7 kya) is a mainly Northeastern cluster with the majority of samples being from Finland, with the exception of one Russian and two Danish samples. H3h2 (3.6 kya) contains samples from Germany and Wales and H3h2a contains samples from Iran and Italy. This suggests a recent migration from Northwestern/Central Europe to the Mediterranean and Near East. Although the sample size in this case is fairly small it allows to distinguish back-migration from the North to the South from lineages that represent the initial source in the South. H3h3 (5.2 kya) is divided in H3h3a, H3h3b and an unnamed cluster. H3h3a (1.5 kya) is an Italian clade, H3h3b (1.7 kya) is found in Denmark, Germany, and the Netherlands, and the third clade (1.3 kya) is found in Finland and Sweden. H3h4 (6.1 kya) is found in Germany. H3h5 (1.1 kya) is found in Finland, Germany and Italy. H3h6 (3.6 kya) is found in the Basque country, Germany and Spain. H3h7 (2.8 kya) contains basal samples from Denmark and Germany. H3h7 is an extremely star-like subclade of H3h, as

all samples directly branch from H3h7 without any clustering. These samples are from Britain, Denmark, Germany and the Netherlands. H3m is mainly found in Ireland, whereas H3n is found in the Basque country and Spain.
The clade with the most samples belonging to one cluster and geographic region is the Sardinian clade (n=218) defined by position 93. Although this subclade dates to 6.6 kya, it could potentially have arrived with the Early Neolithic, dated archaeologically to ∼7.8 kya, followed by heavy drift in the small isolated population, since the branch emerges from the root of H3. This is supported by the unusually strong sharing of haplotypes within the subclade (and others in Sardinia). This drift, making the subclade extremely non-starlike, also renders the age estimate especially imprecise, and consistent with an Early Neolithic origin. The effect is even more pronounced in some of the smaller Sardinian-specific subclades of H3. There are several basal lineages in Spain (and one in Ireland) possibly pointing to an Iberian source, and one Italian tip lineage indicating recent migration to the mainland. This phenomenon has already been discussed in detail by Olivieri et al. (2017) and is one of the main arguments for a Southern European origin of H3. Although this data supports much more a Neolithic arrival, rather than the Mesolithic migration argues by Olivieri et al. (2017).
One Swiss Early Bronze Age sample falls within H3af (5 kya) (Olalde et al., 2018), which is mainly found in Germany, but also in Austria, the Basque country, Bulgaria, France and Spain. Smaller subclades of H3 that include British and Irish samples are:

- H3r (6.6 kya), which contains a basal Finnish sample, a Welsh clade dating to 2.6 kya, a Danish clade dating to 2.6 kya, and H3r1 (1.3 kya), which also contains an Italian sample;
- H3q (6.4 kya), which is mainly Scandinavian with samples from Sweden, Norway and Denmark, but including a clade with a French and an English sample;
- H3v (8.2 kya), which has a Late Neolithic Danish (Allentoft et al., 2015) and an Early Bronze Age sample from the Czech Republic within subclade H3v+16093 (Olalde et al., 2018) and generally comprises of an English, Irish and a Dutch basal sample, as well as some samples from Shetland, Germany, France, Belgium and Scotland;
- H3aa (5.7 kya), which includes Scottish, French and Irish samples;

- H3y (2.6 kya) contains German and a Polish basal sample and further includes British, Spanish, German and Basque samples that mainly all directly branch off the root;
- H3ao (6.5 kya), which contains Irish, German, Italian and Welsh samples and one German Early Bronze Age sample that belongs to H3ao2 (Brotherton et al., 2013);
- the cluster defined by position 16145 (9.7 kya), which contains British, Dutch and USA samples;
- H3ah (2.9 kya), which is mainly South European with samples from Spain, Italy and the Basque country, but also includes one Irish clade dating to 2.6 kya;
- H3ar (4.2 kya), which is a small branch with Italian and Spanish samples, including also one Welsh cluster (1.3 kya);
- H3z (8.4 kya), which is mainly Northwest/Central European found in Scotland, Poland, Britain, Denmark and France, but also in the Basque country and Russia;
- H3au (4.6 kya), which is divided into a Spanish/Basque cluster and a British branch;
- H3ag (6.1 kya), which is found in Belgium and Britain;
- the branch defined by position 16172 (3.5 kya), which contains samples from the Basque country, Ireland and Britain;
- H3ap (4.5 kya), which contains a basal English sample and has samples from Spain, USA and Britain directly branching off, as well as a small cluster of a Polish, a Danish and a Swedish sample.

Two small clusters, H3as (1.7 kya) and H3aq (1.3 kya), are only found in Ireland. Several British and Irish samples directly branch from H3. The phylogenetic analysis alone is inconclusive to determine the origin of haplogroup H3. Neither the tree structure or the distribution of H3 and its frequencies point towards one geographic origin, but most of the Northern European lineages had a source in Iberia. However, the analysis of the subhaplogroups, specifically those carrying only British or Irish samples, has given indications of when and from where lineages arrived in the British Isles and Ireland. Examples for Irish clusters are H3as and H3aq that directly branch off the root

of H3, and a Welsh clade within H3r that possibly arrived to Wales from Finland or generally from Scandinavia. Ancient British samples, such as the basal H3 sample from Neolithic Scotland (∼5.7 kya), support the arrival and establishment of specific H3 lineages on the islands.

C.3.2 Phylogeny of H6

For the phylogenetic reconstruction of mitochondrial haplogroup H6, a total of 456 published and unpublished sequences were used. 303 published samples consisted of 285 modern and 20 ancient mitochondrial genomes. The origin of the published data is shown in Table C.3. A total of 153 newly generated samples was included in the study, of which 31 from the British Isles and Ireland, five from Belgium, three were from France, 71 from Germany, seven from Italy, 14 from the Netherlands and 22 from Spain.

TABLE C.3: All published samples used for the phylogenetic reconstruction of haplogroup H6. Ancient sample numbers are indicated in bold.

Country	Sample Number	Reference
Armenia	3	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*) and Margaryan et al. (2017)
	1	Allentoft et al. (2015)
Basque country	10	Behar et al. (2012b)
British Isles	11	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*) and Behar et al. (2012a)
	4	Martiniano et al. (2016) and Olalde et al. (2018)
Bulgaria	1	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*)
Colombia	1	1000 Genomes Project Consortium (2010)

Continued on next page

Table C.3 – continued from previous page

Country	Sample Number	Reference
Czech Republic	5	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*) and Mielnik-Sikorska et al. (2013)
	2	Olalde et al. (2018)
Denmark	79	Behar et al. (2012a), Li et al. (2014), and Raule et al. (2014)
Egypt	**1**	Schuenemann et al. (2017)
Finland	8	1000 Genomes Project Consortium (2010), Behar et al. (2012a), and Raule et al. (2014)
France	3	Behar et al. (2012a) and Lippold et al. (2014)
Germany	5	Behar et al. (2012a)
	1	Brotherton et al. (2013)
Hungary	**2**	Neparáczki et al. (2017)
Hutterite	1	Pichler et al. (2010)
Iran	2	Roostalu et al. (2007) and Derenko et al. (2013)
Ireland	13	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), Behar et al. (2012a), and Batini et al. (2017)
Israel	6	Behar et al. (2008), Behar et al. (2012a), and Lippold et al. (2014)
Italy	10	1000 Genomes Project Consortium (2010), Achilli et al. (2004), Behar et al. (2012a), Raule et al. (2014), and Olivieri et al. (2017)
Lithuania	1	Behar et al. (2012a)

Continued on next page

Table C.3 – continued from previous page

Country	Sample Number	Reference
Near East	1	Shlush et al. (2008)
Netherlands	2	Behar et al. (2012a)
	1	Olalde et al. (2018)
Norway	6	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), Behar et al. (2012a), and Batini et al. (2017)
Pakistan	1	Lippold et al. (2014)
Palestine	1	Behar et al. (2012a)
Poland	8	Behar et al. (2012a) and Mielnik-Sikorska et al. (2013)
	4	Allentoft et al. (2015), Fernandes et al. (2018), Juras et al. (2018), and Olalde et al. (2018)
Portugal	1	Behar et al. (2012a)
Russia	9	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), Malyarchuk et al. (2010a), Behar et al. (2012a), Mielnik-Sikorska et al. (2013), and Derenko et al. (2014)
	4	Allentoft et al. (2015), Haak et al. (2015), and Mathieson et al. (2015)
Serbia	5	Davidovic et al. (2015) and Batini et al. (2017)
Slovakia	5	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), Behar et al. (2012a), and Mielnik-Sikorska et al. (2013)

Continued on next page

Table C.3 – continued from previous page

Country	Sample Number	Reference
Spain	3	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), Behar et al. (2012a), and Cardoso et al. (2013)
Sweden	3	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*) and Behar et al. (2012a)
Switzerland	1	Behar et al. (2012a)
Ukraine	4	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), Behar et al. (2012a), and Mielnik-Sikorska et al. (2013)
United Arab Emirates	1	Behar et al. (2012a)
Unknown	57	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), Bertolin et al. (2011), Behar et al. (2012a), and Mielnik-Sikorska et al. (2013)
Caucasian USA	17	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), 1000 Genomes Project Consortium (2010), Zaragoza et al. (2010), and Just et al. (2014)

The beginning of the dispersal of haplogroup H6 dates to 17.7 kya, before the LGM (Roostalu et al., 2007). It is the most frequent subclade of H in Central Asia, is found in the Arabian peninsula and is one of the oldest clades in the Near East and Caucasus. It is defined by the mutations 239, 16362 and 16482, and is divided into three subclades: H6a (12.6 kya), which accounts for 90% of the phylogenetic tree, H6b (10.6 kya), and H6c (3.4 kya). While H6a is mainly Central/Northwestern European, H6b is found in Israel, Palestine

and Armenia (50%), Italy (33%) and England (8%), with the British samples being most recent and nested within the Levantine/Armenian samples. H6c on the other hand is entirely Central/Northwest European, found in Germany, the Netherlands, Denmark, Slovakia, Poland, Wales, Scotland, Czech Republic, Norway and Sweden. The ρ age estimates and 95% confidence intervals for the major nodes are given in Table C.4. These age estimates especially for the European branches suggest a much older arrival and expansion within Europe than stated by Roostalu et al. (2007) of only 3.4 kya. Several Near Eastern lineages are interspersed within H6a, as well as Russian lineages. Basal to H6a, H6a1a and H6a1b are Yamnaya samples and H6a1b in addition also has a basal Corded Ware sample. This suggests a spread from the Near East to Russia in the Neolithic, possibly also with some lineages along the Mediterranean into Italy, but was then mixed in the Yamnaya/Corded Ware expansions into Europe at $\sim$4.5-5 kya. However, some European clusters look deeper, possibly originating in the Neolithic. Therefore, we might be seeing two conflated signals.
The major subclade H6a is clustered into H6a1 (10 kya) and H6a2 (4.6 kya), of which H6a1 includes the majority of the samples, and almost all of the ancient data. A Bell Beaker sample from the Czech Republic falls at the root of H6a. H6a1 is divided into H6a1a (6.5 kya) and H6a1b (8.1 kya), by far the two largest subclades within H6, which are about the same size. Whereas H6a1a branches in several clusters, H6a1b mainly consists of two bigger subclades and a few smaller ones. Of the six ancient samples belonging to H6a1a, four are from the Bronze Age (Germany (Brotherton et al., 2013), the Netherlands (Olalde et al., 2018), Poland (Olalde et al., 2018) and Russia (Mathieson et al., 2015)), one sample dates to the Roman period in England (Martiniano et al., 2016), while a Hungarian individual from the 10th Century AD was also identified as belonging to H6a1a (Neparáczki et al., 2017). The subclade of H6a1a1 (7.4 kya) is half Near Eastern and half European, H6a1a5 (2.6 kya) is an Eastern subclade found in Russia and Slovakia, and also subclade H6a1a4 (2.6 kya) is a North European/Eastern clade found in Denmark, Finland, Norway, Poland and Russia. All other subclades are almost entirely European, with most of the samples from Central and Northwest Europe. A Roman individual from England (Martiniano et al., 2016) was identified as H6a1a10 (4.2 kya). Within H6a1a no exclusively British/Irish clade was found. British and Irish samples either cluster with mainly Danish and German samples (as they account for the majority of the samples) or are directly branching off the root of H6a1a.

Polish and Russian samples are found at the root of H6a1b, while, with very minor exceptions, the rest of H6a1b is Northwest/Central European. The majority of ancient samples belonging to clade H6a1b are from the Early Bronze Age in England (Olalde et al., 2018), Poland (Allentoft et al., 2015) and Russia (Allentoft et al., 2015; Haak et al., 2015). This suggests that clade H6a1b was already widespread across the Eastern and Western extents of Europe during the Bronze Age, supporting the arrival of H6a with Near Eastern farmers during the Neolithic. Furthermore, an individual from the 10th century in Hungary belongs to this subclade (Neparáczki et al., 2017). H6a1b can be divided into the major clades H6a1b2 (4.2 kya), H6a1b3 (4.8 kya) and H6a1b4 (4.4 kya) plus several minor clades (including the recently occurring H6a1b1, which is found only in the Druze population (Behar et al., 2008)). The root of H6a1b2 has two samples from England (Martiniano et al., 2016) and Scotland (Olalde et al., 2018), from the Roman period and Middle Bronze Age respectively, as well as modern day Germany, Scotland and Spain. H6a1b2a (2.6 kya) is a British/Danish cluster, while H6a1b2e (3.3 kya) is almost completely Danish, with an Irish exception. Other small, recently occurring clusters of British/Irish samples branch directly off H6a1b2. Examples are an Irish/Scottish cluster (5.2 kya), two Shetland clusters, one of which can be dated to 400 years ago, and a Welsh cluster. Within H6a1b3, subclade H6a1b3a (1.9 kya) is entirely Welsh (excluding one sample from 'Britain'). This clade, and one Welsh individual that branches from the root of H6a1b3, are the only British samples within this clade, with the majority of samples being from Denmark and Germany. H6a1b4 had one mixed clade of Dutch, Irish and Italian samples (2.2 kya), with English, Irish and Shetland samples branching directly from the root. An Early Bronze Age sample from the Czech Republic also belongs to H6a1b4 (Olalde et al., 2018).

H6a2 includes a few Irish samples and is otherwise a mix of an ancient Russian from the Middle Bronze Age (Mathieson et al., 2015) at the root, and modern Basque, Danish, Iberian and Serbian samples. In the Near Eastern/European clade of H6b, one English sample is found in H6b1 (7.9 kya) on a branch dating to 3.9 kya, whereas, in H6c one Scottish and one Welsh sample can be found, indicating very recent minor gene flow from the Near East to the British Isles. Out of the 18 ancient samples in this clade, only two belong to H6b; a Late Bronze Age Armenian (Allentoft et al., 2015) and a Ptolemaic individual from Egypt (Schuenemann et al., 2017).

TABLE C.4: Rho age estimates and 95% confidence intervals of main nodes of the H6 phylogenetic tree.

Node	Rho	Standard error	Age	95% confidence interval
> H6	6.517	1.726	17,636	8255 - 27,464
>> H6a	4.728	1.238	12,615	6017 - 19,456
>>> H6a1	3.773	0.783	9988	5850 - 14,226
>>>> H6a1a	2.500	0.274	6549	5119 - 7991
>>>>> H6a1a1	2.833	0.833	7443	3108 - 11,893
>>>>>> H6a1a1a	0.500	0.500	1287	0 - 3843
>>>>> H6a1a2	4.529	1.218	12,066	5592 - 18,776
>>>>>> H6a1a2a	2.857	0.969	7507	2476 - 12,694
>>>>>> H6a1a2b	3.000	1.158	7892	1886 - 14,120
>>>>>>> H6a1a2b1	2.222	0.801	5807	1680 - 10,043
>>>>> H6a1a3	0.833	0.278	2151	742 - 3574
>>>>>> H6a1a3a	1.000	0.707	2585	0 - 6244
>>>>> H6a1a4	1.000	0.354	2585	789 - 4404
>>>>> H6a1a5	1.000	0.529	2585	0 - 5315
>>>>> H6a1a6	1.000	0.577	2585	0 - 5566
>>>>> H6a1a7	0.600	0.346	1546	0 - 3315
>>>>> H6a1a8	2.600	1.217	6816	553 - 13,327
>>>>>> H6a1a8a	0.750	0.433	1935	0 - 4155
>>>>> H6a1a9	1.250	0.559	3239	396 - 6136
>>>>> H6a1a10	1.600	0.894	4159	0 - 8847
>>>> H6a1b	3.053	0.636	8035	4704 - 11,433

Continued on next page

Table C.4 – continued from previous page

Node	Rho	Standard error	Age	95% confidence interval
>>>>> H6a1b2	1.608	0.245	4179	2917 - 5450
>>>>>> H6a1b2a	1.000	0.577	2585	0 - 5566
>>>>>> H6a1b2d	0.333	0.333	857	0 - 2551
>>>>>> H6a1b2e	1.286	0.892	3333	0 - 7984
>>>>> H6a1b3	1.855	0.402	4831	2761 - 6928
>>>>>> H6a1b3a	0.727	0.257	1876	573 - 3190
>>>>>> H6a1b3b	1.400	0.663	3632	256 - 7085
>>>>> H6a1b4	1.577	0.409	4098	2001 - 6223
>>> H6a2	1.778	0.497	4628	2075 - 7223
>> H6b	4.000	0.943	10,610	5620 - 15,743
>>> H6b2	3.750	1.299	9927	3118 - 17,009
>>> H6b1	3.000	1.106	7892	2151 - 13,835
>> H6c	1.304	0.359	3381	1550 - 5235
>>> H6c1	1.000	0.577	2585	0 - 5566

C.3.3 Phylogeny of H7

For the phylogenetic reconstruction of mitochondrial haplogroup H7, an overall total of 329 published and unpublished sequences was used. The samples comprising the published dataset are displayed in Table C.5. Only seven ancient samples have been reported to belong to haplogroup H7 so far, and all were included in the analysis. A total of 95 newly generated modern sequences, including two from Belgium, 20 from the British Isles and Ireland, four from France, 40 from Germany, four from Greece, 12 from Italy, one from Luxembourg, six from the Netherlands and six from Spain, were used in this study.

TABLE C.5: All published samples used for the phylogenetic reconstruction of haplogroup H7. Ancient sample numbers are indicated in bold.

Country	Sample Number	Reference
Afghanistan	1	Lippold et al. (2014)
Armenia	2	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*) and Schönberg et al. (2011)
Austria	1	Behar et al. (2012a)
Basque country	2	Behar et al. (2012b)
Brazil	3	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*) and de Araujo et al. (2015)
Bulgaria	1	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*)
	2	Mathieson et al. (2018)
Canada	1	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*)
China	1	Derenko et al. (2014)
Croatia	3	Behar et al. (2012a) and Fu et al. (2012)
	1	Mathieson et al. (2018)
Czech Republic	1	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*)
	1	Olalde et al. (2018)
Denmark	62	Behar et al. (2012a), Li et al. (2014), and Raule et al. (2014)

Continued on next page

Table C.5 – continued from previous page

Country	Sample Number	Reference
British Isles & Ireland	20	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), Behar et al. (2012a), Lloyd et al. (2015), and Batini et al. (2017)
Europe	6	Coble et al. (2004)
Finland	6	1000 Genomes Project Consortium (2010), Behar et al. (2012a), and Raule et al. (2014)
France	4	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*) and Behar et al. (2012a)
Georgia	1	Schönberg et al. (2011)
Germany	7	Zsurka et al. (2007), Behar et al. (2012a), and Batini et al. (2017)
	2	Brotherton et al. (2013)
Greece	3	Behar et al. (2012a), Raule et al. (2014), and Batini et al. (2017)
Hungary	1	Behar et al. (2012a)
	1	Lipson et al. (2017)
India	1	Palanichamy et al. (2015)
Iran	4	Schönberg et al. (2011) and Derenko et al. (2013)
Israel	6	Shlush et al. (2008) and Lippold et al. (2014)

Continued on next page

Table C.5 – continued from previous page

Country	Sample Number	Reference
Italy	34	1000 Genomes Project Consortium (2010), Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), Achilli et al. (2004), Gasparre et al. (2007), Behar et al. (2012a), Bodner et al. (2015), Raule et al. (2014), Lippold et al. (2014), Batini et al. (2017), and Olivieri et al. (2017)
Lithuania	2	Behar et al. (2012a)
Netherlands	1	Behar et al. (2012a)
Norway	1	Behar et al. (2012a)
Poland	4	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*) and Behar et al. (2012a)
Romania	1	Behar et al. (2012a)
Russia	3	Malyarchuk et al. (2010a), Derenko et al. (2014), and Lippold et al. (2014)
Serbia	1	Batini et al. (2017)
Spain	1	Fregel et al. (2009)
Sweden	1	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*)
Switzerland	1	Behar et al. (2012a)
Ukraine	1	Behar et al. (2012a)
Unknown	31	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), Pereira et al. (2007), Gómez-Carballa et al. (2011), Behar et al. (2012a), and Lippold et al. (2014)

Continued on next page

Table C.5 – continued from previous page

Country	Sample Number	Reference
USA	13	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*), 1000 Genomes Project Consortium (2010), Just et al. (2014), and Sequeira et al. (2015)

Haplogroup H7 has an age estimate of 9.8 kya, is defined by the mutation at position 4793, and is divided into several large and small subclades. The biggest subclades are H7a (8.5 kya), H7b (5.9 kya), H7c (8.3 kya) and H7d (6.8 kya). Given the structure of the tree, the most likely origin of haplogroup H7 is Southeastern Europe. This assumption is based on basal Neolithic samples from Hungary and Bulgaria, suggesting a Neolithic dispersal into the Central and Eastern Mediterranean and into Central Europe, presumably with the *Linearbandkeramik* (LBK). The age estimates and 95% confidence intervals for all branches of H7 are given in Table C.6.

TABLE C.6: Rho age estimates and 95% confidence intervals of main nodes of the H7 phylogenetic tree.

Node	Rho	Standard error	Age	95% confidence interval
> H7	3.687	0.447	9753	7382 - 12,157
>> H7a	3.429	0.901	9051	4325 - 13,911
>>> H7a1	2.500	0.420	6549	4359 - 8768
>>>> H7a1a	5.500	1.837	14,766	4949 - 25,104
>>>> H7a1b	1.500	0.343	3895	2139 - 5672
>>>> H7a1c	1.000	0.333	2585	891 - 4299

Continued on next page

Table C.6 – continued from previous page

Node	Rho	Standard error	Age	95% confidence interval
>>>> H7a1d	0.500	0.354	1287	0 - 3090
>>> H7a2	1.000	0.707	2585	0 - 6244
>> H7b	2.420	0.396	6335	4274 - 8423
>>> H7b1	1.867	0.422	4863	2690 - 7066
>>> H7b2	1.706	0.428	4438	2238 - 6670
>>>> H7b2a	1.250	0.559	3239	396 - 6136
>>> H7b4	1.000	0.577	2585	0 - 5566
>>> H7b5	0.500	0.500	1287	0 - 3843
>>> H7b6	1.500	0.866	3895	0 - 8425
>> H7c	3.339	0.701	8809	5127 - 12,572
>>> H7c1	2.467	0.693	6459	2869 - 10,130
>>> H7c2	0.250	0.250	642	0 - 1909
>>> H7c3	1.000	0.745	2585	0 - 6444
>>> H7c4	2.667	1.247	6995	570 - 13,680
>>> H7c6	2.400	1.095	6281	650 - 12,115
>> H7d	2.811	0.649	7382	3999 - 10,835
>>> H7d1	0.857	0.404	2213	167 - 4288
>>> H7d2	1.667	0.782	4334	345 - 8429
>>>> H7d2a	0.667	0.471	1719	0 - 4134
>>> H7d3	3.000	1.225	7892	1544 - 14,488
>>> H7d4	1.000	0.707	2585	0 - 6244
>>> H7d5	1.000	0.707	2585	0 - 6244

Continued on next page

Table C.6 – continued from previous page

Node	Rho	Standard error	Age	95% confidence interval
>> H7e	1.143	0.404	2959	902 - 5044
>> H7f	1.333	0.667	3457	68 - 6924
>> H7g	0.800	0.490	2065	0 - 4581
>> H7h	1.250	0.500	3239	694 - 5828
>> H7i	1.857	0.958	4838	0 - 9887

Two ancient samples are found basal to H7. One sample is from Bulgaria (Mathieson et al., 2018) and the other from Hungary (Lipson et al., 2017), from the Eneolithic and the Middle Neolithic, respectively. They cluster with modern Danish and French samples. Another ancient sample from Chalcolithic Bulgaria (Mathieson et al., 2018) falls at the root of H7a1 (6.0 kya), together with modern Danish and German samples. A Neolithic sample from Croatia (Mathieson et al., 2018) is basal to H7c and a late Neolithic sample from Czech Republic (Olalde et al., 2018) is basal to H7d, together with an ancient sample from Germany dating to the Late Neolithic (Brotherton et al., 2013). A sample from the Unetice culture in Germany (Brotherton et al., 2013) falls at the root of H7h together with a modern sample from Spain, which are generally very few and possibly arrived with the Bronze Age or after. Similarly no ancient lineages have been reported from the British Isles or any other region of the Atlantic fringe.

Subclade H7a dates to ∼8.5 kya and is almost entirely European. The majority of samples within H7a fall into H7a1 (6.0 kya), and only two samples fall into H7a2 (2.6 kya). All other samples branch directly off the root of H7a. Subclade H7a1 is further divided into six larger subclades and two small subclades, with several samples just branching off the root of H7a1. H7a1c (2.3 kya) is entirely Central/Northwestern European, and is present in the British Isles (England, Scotland and Wales), Germany and the Netherlands. H7a1b (4.0 kya) is the largest of the subclusters within H7a1 and is mainly found in Denmark, England and Scotland, but also in Finland and Sardinia. H7a1a (5.3 kya) is entirely Italian, while sub-branch H7a1d (1.3 kya) is completely Danish. A very small and very recent clade is present in Britain and Russia.

Another very small clade dates to 2.6 kya and is entirely German. A small Danish/German clade dates to 2.6 kya, and a slightly bigger cluster dating to 5.2 kya is found in Denmark, Germany, Italy and the Netherlands.
Subclade H7b (5.9 kya) is found both in Europe and the Near East. It has two larger subclusters, H7b1 (4.8 kya) and H7b2 (4.4 kya), as well as several small clusters and branches that are all basal to the root of H7b. H7b1 is ∼40% Near Eastern (samples from Armenia and Israel), ∼40% South/Southeast European (samples from Italy, Greece and Sardinia) and ∼20% Central European (samples from Czech Republic and the Netherlands). The subclades of H7b suggest a Middle or Late Neolithic migration into the Near East and Central and Southern Europe.
The subclade of H7c (8.3 kya) is generally more South European. It is divided into one large subcluster, H7c1 (5.8 kya), and seven small clusters (H7c2 to H7c6 and two unnamed clusters). Around 50% of H7c1 are Italian and 10% from Israel, while the rest are from France, Germany, Russia, Scotland, Switzerland and Wales. One subbranch of H7c1 seems entirely British, with samples from Scotland and Wales, dating to 1.3 kya. Except for one German cluster (2.6 kya), all other subclusters of H7c are found in both North/Central and South Europe. H7c5 (2.6 kya) is a mix of Iranian and Italian samples. Generally, this cluster looks South European and suggests a spread to the Northwest of Europe during the Bronze Age.
The subclade of H7d (6.8 kya) is almost entirely Northwest European (∼ 77%), with other samples being from Belgium, Germany and the Netherlands (∼13%) or Spain. The largest subclade (1.8 kya), with ∼50% of the samples, is only present in Denmark, Norway and Shetland, with the samples from Shetland branching off recently from the Danish/Norwegian cluster. Subclades H7d1 (2.2 kya) and H7d2a (1.7 kya) are both Irish/British clusters. This suggests a Neolithic expansion, which is would be supported by the Middle Neolithic and Late Neolithic samples from Germany and the Czech Republic.
Most of the small subclades branching off the root of H7 are completely European, with only a very few samples from Asia, the Near East or Caucasus spread among them. H7f (3.5 kya) is another entirely British/Irish cluster, and an unnamed cluster dating to 1.3 kya, and identified by mutations 11689 and 14257, is only found in Wales. These British clusters seem to be very recent, possibly from the Iron Age.

C.3.4 Phylogeny of H8

For the phylogenetic reconstruction of mitochondrial haplogroup H8, a total of 124 published and unpublished sequences was included. The published dataset is given in Table C.7, and includes four ancient samples (three Armenian and one Moldovian) (Allentoft et al., 2015; Juras et al., 2017; Margaryan et al., 2017). The 78 newly generated sequences are from the British Isles (n=4), France (n=1), Germany (n=2), Greece (n=67, mainly due to a localised sampling/founder effect in Crete) and Italy (n=4).

TABLE C.7: All published samples used for the phylogenetic reconstruction of haplogroup H8. Ancient sample numbers are indicated in bold.

Country	Sample Number	Reference
Armenia	5	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*) and Margaryan et al. (2017)
British Isles & Ireland	2	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*) and Behar et al. (2012a)
China	2	Derenko et al. (2014) and Lippold et al. (2014)
Croatia	1	Fu et al. (2012)
Denmark	1	Raule et al. (2014)
France	2	Behar et al. (2012a) and Batini et al. (2017)
Germany	1	Behar et al. (2012a)
Greece	1	Batini et al. (2017)
Italy	8	Achilli et al. (2004), Behar et al. (2012a), Raule et al. (2014), and Olivieri et al. (2017)
Poland	1	Behar et al. (2012a)
Portugal	1	Pereira et al. (2007)

Continued on next page

Table C.7 – continued from previous page

Country	Sample Number	Reference
Near East	1	Shlush et al. (2008)
Russia	1	Ingman et al. (2007), Sukernik et al. (2012), Duggan et al. (2013), and Derenko et al. (2014)
Spain	3	1000 Genomes Project Consortium (2010), Gómez-Carballa et al. (2011), and Batini et al. (2017)
Sweden	1	Behar et al. (2012a)
Unknown	3	Mishmar et al. (2003) and Behar et al. (2012a)
Caucasian USA	2	1000 Genomes Project Consortium (2010)

Haplogroup H8 appears to be one of the oldest subhaplogroups of haplogroup H, with an age estimate of 17.6 kya. It is defined by four mutations: 709, 13101, 16288 and 16362, which divide the phylogenetic tree into three large subclades: H8a (5.5 kya), H8b (10.2 kya) and H8c (10 kya). The age estimates and 95% confidence intervals for all branches of H8 are given in Table C.8.

TABLE C.8: Rho age estimates and 95% confidence intervals of main nodes of the H8 phylogenetic tree.

Node	Rho	Standard error	Age	95% confidence interval
> H8	6.520	1.886	17,645	7418 - 28,405
>> H8a	2.100	0.900	5482	864 - 10,238
>>> H8a1	1.000	0.433	2585	388 - 4815

Continued on next page

Table C.8 – continued from previous page

Node	Rho	Standard error	Age	95% confidence interval
>> H8b	3.877	1.596	10,272	1928 - 19,029
>>> H8b1	2.833	1.106	7443	1718 - 13,371
>> H8c	3.692	0.944	9769	4800 - 14,884
>>> H8c1	1.667	0.745	4334	528 - 8236
>>> H8c2	0.600	0.346	1546	0 - 3315

Given the structure of the tree, the most likely origin of haplogroup H8 is in Southeast Europe, as the older subclades look South and Eastern European, H8b and H8c respectively. The much younger subclade H8a, is mainly Near Eastern and suggests a migration from Southeast Europe into the Near East ∼5.5 kya. This is further highlighted by the almost extreme separation of geographic areas based on the subclades of H8. H8a is the youngest subclade (5.5 kya) and includes an ancient sample from Armenia and the majority of its samples are from the Near Eastern Druze population. Its subbranch H8a1 (2.6 kya) is mainly found in Armenia (56%), including one ancient sample from the Late Bronze Age (Allentoft et al., 2015) and two from the Early and Late Bronze Age (Margaryan et al., 2017), and other samples are from Italy (22%), and Greece and Germany (each 11%). The subclade of H8b can be divided into the Asian/European Russian branch of H8b1 (7.4 kya) and a branch of solely Cretan samples. As the Cretan samples almost all have the same sequence it appears like a founder effect, especially as the branch is very recent. H8b1 consists of 18% Chinese and 82% Russian samples. A few other samples that belong to haplogroup H8b cluster separately and originate in France, Sardinia and Wales. Subclade H8b is the oldest cluster within H8 and extends into Siberia and China, with the exception of a newly reported Mediterranean clade found in Greece (Foody et al. in preparation). Subclade H8c has a basal Sardinian sample and can generally be divided into different European branches, suggesting a European origin. H8c1 (4.3 kya) is Central/Northern European, present in Denmark, Poland and Sweden. H8c2 is recent (∼1.5 kya) and entirely British, with samples from England, Scotland and Wales. Another fairly large subclade (23% of H8c) is completely South

European, present in Croatia, Italy and Spain, with the majority of the samples being from Italy. Four smaller branches are found in France and Spain, France and Wales, Germany and Italy, and Italy, respectively.
Among Asian populations haplogroups H6 and H8 are distinctly divergent from European haplotypes, which was stated to highlight a long-time separation of Asian and European H6 and H8 mtDNA genome pools (Loogväli et al., 2004). However, no deep Asian subclades could be identified in H6 to support this.

C.3.5 Phylogeny of H31

For the phylogenetic reconstruction of mitochondrial haplogroup H31, a total of 55 sequences was used. Of this, 39 are published data (see Table C.9) and 16 sequences (Belgium 1; British Isles 2; Germany 11; Netherlands 2; British Isles 2) are newly generated.

TABLE C.9: All published samples used for the phylogenetic reconstruction of haplogroup H31.

Country	Sample Number	Reference
British Isles & Ireland	8	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*) and Behar et al. (2012a)
Denmark	11	Li et al. (2014) and Raule et al. (2014)
Finland	3	Behar et al. (2012a), Soini et al. (2012), and Raule et al. (2014)
Germany	1	Behar et al. (2012a)
Italy	1	Gasparre et al. (2007)
Netherlands	1	Batini et al. (2017)
Norway	1	Behar et al. (2012a)
Sweden	2	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*)

Continued on next page

Table C.9 – continued from previous page

Country	Sample Number	Reference
Caucasian USA	7	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*) and Just et al. (2014)
Unknown	4	Greenspan (*Direct Submission - Family Tree DNA - Genealogy by Genetics*) and Behar et al. (2012a)

Given the structure of the tree, the most likely origin of haplogroup H31 is Central or Northwestern Europe. This haplogroup is generally very small and not very frequent, given that it so far contains only 55 reported samples. To date, no ancient individuals have been designated as belonging to H31. It has an age estimate of 3.9 kya, which supports its recent origin, and also explains why it is not very common. It is defined by mutations 7930t and 10771, and is divided into three subbranches: H31a (3.6 kya), H31b (1.7 kya), and a branch defined by mutations at position 198 and 2789 (2.6 kya). The age estimates and 95% confidence intervals for all H31 branches are given in Table C.10.

TABLE C.10: Rho age estimates and 95% confidence intervals of main nodes of the H31 phylogenetic tree.

Node	Rho	Standard error	Age	95% confidence interval
> H31	1.491	0.443	3871	1604 - 6173
>> H31a	1.375	0.956	3567	0 - 8564
>> H31b	0.667	0.333	1719	34 - 3422

Subclade H31a is the largest of the subclades and of fairly mixed European

origin, with samples from Denmark, Netherlands, Northern Wales, Scotland and Southeast England. Subclade H31b is much smaller with samples from England, Germany, Netherlands and Norway. The yet unnamed and smallest third subclade, defined by positions 198 and 2789, solely consists of samples from Denmark and Sweden, suggesting a Scandinavian origin. It is not possible to determine a more precise geographic origin for H31a and H31b, as they both have a mixed distribution within Northwestern and Central Europe. Based on the origin of the majority of basal samples from H31, a general origin within Denmark or Germany could be argued for, with a main spread into the British Isles and the Netherlands. Haplogroup H31 appears to have spread from Scandinavia across Northern Europe, possibly during the Bronze Age dispersals.

C.4 Conclusion

The phylogenetic reconstruction of specific sublineages of haplogroup H has facilitated a more detailed picture of the distribution and origin of the haplogroup. Several lineages can be traced to their origin, and the inclusion of ancient data in the reconstruction can help to understand arrival times of specific lineages, as well as the entire haplogroup H itself. Due to the large scope of this project, the number of people involved, and the time it takes to reconstruct all phylogenetic trees of all haplogroup H lineages, this chapter unfortunately only shows a small part of the overall research, which should show a lot more details of the overall evolution of haplogroup H. However, the presented phylogenetic analyses already give a lot of information on the individual subhaplogroups. As the work on this project is still ongoing, the full reconstruction and Founder Analysis of haplogroup H are still to be completed. The results of these analyses should throw light onto the haplogroup H enigma, and help us to understand the origin of H and its arrival and spread in Europe.

Appendix D

Negative control analysis and DNA damage patterns

D.1 Negative controls

To assess anti-contamination measures, negative controls were introduced during each stage of sample processing for all samples. An air control and a water control were introduced during the drilling stage. The air control consisted of an empty 1.5ml Eppendorf tube with an open lid, which was either placed for one hour in the cleaned work area between sample drilling (TCD approach) or placed during drilling of one random sample within the drilling hood (University of Huddersfield approach). For the water control, 2ml of $dd H_2O$ were shaken, minus the steel ball, inside a metal shaker in the Mixer Mill (Retsch MM400). 1ml of the water control then went through the whole extraction process.
None of the controls that accompanied the three samples prepared at TCD were sequenced. Controls introduced during the extraction process at the University of Huddersfield were quantified using a Qubit 3.0 Fluorometer, and the DNA content of all controls was too low to be measured. However, negative controls from the extraction, USER treatment step, and library amplification step were sequenced alongside the samples on an Illumina HiSeq 2000 (NBAF Liverpool). For the sequencing runs undertaken on the Illumina HiSeq 4000 (Macrogen, South Korea), one extraction control and one library control were sequenced alongside the samples. The sequencing results for all controls are displayed in Table D.1.
The mitochondrial contamination of all controls was estimated to be lower than 0.001%.

TABLE D.1: Sequencing results of the controls from all sequencing runs.

	Control ID	No. of mapped reads	Mismatches	Mean mapping quality	Mean coverage	Std coverage
Liverpool	USER	82	9	6.47	0	0.0012
	Negative	7	1	0.61	0	0.0004
	Library	0	-	-	-	-
Macrogen	Library	369,809	115,634	35.58	0.0061	0.0795
	Extraction	62,025	25,066	35.57	0.0008	0.0291

D.2 DNA damage patterns

All analysed samples were treated with uracil-DNA-glycosylase (UDG) during the library preparation, which reduces the number of ancient DNA errors by cleaving uracils (deaminated cytosines). Therefore, the samples displayed reduced damage patterns, which were observed using mapDamage (Jónsson et al., 2013) and bamdamage (Malaspinas et al., 2014).

D.3 Sequence length distribution

The analysed samples displayed a sequence length distribution with a peak between 45-55 bp for each sample. Therefore, all samples display sequence lengths previously reported from ancient samples displaying an average sequence length of less than 100 bp (Shapiro et al., 2014) . A peak at 100 bp could be observed as well, representing all sequences greater than this length that were truncated in their sequencing runs.

D.4 Gamba calculations

As expected, the number of reads from the controls aligning to the human reference were much higher for the blanks run alongside the samples sequenced on a HiSeq 4000 (Macrogen, South Korea) than the number of reads found in the controls sequenced on a HiSeq 2500 (NBAF Liverpool, United Kingdom). Gamba calculations Gamba et al. (2014) were used to estimate the contamination of these blanks, and thus estimate the amou7nt of possible contamination across all samples. An overview of the number of raw and mapped reads of the samples on two lanes (lane one and lane three) can be found in Tables D.2 and D.3. The calculations of the contamination fraction of the lanes including controls (lane one and lane three) can be found in Tables D.4 and D.5.

The number of reads alignable to the human reference genome obtained for the controls were compared to the number of reads obtained for the samples

run on the same lane, and normalised by the total number of reads obtained in each HiSeq run, in order to estimate the laboratory contamination at each stage. This contamination estimation followed Gamba et al. (2014), calculating the contamination fraction (C) as follows: $C = (R_b/TV_b)/(R_s/TV_s)$ with R_b denoting the number of control reads aligning to the human genome, R_s the number of reads aligning to the human genome for the sample analysed in parallel with the control, T the total number of reads obtained in the specific HiSeq run and V the volume added to the pool loaded on the sequencer for controls (V_b) and samples (V_s).
For the two lanes (libraries K1-K12 and K25-K36) I calculated a C value of 0.9145% and 0.4951%, respectively. In comparison Gamba et al. (2014) reported 0% and 0.5%. The maximum number of reads for a blank in Gamba et al. (2014) was 282, which equates to ∼6316 reads on a HiSeq machine based on the 22.4 fold difference in read numbers. However, a lot also depends on the number of samples run per lane.

TABLE D.2: Number of raw reads and mapped reads per sample on lane one.

Library name	DNA ID	No. of raw reads	No. of mapped reads
K1	KD039	75,917,571	11,945,233
K2	KD040	70,078,405	12,546,601
K3	KD041	81,688,763	17,317,745
K4	KD042	47,347,014	13,145,087
K5	KD043	54,376,401	29,435,049
K6	KD044	90,937,892	39,830,659
K7	KD045	59,194,894	14,192,261
K8	KD046	83,482,447	14,412,344
K9	KD068	71,086,325	15,333,355
K10	KD069	72,140,138	12,412,080
K11	CE048M(B)	63,038,900	11,846,116
K12	Library Blank	57,344,199	13,441,923
total		826,632,949	205,858,453

TABLE D.3: Number of raw reads and mapped reads per sample on lane three.

Library name	DNA ID	No. of raw reads	No. of mapped reads
KD25	Extraction Blank	55,793,320	740,834
KD26	KD047	63,542,862	1,849,649
KD27	KD048	62,890,464	241,456
KD28	KD049	55,679,898	2,854,560
KD29	KD050	89,513,406	3,947,728
KD30	KD051	46,069,394	183,022
KD31	KD052	53,124,892	210,997
KD32	KD053	52,647,474	147,194
KD33	KD055	59,521,362	2,887,903
KD34	KD001	105,144,736	4,627,233
KD35	KD070	53,129,192	1,290,469
KD36	KD073	109,825,696	1,347,082
total		806,882,696	20,328,127

TABLE D.4: Overview of read numbers and the calculated contamination fraction of the lane with libraries K1-K12.

Control mapped reads	13,441,923
Total reads for lane	826,632,949
Total reads for lane/12	68,886,079
All mapped reads for 11 samples	192,416,530
Total reads for lane	826,632,949
Total reads for lane x 11/12	901,781,399
Control mapped reads / (total reads x 1/12)	0.1951
All mapped reads for 11 samples / (total reads per lane x 11/12)	0.2134
Contamination fraction (C)	**0.9145**

TABLE D.5: Overview of read numbers and the calculated contamination fraction of the lane with libraries K25-K36.

Control mapped reads	740,834
Total reads for lane	806,882,696
Total reads for lane/12	67,240,224
All mapped reads for 11 samples	19,587,293
Total reads for lane	806,882,696
Total reads for lane x 11/12	880,235,668
Control mapped reads / (total reads x 1/12)	0.0110
All mapped reads for 11 samples / (total reads per lane x 11/12)	0.0223
Contamination fraction (C)	**0.4951**

Appendix E

Sex determination of low coverage samples

A number of samples had a very low genome coverage and no genetic sex could be assigned based using the script by Skoglund et al. (2013). Therefore, the number of reads with "Mapping Quality 30" was plotted against the chromosome length. The number of reads with "Mapping Quality 30" mapping to the X-chromosome were compared to the number of reads which aligned to chromosome 7 (as they are of similar length) and normalised by the length of the reference sequence following Skoglund et al. (2015). Two better coverage samples, one female and one male individual, were plotted alongside, and the Skoglund score was calculated as comparison for the low coverage samples.

Equations to calculate the percentage of single or double copies of chromosome X are given in the following:
KD001 - male control - Rosemarkie:

$$(382,142/155.270,560)/(835,025/159.138,663) = 46.9\% \tag{E.1}$$

CE005 - female control - Oakridge:

$$(903,995/155.270,560)/(1.038,778/159.138,663) = 89.2\% \tag{E.2}$$

KD039:

$$(24,226/155.270,560)/(45010/159.138,663) = 55.2\% \tag{E.3}$$

KD040:

$$(33,376/155.270,560)/(67330/159.138,663) = 50.8\% \tag{E.4}$$

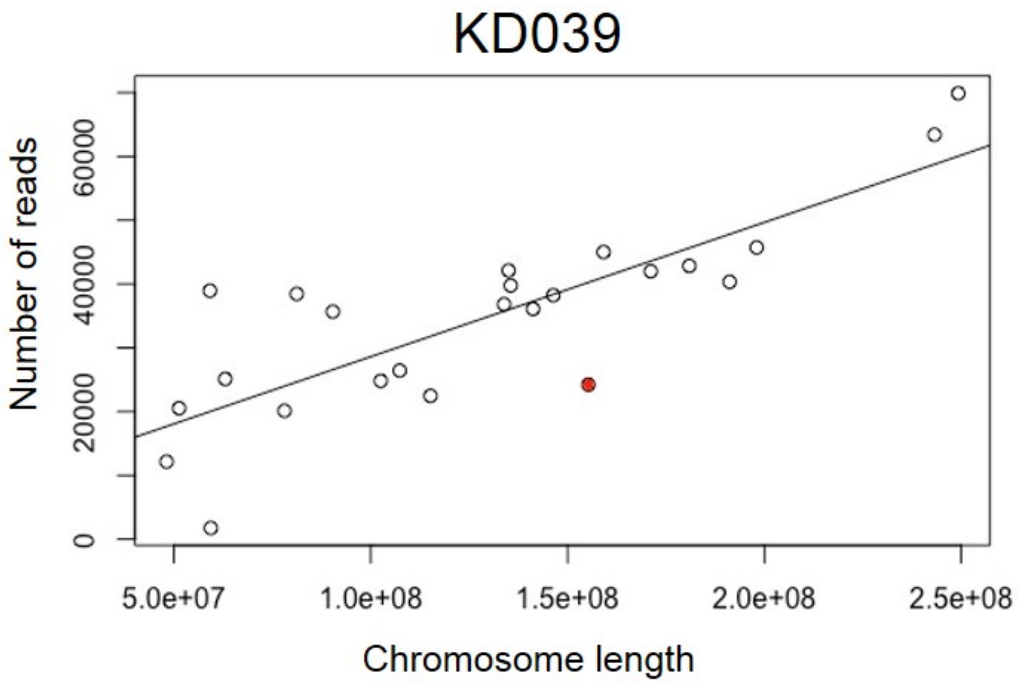

FIGURE E.1: Karyotype plot of KD039.

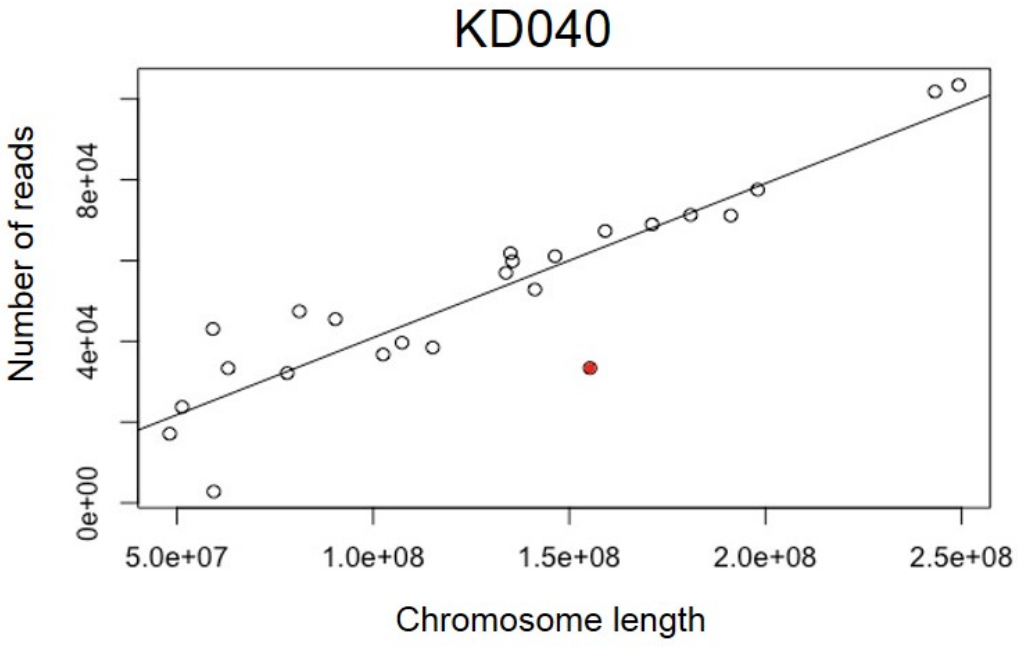

FIGURE E.2: Karyotype plot of KD040.

KD046:

$$(18,568/155.270,560)/(28,722/159.138,663) = 66.3\% \tag{E.5}$$

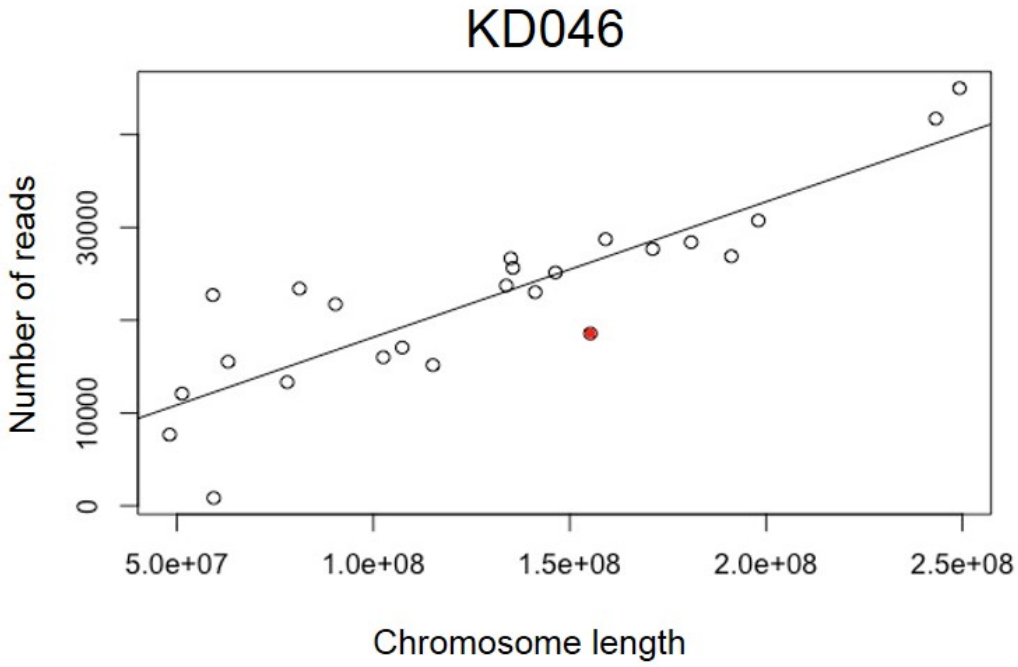

FIGURE E.3: Karyotype plot of KD046.

KD068:

$$(34,863/155.270,560)/(42,964/159.138,663) = 83.2\% \tag{E.6}$$

KD069:

$$(18,463/155.270,560)/(30,861/159.138,663) = 61.3\% \tag{E.7}$$

KD023:

$$(17,864/155.270,560)/(23,283/159.138,663) = 78.6\% \tag{E.8}$$

KD024:

$$(25,517/155.270,560)/(31,706/159.138,663) = 82.5\% \tag{E.9}$$

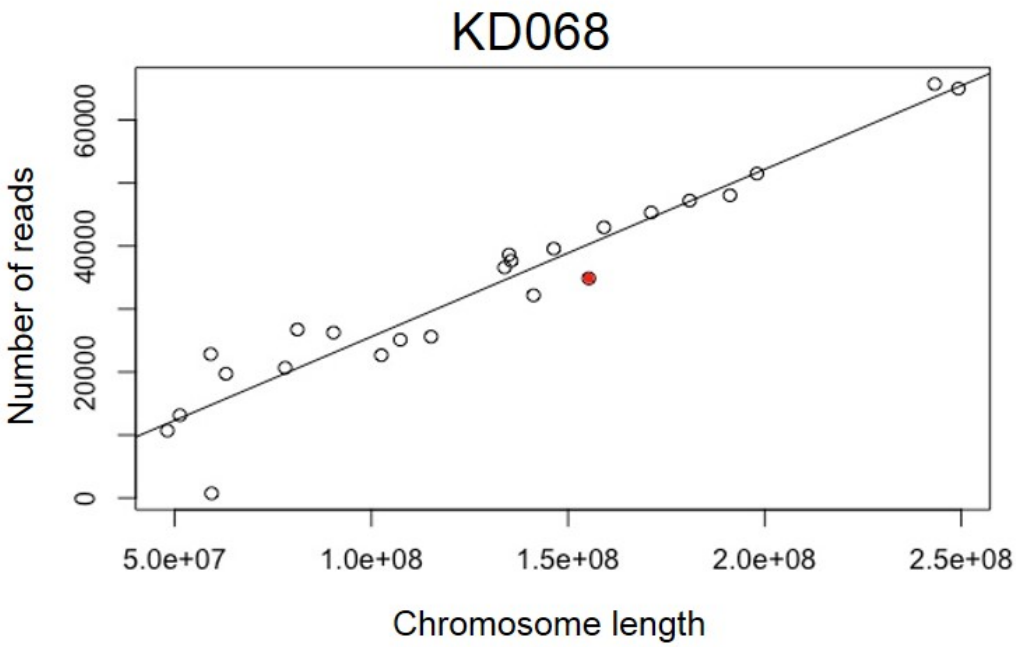

FIGURE E.4: Karyotype plot of KD068.

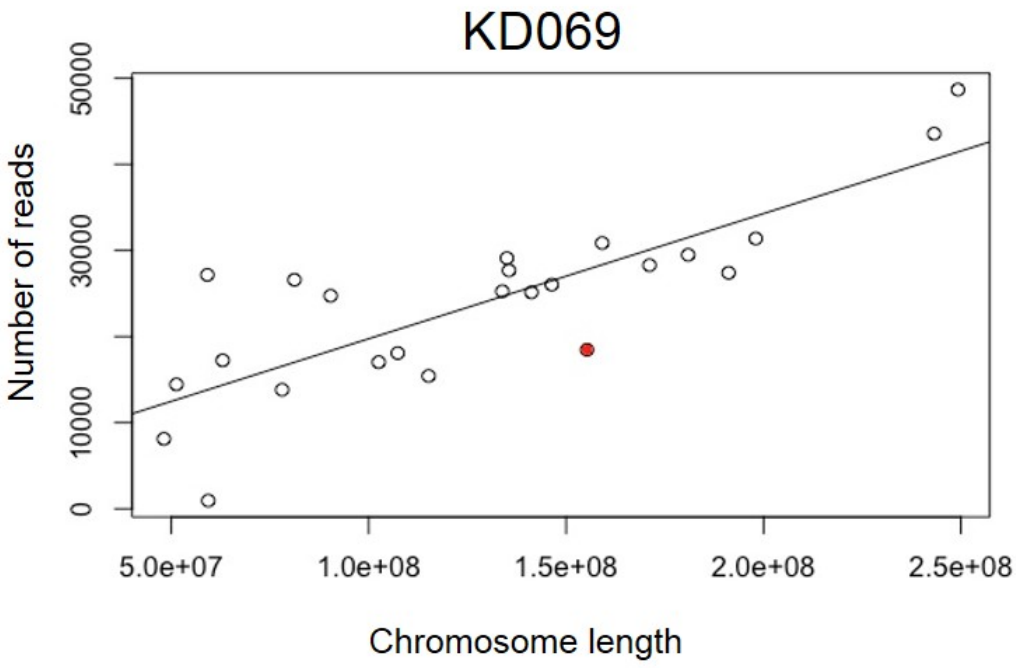

FIGURE E.5: Karyotype plot of KD069.

KD025:

$$(8,928/155.270,560)/(15,246/159.138,663) = 60.0\% \qquad \text{(E.10)}$$

KD027:

$$(31,761/155.270,560)/(40,570/159.138,663) = 53.2\% \tag{E.11}$$

KD028:

$$(14,479/155.270,560)/(20,232/159.138,663) = 73.3\% \tag{E.12}$$

KD029:

$$(20,889/155.270,560)/(24,537/159.138,663) = 87.3\% \tag{E.13}$$

KD071:

$$(64,939/155.270,560)/(85,128/159.138,663) = 78.2\% \tag{E.14}$$

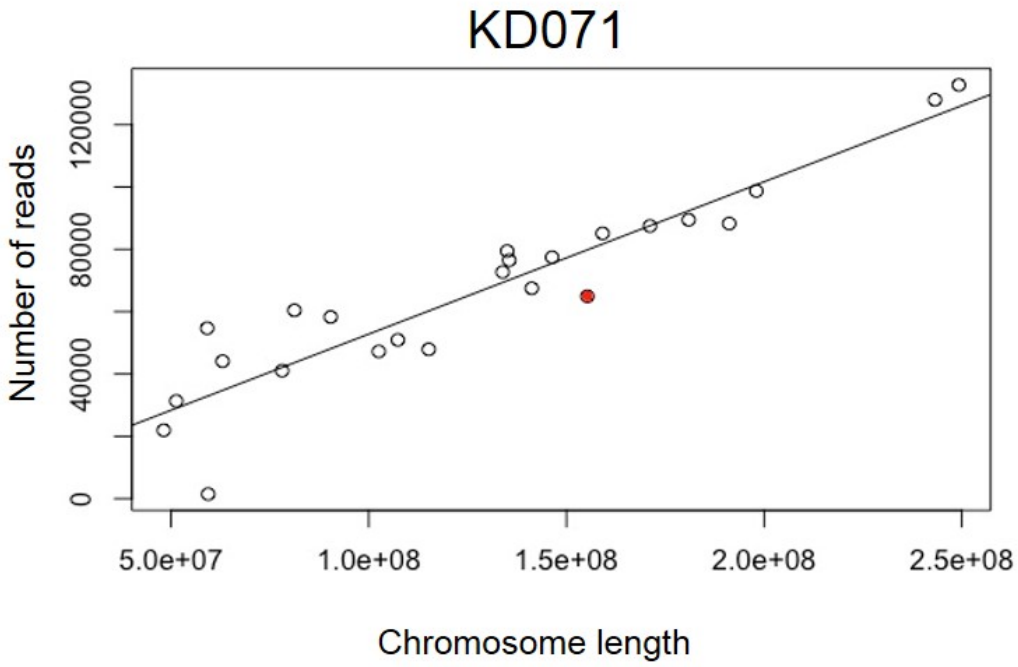

FIGURE E.6: Karyotype plot of KD071.

KD072:

$$(54,595/155.270,560)/(74,558/159.138,663) = 75.0\% \tag{E.15}$$

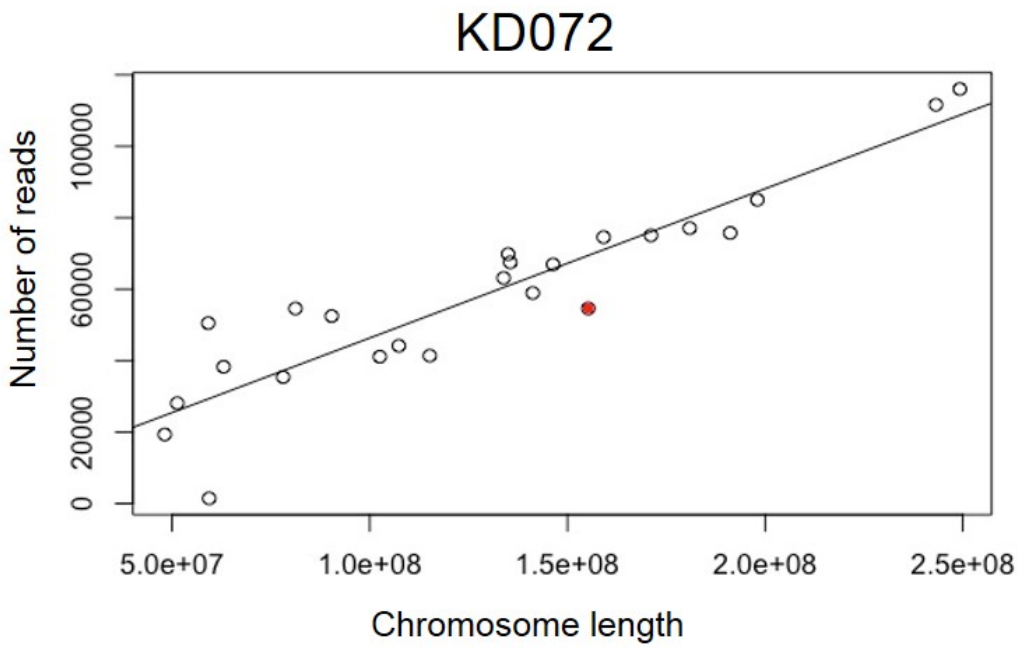

FIGURE E.7: Karyotype plot of KD072.

KD048:

$$(8,181/155.270,560)/(12,053/159.138,663) = 69.6\% \tag{E.16}$$

KD051:

$$(5,961/155.270,560)/(8,987/159.138,663) = 68.0\% \tag{E.17}$$

KD052:

$$(7,772/155.270,560)/(10,650/159.138,663) = 74.8\% \tag{E.18}$$

KD053:

$$(4,405/155.270,560)/(7,369/159.138,663) = 61.3\% \tag{E.19}$$

KD073:

$$(42,195/155.270,560)/(61,494/159.138,663) = 70.3\% \tag{E.20}$$

KD057:

$$(27,024/155.270,560)/(54,293/159.138,663) = 51.0\% \tag{E.21}$$

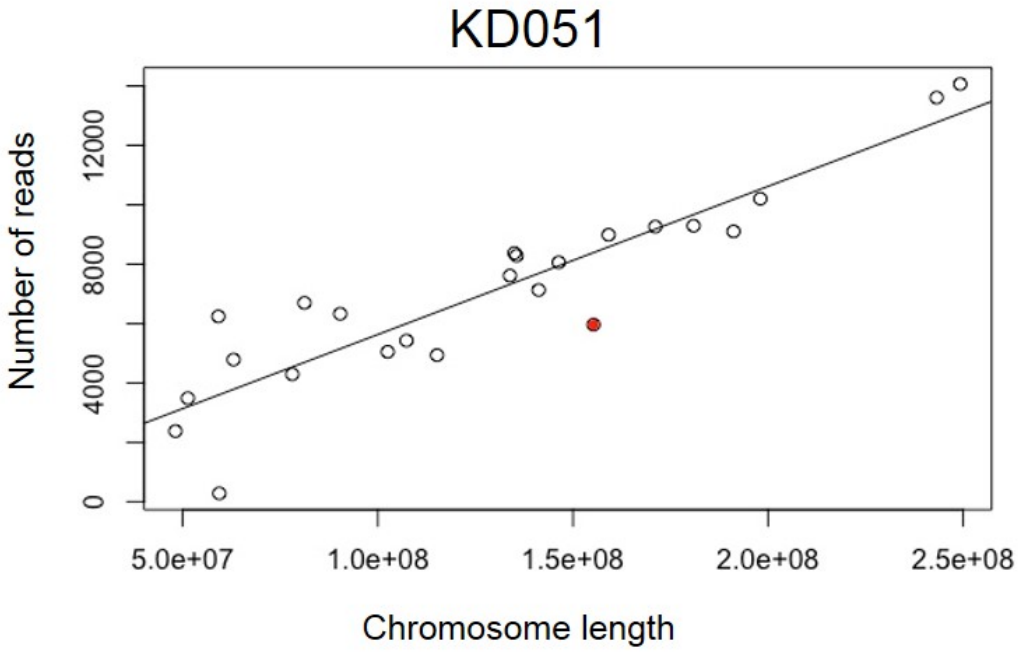

FIGURE E.8: Karyotype plot of KD051.

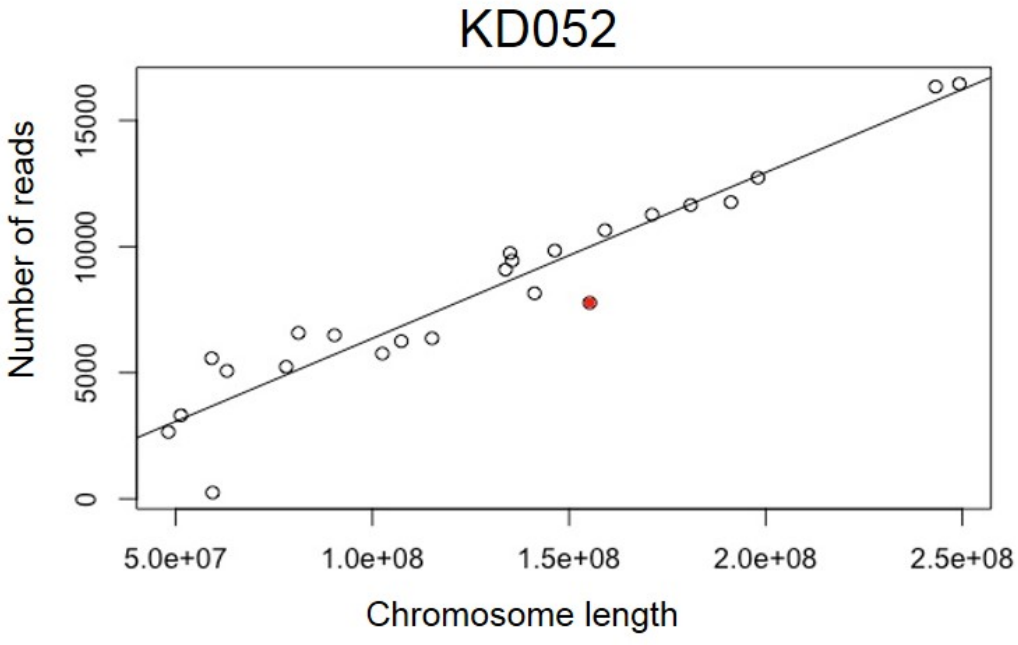

FIGURE E.9: Karyotype plot of KD052.

KD058:

$$(59,156/155.270,560)/(73,865/159.138,663) = 82.1\% \tag{E.22}$$

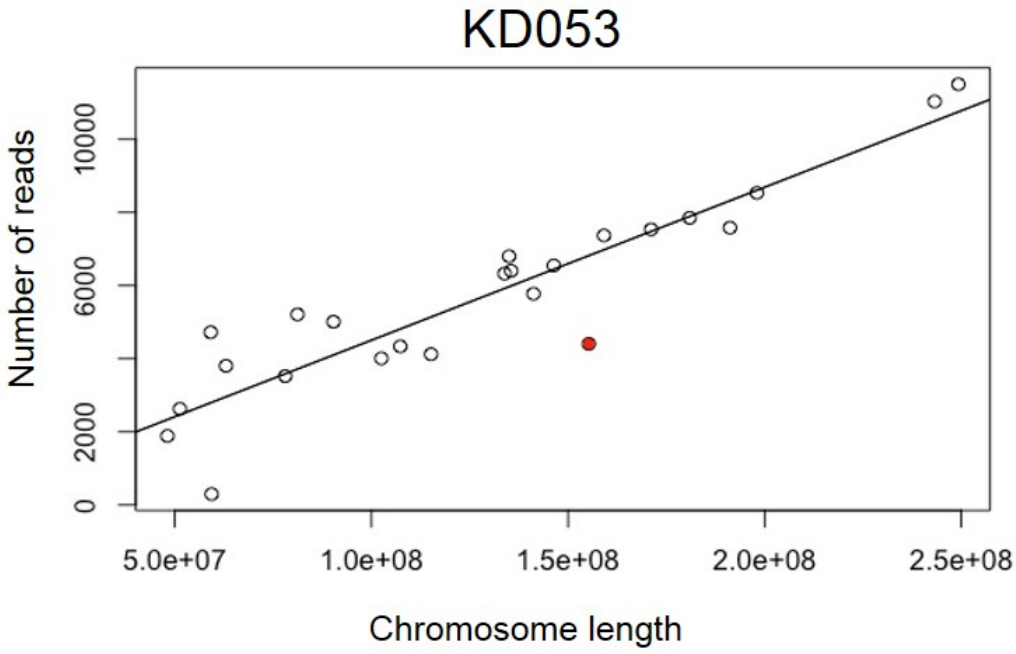

FIGURE E.10: Karyotype plot of KD053.

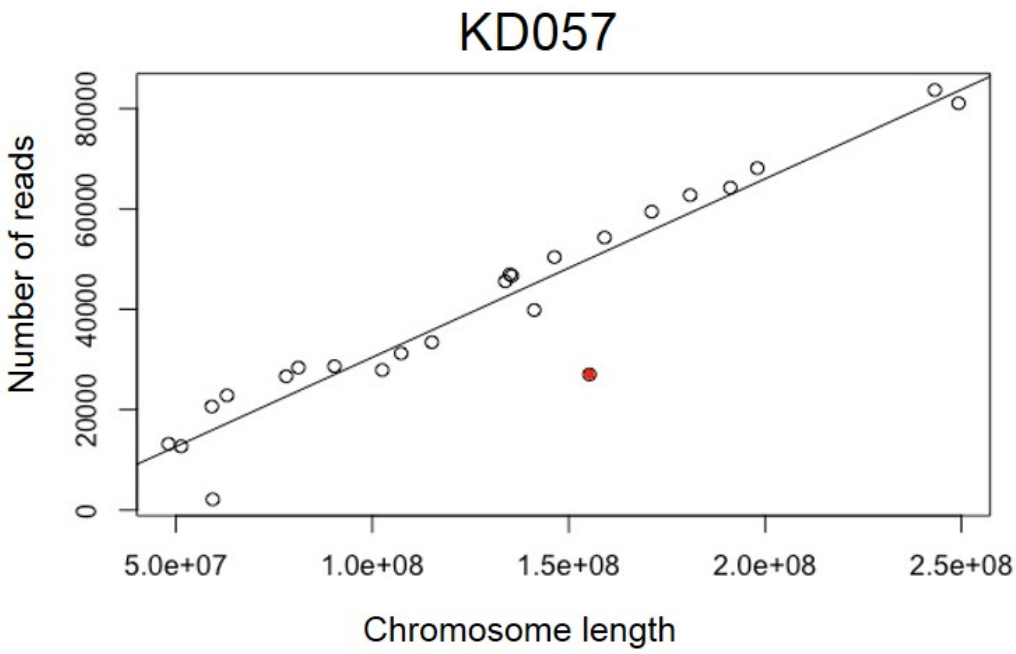

FIGURE E.11: Karyotype plot of KD057.

KD059:

$$(21,070/155.270,560)/(40,586/159.138,663) = 53.2\% \tag{E.23}$$

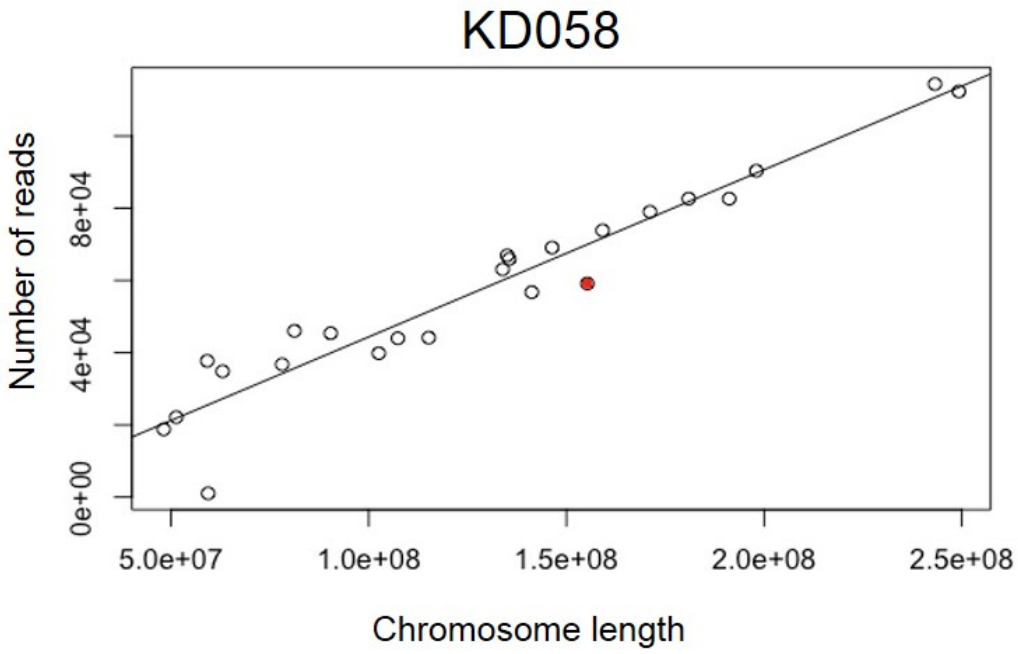

FIGURE E.12: Karyotype plot of KD058.

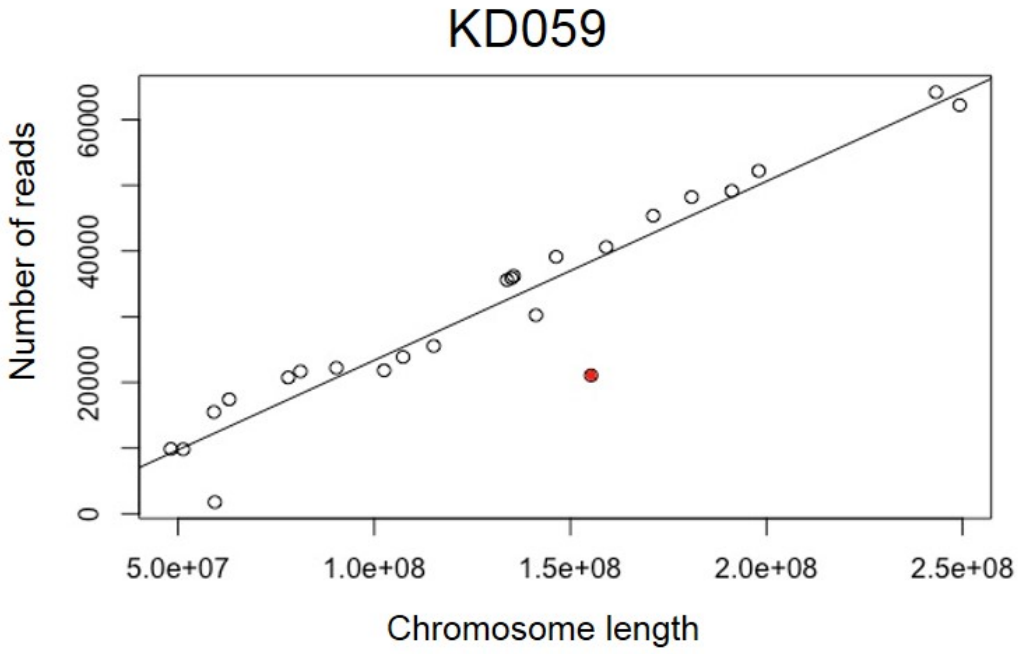

FIGURE E.13: Karyotype plot of KD059.

KD066:

$$(21,271/155.270,560)/(29,473/159.138,663) = 73.9\% \tag{E.24}$$

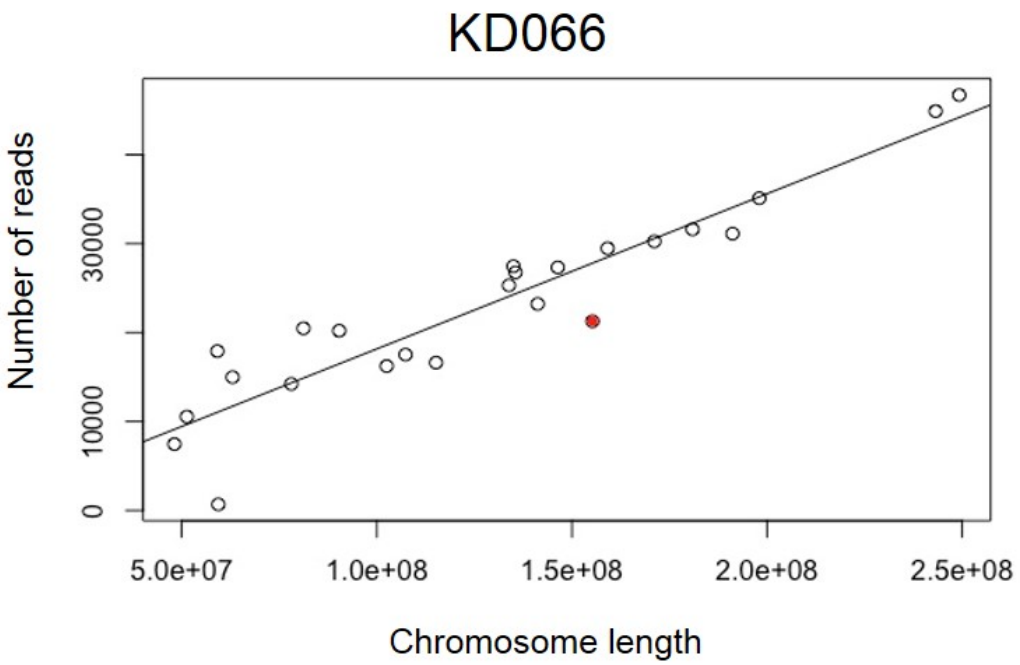

FIGURE E.14: Karyotype plot of KD066.

Low Hauxley - KD070:

$$(172,415/155.270,560)/(307,594/159.138,663) = 57.4\% \tag{E.25}$$

Figure see Chapter 3.

Appendix F

Shot-gun sequenced mitochondrial haplotypes

TABLE F.1: All retrieved shot-gun sequenced mitochondrial haplotypes and their haplogroup assignments.

Sample ID	Haplogroup	Overall rank	Polymorphisms
KD001	J1b1a1a	1.15	73G 242Y 263G 295Y 462T 489C 750G 1438G 2158Y 2706G 3010A 4216C 4769G 5460R 5463Y 6911Y 7028T 8269R 8557A 8860G 10398G 11251G 11719A 12007A 12612R 13708A 13879C 14766T 15326G 15452A 15928R 16069T 16126C 16145A 16172C 16192T 16222T 16261T
KD002	H6a1a	0.96	239C 263G 750G 1438G 3915A 4727G 4769G 8860G 9380A 11253C 15326G 16362C 16482G
KD003	T2e	0.97	73G 150T 263G 709A 750G 1438G 1888A 2706G 4216C 4769G 4917G 7028T 8697A 8860G 10463C 11251G 11719A 11812G 13368A 14233G 14766T 14905A 15326G 15452A 15607G 15928A 16126C 16153A 16294T 16296T
KD004	H1b	0.95	263G 750G 1438G 3010A 4769G 8860G 15326G 16189C 16356C
KD005	H7a1b	1.00	263G 750G 1393A 1438G 1719A 4769G 4793G 8860G 15326G 16261T
KD006	T2a1b1a	1.00	73G 263G 709A 750G 1438G 1888A 2141C 2706G 4216C 4769G 4917G 7028T 8697A 8860G 9117C 10463C 11251G 11719A 11812G 13368A 13965C 13966G 14233G 14687G 14766T 14905A 15326G 15452A 15607G 15928A 16126C 16294T 16296T 16324C 16519C
KD023	J1c1	1.00	73G 228A 263G 295T 462T 482C 489C 750G 1438G 2706G 3394C 4216C 4769G 7028T 8860G 10398G 11251G 11719A 12612G 14766T 14798C 15326G 15452A 16069T 16126C
KD024	T2b	0.99	73G 263G 709A 750G 930A 1438G 1888A 2706G 4216C 4769G 4917G 5147A 7028T 8697A 8860G 10463C 11251G 11719A 11812G 13368A 14233G 14766T 14905A 15326G 15452A 15607G 15928A 16126C 16294T 16296T

Table F.1 continued from previous page

Sample ID	Haplogroup	Overall rank	Polymorphisms
KD025	H1c	0.61	263G 477C 1438G 8860G
KD026	U5b2c	0.92	73G 150T 263G 723G 750G 1438G 1721T 2706G 3197C 4769G 7028T 7768G 8860G 9477A 11467G 11719A 12308G 12372A 13017G 13617C 13637G 14182C 14766T 15326G 16192T 16270T
KD027	U5b2c	0.93	73G 150T 263G 723G 750G 1438G 2706G 3197C 4769G 7028T 7768G 8860G 9477A 11467G 11719A 12308G 12372A 13017G 13617C 13637G 14766T 15326G 16192T 16270T
KD028	T2b	0.99	73G 263G 709A 750G 930A 1438G 1888A 2706G 4216C 4769G 4917G 5147A 7028T 8697A 8860G 10463C 11251G 11719A 11812G 13368A 14233G 14766T 14905A 15326G 15452A 15607G 15928A 16126C 16294T 16296T
KD029	H5b	1.00	263G 456T 750G 1438G 4769G 5471A 8860G 15326G 16304C
KD040	T2b4h	0.78	214G 263G 709A 750G 930A 1438G 2706G 4216C 4769G 4917G 5147A 8697A 10463C 11251G 11719A 13368A 14766T 14905A 15452A 15607G 15928A 16126C
KD041	U5a1a2a	1.03	73G 263G 750G 1438G 2706G 3197C 4769G 5319G 8860G 9477A 11467G 12308G 12372A 13617C 14766T 14793G 15326G 16192T! 16256T 16270T 16399G
KD042	U5a1b1a	0.96	73G 263G 750G 1438G 2706G 3197C 4553C 4769G 7028T 8860G 9477A 9667G 11719A 12308G 12372A 13617C 14766T 14793G 15326G 16291T 16399G
KD043	H1b	0.95	263G 750G 1438G 3010A 4769G 8860G 15326G 16189C 16356C
KD044	U5b2a3	1.01	73G 150T 263G 750G 1438G 1721T 2706G 3197C 4732G 4769G 7028T 7768G 8860G 9477A 10031C 11467G 11719A 12308G 12372A 13617C 13637G 14182C 14766T 16189C 16192T 16270T

Table F.1 continued from previous page

Sample ID	Haplogroup	Overall rank	Polymorphisms
KD045	J1c2a	0.98	73G 185A 188G 228A 263G 295T 462T 489C 750G 1438G 2706G 3010A 4216C 4769G 6293C 7028T 8860G 10398G 11251G 11719A 12612G 13708A 14766T 14798C 15326G 15452A 16069T 16126C
KD046	T2a1b1a	1.01	73G 263G 750G 1438G 1888A 2141C 2706G 4216C 4769G 4917G 7028T 8697A 8860G 9117C 10463C 11251G 11719A 11812G 13368A 13965C 13966G 14233G 14687G 14766T 14905A 15326G 15452A 15607G 15928A 16126C 16294T 16296T 16324C
KD047	H39	0.90	263G 750G 1438G 4769G 8860G 15326G 16299G
KD047	J1c3	0.94	73G 185A 228A 263G 295T 462T 489C 750G 2706G 3010A 11719A 12612G 13708A 13934T 14766T 14798C 15326G 15452A 16069T 16126C
KD048	H39	0.77	750G 1438G 4769G 8860G 15326G 16299G
KD049	H39	0.93	263G 750G 1438G 4769G 8860G 15326G 16299G
KD050	H39	0.93	263G 750G 1438G 4769G 8860G 15326G 16299G
KD052	K1a29a	1.08	750G 1438G 4769G 7028T 8860G 12372A 14798C
KD053	J1c2a	0.99	73G 185A 188G 263G 295T 462T 489C 750G 1438G 2706G 3010A 4216C 4769G 6293C 7028T 8860G 10398G 11251G 11719A 12612G 13708A 14766T 14798C 15326G 15452A 16069T 16126C
KD057	H1n1	1.00	146C 263G 750G 1438G 2098A 3010A 4769G 8860G 15326G 15742A
KD058	K1a3a	0.93	73G 263G 497T 750G 1189C 1438G 1811G 2706G 3480G 4769G 7559G 8860G 9055A 10398G 10550G 11299C 12372A 14766T 15326G 16224C
KD059	T2b21	1.03	152C 263G 709A 750G 930A 1438G 1888A 2706G 4216C 4769G 4917G 7028T 8697A

Table F.1 continued from previous page

Sample ID	Haplogroup	Overall rank	Polymorphisms
			8860G 10463C 11719A 11812G 13368A 14233G 14766T 14836G 14905A 15326G 15452A 15607G 15928A 16126C 16294T 16304C
KD060	H1n1	1.01	263G 750G 1438G 2098A 3010A 4769G 8860G 15326G 15742A
KD061	K1c2	1.00	73G 146C 152C 263G 750G 1189C 1438G 1811G 2706G 3480G 4769G 7028T 8860G 9006G 9055A 10398G 10550G 11299C 11467G 11719A 12308G 12372A 14002G 14040A 14167T 14766T 14798C 15326G 16224C 16311C 16320T
KD062	U5b2a3	1.04	73G 150T 263G 750G 1438G 1721T 2706G 3197C 4769G 7028T 7768G 8860G 10031C 11467G 11719A 12308G 12372A 13617C 13637G 14182C 14766T 15326G 16189C 16192T 16270T
KD063	H58a	1.00	263G 750G 1438G 4769G 8860G 15326G 15466A 15721C
KD064	T2b21	0.98	73G 152C 263G 709A 750G 930A 1438G 1888A 2706G 4216C 4769G 4917G 5147A 7028T 8697A 8860G 10463C 11251G 11719A 11812G 13368A 14233G 14766T 14836G 14905A 15326G 15452A 15607G 15928A 16126C 16294T 16304C
KD065	H39	0.98	263G 750G 1438G 4769G 8860G 15326G
KD066	T2a1b1a	1.11	263G 709A 750G 1438G 1888A 2706G 4769G 4917G 7028T 8697A 8860G 9117C 11719A 13368A 13965C 13966G 14233G 14766T 15326G
KD067	H+195	0.93	195C 263G 750G 1438G 4769G 8860G 15326G
KD069	H7a1b	0.66	263G 750G 1393A 1719A 4769G 8860G
KD070	T2e1a	1.08	41T 263G 750G 1438G 2706G 4216C 4769G 4917G 7028T 8697A 8860G 10463C 11251G 11719A 11812G 13368A 14233G 14766T 14905A 15326G 15452A 15928A 16126C 16294T

Table F.1 continued from previous page

Sample ID	Haplogroup	Overall rank	Polymorphisms
KD071	H1b1+16362	0.78	1438G 3796G 4769G 8860G 16189C 16356C 16362C
KD072	H1b1+16362	0.91	263G 750G 3796G 4769G 8860G 15326G 16356C 16362C
KD073	J1b1a1	0.99	73G 242T 263G 295T 489C 750G 1438G 2158C 2706G 4216C 4769G 7028T 8269A 8557A 8860G 10398G 11251G 12612G 14766T 15326G 15452A 16069T 16126C 16222T 16261T
Low Hauxley	T2e1a	1.02	41T 73G 150T 263G 750G 1438G 1888A 2706G 4216C 4769G 4917G 7028T 8697A 8860G 10463C 11251G 11719A 11812G 13368A 14233G 14766T 14905A 15326G 15452A 15607G 15928A 16126C 16294T
Oakridge	H3g1a	0.9386	152C 263G 750G 1438G 3992T 4418C 4769G 6776C 8860G 10754C 15326G 15377G

Appendix G

Y-chromosome haplogroup identification

The complete lists of all derived mutations of Y-chromosome predictive markers used in Yleaf are given in the following tables. Mutations at the first or last three bases of a sequence read were excluded, as well as mutations with a quality score of less than 30.

G.1 Strathglebe

TABLE G.1: Yleaf output of KD026, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
6847637	L1098	A0-T	C->A	1	A
22957136	L1084	A1	G->A	1	A
4957608	V174	A1	A->G	1	G
8466995	L1112	A1	G->A	1	A
14122034	M9086	BT	T->C	1	C
14786996	M9116	BT	A->T	1	T
15569966	M9143	BT	T->C	1	C
17670179	M9226	BT	C->T	1	T
17670193	M9227	BT	G->A	1	A

Table G.1 continued from previous page

19416000	M9293	BT	C->G	1	G
21470547	M9322	BT	T->G	1	G
22817840	M9365	BT	G->C	1	C
24377975	M9406	BT	G->T	1	T
28772048	Y10888	BT	C->T	1	T
6719048	M8961	BT	T->G	1	G
6779907	M8967	BT	G->A	1	A
6859691	M8969	BT	C->T	1	T
6908261	V29	BT	A->G	1	G
7711024	M8995	BT	A->G	1	G
7735760	M8997	BT	C->G	1	G
8330559	M9019	BT	T->C	1	C
9381562	Y8311	BT	G->A	1	A
14074463	Y1525	CT	C->T	1	T
17909505	M5738	CT	G->A	1	A
18961511	CTS9555	CT	G->C	1	C
19039750	M5762	CT	G->C	1	C
21413635	M5782	CT	A->C	1	C
21683734	M5788	CT	C->T	1	T
22529475	Y1567	CT	C->G	1	G
22639802	PF1205	CT	C->T	1	T
6975295	L1462	CT	C->G	1	G
17595603	F3964	D1a1a1a1a2	T->C	1	C
16368310	L468	F	A->G	1	G

Table G.1 continued from previous page

17256018	P166	F	C->T	1	T
18765649	F2837	F	T->C	1	C
2756471	M3637	F	A->T	1	T
8442341	F1209	F	A->G	1	G
15023364	M258	I	T->C	1	C
15389836	CTS4088	I	T->C	1	C
15937959	CTS4982	I	C->T	1	T
18394743	L751	I	A->G	1	G
14031334	M429	IJ	T->A	1	A
16422613	PF3542	IJ	C->T	1	T
17040447	F2366	IJ	C->A	1	A
19166861	F4281	IJ	T->C	1	C
22909101	F3368	IJ	C->T	1	T
16041510	Y30471	O2a2b1a2a1d1	G->A	1	A
7256448	Y1178	Q1b2	C->T	1	T
10008791	L150.1	R1b1a1a2	C->T	1	T

G.2 Low Hauxley

G.3 Links of Noltland

TABLE G.2: Yleaf output of the Low Hauxley individual filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
6788449	Z11900	A1b	T->C	1	C
8319368	Z5287	B3∼	G->A	1	A
13840089	Z40402	BT	T->C	1	C
16542464	M9174	BT	A->C	1	C
19445252	M9295	BT	A->C	1	C
21473970	M9323	BT	A->G	1	G
21490191	M9326	BT	A->G	1	G
22521638	Z17385	BT	C->T	1	T
9930675	Z40372	BT	C->T	2	T
15344987	M5675	CT	T->C	1	C
7008958	M5589	CT	T->C	1	C
7597185	M5603	CT	G->A	1	A
7646619	M5605	CT	G->C	1	C
8183439	M5627	CT	G->A	1	A
19040273	Z22098	E1b1a1b	C->T	1	T
22156498	M3752	F	C->A	1	A
8521791	Z39592	G2a2b2a1a1b1a1a2b1a1e∼	T->A	1	A
14989342	FGC12147	G2a2b2a1a1c2a	A->C	1	C
21375113	Y21735	N1a1a1a1a1a2a1a2i1∼	C->T	1	T
9093383	M1195	P	G->C	1	C
17723850	CTS7880	R	C->T	1	T
16426937	CTS5676	R1b1	C->G	1	G
19504659	L820	R1b1a	T->A	1	A
28733101	L389	R1b1a1	C->G	1	G
28771116	CTS12972	R1b1a1a2	C->G	1	G

TABLE G.3: Yleaf output of KD006, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
15425676	L1123	A0-T	C->T	1	T
18147303	L1135	A0-T	C->A	1	A
21593345	L1143	A0-T	A->G	1	G
15442970	M11301	A1b	T->C	1	T
22111732	Z17899	A1b	T->C	1	C
16002481	Z38086	B2b1a2a∼	C->T	1	T
13202478	Z40398	BT	G->A	1	A
14980170	M9121	BT	C->T	1	T
15256839	M9133	BT	A->G	1	G
16538095	M9173	BT	C->T	1	T
17353147	L1220	BT	A->G	1	G

Table G.3 continued from previous page

Position	Marker name	Haplogroup	Mutation	Reads	Called base
17840394	M9235	BT	C->T	1	T
18152393	M9251	BT	C->T	1	T
18628317	M9260	BT	G->A	1	A
19445252	M9295	BT	A->C	1	C
21259569	M9311	BT	C->T	1	T
22062727	M9347	BT	C->T	1	T
22180636	M9352	BT	C->A	2	A
23757465	M9397	BT	G->C	1	C
23991003	M9405	BT	G->A	1	A
2899711	M8955	BT	A->G	1	G
7037901	M8976	BT	G->C	1	C
7269376	M8980	BT	C->T	1	T
7944491	M9008	BT	C->A	1	A
16346371	CTS5532	CT	C->T	1	T
17517587	CTS7517	CT	T->A	1	A
17858228	CTS8166	CT	A->G	1	G
17894575	CTS8243	CT	C->T	1	T
18948988	M5756	CT	T->C	1	C
22067044	M5794	CT	A->T	1	T
23407446	CTS11991	CT	T->G	1	G
28789895	Y1580	CT	A->T	1	T
7707484	M5607	CT	C->A	1	A
8543804	M5633	CT	G->A	1	A
13642737	Z14835	D1b1c1a1	G->A	1	A
14149772	F1704	F	C->T	1	T
14199284	P138	F	T->C	1	C
14334396	F1767	F	G->T	1	T
18178091	P161	F	G->A	1	A
21147336	F3136	F	T->C	1	C
23040647	P136	F	T->G	1	G
7350493	Z31447	G2a2a1a2a1b1a2∼	G->A	1	A
14974451	L1197	I	C->T	1	T
15089989	CTS3641	I	T->C	1	C
15595624	CTS4340	I	G->A	1	A
15799074	CTS4752	I	C->T	1	T

Table G.3 continued from previous page

Position	Marker name	Haplogroup	Mutation	Reads	Called base
16039881	CTS5150	I	C->T	1	T
16397716	CTS5622	I	C->A	1	A
17548890	CTS7593	I	G->A	1	A
21452125	PF3803	I	A->G	1	G
7853028	PF3645	I	C->A	1	A
14539161	A13926	I1a2b1a∼	C->T	1	T
17359886	CTS7218	I2a1b	A->C	2	C
17893806	CTS8239	I2a1b	A->G	1	G
14144593	P129	IJ	A->G	1	G
8558969	P130	IJ	A->T	1	T
16208786	CTS5323	O2a2b1a2a1a1b	G->A	1	A
16242725	Y30494	O2a2b1a2a1d1a	G->A	1	A
7256448	Y1178	Q1b2	C->T	1	T

TABLE G.4: Yleaf output of KD045, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
21288562	L1142	A0-T	C->T	1	T
23572106	Z17900	A1b	G->A	1	A
14762104	M9114	BT	G->C	1	C
18975603	M9271	BT	C->T	1	T
23049011	M9374	BT	T->G	1	G
6691986	M8959	BT	G->A	1	A
8635293	M9032	BT	C->A	1	A
23090404	M5809	CT	G->A	1	A
7740081	M5610	CT	G->A	1	A
8526565	M5632	CT	G->A	1	A
21603164	M1160	P	T->G	1	G

TABLE G.9: Yleaf output of KD060, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
21657017	FGC27483	A00	G->A	1	A
15442970	M11301	A1b	T->C	1	T

Table G.9 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
20820626	Z17898	A1b	A->G	1	A
23551003	L989	A1b	T->A	1	A
4501479	Z11901	A1b	A->T	1	A
14013228	M9075	BT	T->C	1	C
18069115	M9246	BT	G->A	1	A
19149592	M9280	BT	C->A	1	A
19325165	M9290	BT	C->T	1	T
22625732	M9356	BT	C->T	1	T
2790584	V202	BT	G->C	1	C
28746408	Z17390	BT	G->C	1	C
7221998	M8979	BT	C->A	1	A
7531097	M8988	BT	T->C	1	C
9930675	Z40372	BT	C->T	1	T
13213798	Z17710	CT	G->T	2	T
15097073	CTS3662	CT	G->A	1	A
17589395	M5723	CT	G->A	1	A
17639829	CTS7741	CT	C->T	1	T
17894575	CTS8243	CT	C->T	1	T
23111277	M5813	CT	C->A	1	A
2733618	M8948	CT	C->A	1	A
16368116	CTS5568	E1b1a1a1a1c1a1a3d4d∼	G->A	1	A
16839641	P316	F	A->T	1	T
17762668	F2587	F	T->C	1	C
19113388	FGC12299	G2b1b2a1c∼	G->A	1	A
13835003	FGC2415	I	T->C	1	C
14286853	CTS2387	I	T->C	1	C
15479899	CTS4209	I	T->A	1	A
17497181	CTS7469	I	C->A	1	A
8262092	FGC2413	I	C->T	1	T
9900057	FGC2411	I	A->G	1	G
16422613	PF3542	IJ	C->T	1	T
7694266	FGC1564	IJ	G->A	1	A
7702973	M2683	IJK	T->A	1	A
8771750	F3694	N	G->A	1	A
17617938	Y691	Q1a1a1	G->A	1	A

TABLE G.5: Yleaf output of KD047, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
14467641	Z11906	A1b	C->G	1	G
15112480	M9128	BT	C->T	1	T
19235631	M9285	BT	C->T	1	T
19555949	Z17372	BT	T->G	1	G
21646196	M9336	BT	G->A	1	A
22625732	M9356	BT	C->T	1	T
23201899	Z40411	BT	A->C	1	C
21159234	FGC22190.2^	C1b1a1a1a1a1a1b~	C->A	1	A
23040647	P136	F	T->G	1	G
7321418	CTS1301	I	C->T	1	T
22974926	Z124	I2a1a1a1~	T->C	1	C
18647146	M1240	P	A->C	1	C

TABLE G.10: Yleaf output of KD061, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
17947672	V168	A1	G->A	1	A
16243116	Z11919	A1b	C->T	1	T
14401169	M9098	BT	T->C	1	C
15261255	Z40392	BT	C->T	1	T
17041558	M9193	BT	G->A	1	A
17147159	M9197	BT	C->T	1	T
17571812	M9220	BT	T->A	1	A
18005126	Y8319	BT	T->A	1	A
18585407	M9255	BT	T->G	1	G
18643546	M9262	BT	T->G	1	G
19555949	Z17372	BT	T->G	1	G
22798764	M9362	BT	G->A	1	A
22903616	M9369	BT	T->C	1	C
23558260	M9393	BT	G->A	1	A
2756519	V187	BT	T->C	1	C
2906950	M8956	BT	C->T	1	T
7890655	M9003	BT	C->G	1	G
13867088	Z12163	C2	A->T	1	T
16863259	CTS6376	CF	C->G	1	G
14079528	L957	CT	C->T	1	T

Table G.10 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
14516781	M5661	CT	T->C	1	C
14943290	CTS3460	CT	C->T	1	T
15792531	CTS4740	CT	G->A	1	A
16984364	M5711	CT	G->T	1	T
17301733	M5716	CT	T->C	1	C
18905190	CTS9458	CT	G->A	1	A
19407727	M5769	CT	C->G	1	G
21258894	M5776	CT	T->C	1	C
21349119	M5780	CT	G->T	1	T
7153375	M5591	CT	C->T	1	T
7256433	CTS1181	CT	A->T	1	T
7796647	M5613	CT	C->T	1	T
15651438	CTS4443	F	A->T	1	T
16178042	F2075	F	A->T	1	T
18097251	P159	F	C->A	1	A
19205814	L470	F	A->T	1	T
8131538	F3692	F	G->T	1	T
8572150	F1320	F	G->A	1	A
8602415	P146	F	C->T	1	T
14373487	FGC58151	G2a2a1a1a	G->A	1	A
22228628	L672	I2a1a1a	T->A	1	A
7702973	M2683	IJK	T->A	1	A
14864191	CTS3358	P	T->C	1	C
15111148	CTS3697	P	G->C	1	C
17502468	CTS7481	P	C->T	1	T
19246561	CTS10085	P	A->G	1	G
21298239	M1256	P	C->G	1	G
22003770	M1264	P	G->A	1	A
2803717	CTS196	P	C->T	1	T
9921064	PF5481	P	G->A	1	A
17716251	P286	R1	C->T	1	T
8633545	P245	R1	T->C	1	C
21801722	PF6506	R1b1a1a	G->A	1	A
15732786	CTS4608	R1b1a1a2	T->C	1	C
18095336	CTS8591	R1b1a1a2	A->C	1	C

Table G.10 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
22796697	CTS10834	R1b1a1a2	T->C	1	C
8667179	PF6435	R1b1a1a2	A->G	1	G

TABLE G.11: Yleaf output of KD064, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
21593345	L1143	A0-T	A->G	1	G
22191266	L1155	A0-T	G->C	1	C
2790726	L1085	A0-T	T->C	1	C
16070726	FGC26595	A00	G->A	1	A
17645335	V250	A1	A->G	1	G
8087059	L986	A1	G->A	1	A
23125972	Z11907	A1b	A->C	1	C
23446521	Z11915	A1b	A->G	1	G
23572106	Z17900	A1b	G->A	1	A
8370730	Z11904	A1b	A->G	1	A
10008803	L418	BT	C->G	1	G
13592206	Z40378	BT	C->A	1	A
13665248	Z40379	BT	G->A	1	A
13826087	Y8488	BT	G->A	1	A
14461702	M9099	BT	G->A	1	A
14720802	M9110	BT	G->A	1	A
15068525	M9125	BT	A->T	1	T
15112480	M9128	BT	C->T	1	T
16676762	M9178	BT	T->C	1	C
17147159	M9197	BT	C->T	1	T
17287394	V64	BT	T->A	1	A
17966904	M9242	BT	C->A	1	A
18958429	M9269	BT	C->T	1	T
19149592	M9280	BT	C->A	1	A
19555949	Z17372	BT	T->G	1	G
23005081	M9372	BT	G->T	1	T
23117154	L604	BT	C->A	1	A
23201899	Z40411	BT	A->C	1	C

Table G.11 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
23237493	Z12129	BT	T->A	1	A
28746408	Z17390	BT	G->C	1	C
6936412	M8971	BT	G->C	1	C
7763132	M8999	BT	C->T	1	T
7815370	M9000	BT	T->A	1	A
7907933	M9006	BT	G->A	1	A
8057835	M9011	BT	G->A	1	A
8330559	M9019	BT	T->C	1	C
8511682	M9026	BT	T->C	1	C
9066684	M9045	BT	C->T	1	T
16863259	CTS6376	CF	C->G	1	G
10049308	Z17708	CT	A->G	1	G
14079528	L957	CT	C->T	1	T
14149520	M5652	CT	T->A	1	A
14172842	CTS2077	CT	T->G	1	G
14943290	CTS3460	CT	C->T	1	T
15344987	M5675	CT	T->C	1	C
15792531	CTS4740	CT	G->A	1	A
16822011	CTS6327	CT	A->G	2	G
18106506	M5745	CT	G->A	1	A
21413635	M5782	CT	A->C	1	C
23106149	Z40572	CT	C->T	1	T
23111277	M5813	CT	C->A	1	A
7210130	M5593	CT	G->A	1	A
7597185	M5603	CT	G->A	1	A
7738840	M5609	CT	T->G	1	G
7778691	M5611	CT	C->T	1	T
9966488	PF1431	DE	G->A	1	A
17398598	P14	F	C->T	1	T
19205814	L470	F	A->T	1	T
22156498	M3752	F	C->A	1	A
8424089	P145	F	G->A	1	A
22937244	Z40494	G2a2b2a3a	G->A	1	A
17421318	Z19055	H2a1a∼	C->T	1	T
14484379	P38	I	A->C	1	C

Table G.11 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
15615533	L772	I	C->A	1	A
16751000	CTS6231	I	C->T	1	T
17497181	CTS7469	I	C->A	1	A
17525137	CTS7540	I	A->G	1	G
21525069	PF3806	I	G->A	1	A
22115103	YSC0000272	I	G->A	1	A
22200336	PF3822	I	G->A	1	A
22573702	PF3837	I	G->A	1	A
2707072	CTS70	I	C->T	1	T
6943522	CTS674	I	C->T	1	T
14074218	CTS1802	I2a1b	A->T	1	T
15055079	CTS3597	IJ	A->T	1	T
16244914	CTS5387	IJ	G->A	1	A
18633084	F2794	IJ	C->A	1	A
7173143	L16	IJK	G->A	2	A
13863007	F4906	Q1a1b1∼	T->A	1	A

G.4 Knowe of Skea

TABLE G.12: Yleaf output of KD004, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
14231291	L1118	A0-T	T->C	1	C
15426344	L1124	A0-T	T->C	1	C
16428812	L1128	A0-T	C->T	1	T
16596846	L1129	A0-T	T->C	1	C
16661010	L1130	A0-T	T->G	1	G
18147303	L1135	A0-T	C->A	1	A
7590048	L1105	A0-T	C->T	1	T
16773792	L1004	A1	T->C	1	C
7651556	V238	A1	G->T	1	T
14774366	Z17894	A1b	C->T	1	T
15130659	Z11910	A1b	A->G	1	G

Table G.12 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
19390790	Z11917	A1b	A->C	1	A
23125773	Z11892	A1b	C->A	1	A
23551003	L989	A1b	T->A	1	A
13668726	Z40380	BT	T->C	1	C
14386666	M9097	BT	G->A	1	A
14723798	M9111	BT	T->C	1	C
14726915	M9113	BT	C->T	1	T
14980170	M9121	BT	C->T	1	T
15022370	M251	BT	G->A	1	A
15256839	M9133	BT	A->G	1	G
15569966	M9143	BT	T->C	1	C
16016838	M9159	BT	G->A	2	A
16913677	M9187	BT	T->C	1	C
17063750	M9195	BT	T->A	1	A
17453361	M9216	BT	A->G	1	G
17728330	M9230	BT	T->C	1	C
17753892	M9231	BT	G->A	1	A
18067088	M9245	BT	T->A	1	A
18179224	M9253	BT	G->A	1	A
18242832	M9254	BT	G->A	1	A
18262817	Y8320	BT	G->A	1	A
18269175	Z40405	BT	A->G	1	G
18592027	M9258	BT	G->A	1	A
18633816	M9261	BT	C->T	2	T
18917853	M9267	BT	C->T	1	T
19391922	M9292	BT	T->C	1	C
19445252	M9295	BT	A->C	1	C
21238206	Z40406	BT	C->T	1	T
21391240	M9321	BT	G->A	1	A
21866840	M42	BT	A->T	1	T
23005081	M9372	BT	G->T	1	T
23237493	Z12129	BT	T->A	1	A
23244049	M9379	BT	G->C	1	C
23558260	M9393	BT	G->A	1	A

Table G.12 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
28602059	M9417	BT	A->G	1	G
28787653	Z40384	BT	G->C	1	C
2899711	M8955	BT	A->G	1	G
6634305	M8957	BT	A->G	1	G
6646548	M8958	BT	T->C	1	C
7269376	M8980	BT	C->T	1	T
7711024	M8995	BT	A->G	1	G
8021764	M9010	BT	G->A	1	A
8180027	Z40388	BT	A->G	1	G
8226191	M9017	BT	C->A	1	A
8330559	M9019	BT	T->C	1	C
8869241	M9041	BT	G->A	1	A
9381562	Y8311	BT	G->A	2	A
9869354	M11760	BT	C->A	1	A
7430665	Z32459	C1b1a2a	C->T	1	T
16863259	CTS6376	CF	C->G	1	G
13594339	Y1791	CT	C->T	1	T
14439817	CTS2711	CT	G->A	2	A
14516781	M5661	CT	T->C	1	C
14813991	M168	CT	C->T	1	T
15344987	M5675	CT	T->C	1	C
15932327	Y1528	CT	A->G	1	G
16131945	Y1531	CT	C->G	1	G
16203547	CTS5318	CT	G->T	1	T
16325663	M5692	CT	A->C	1	C
16770207	CTS6252	CT	G->A	1	A
17174326	CTS6890	CT	G->C	1	C
17639829	CTS7741	CT	C->T	1	T
17751557	CTS7936	CT	A->T	1	T
17853378	M5734	CT	C->T	1	T
18961511	CTS9555	CT	G->C	1	C
18974195	M5760	CT	C->T	1	T
19169483	PF970	CT	T->C	3	C
19421388	CTS10362	CT	G->A	3	A

Table G.12 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
19503081	CTS10512	CT	C->A	1	A
21683734	M5788	CT	C->T	1	T
22594299	M5801	CT	C->T	1	T
23081753	CTS11358	CT	A->G	1	G
23466468	M5819	CT	C->T	2	T
28790215	PF1337	CT	G->C	1	C
7233061	Z40571	CT	G->A	1	A
7358315	M5599	CT	A->G	1	G
7740081	M5610	CT	G->A	1	A
8138589	M5625	CT	A->G	1	G
16413930	F2160	D1a1a	G->A	1	A
23898115	S25934	E1b1a1a1a1c1a1a3a1d1b1b1a3∼	C->A	1	A
14334396	F1767	F	G->T	1	T
16690780	CTS6135	F	T->C	1	C
17256018	P166	F	C->T	1	T
18178091	P161	F	G->A	1	A
18765649	F2837	F	T->C	1	C
22108101	M3751	F	G->A	1	A
22673903	F3335	F	T->G	1	G
8227605	M3652	F	A->G	1	G
8623281	Z44138	G2a2b2a1a1b1a1a2a5c∼	G->A	1	A
21135905	Z31211	G2a2b2a1a1c2b1a∼	G->A	1	A
14852708	Z31213	G2a2b2a1a1c2b1b∼	C->A	1	A
13862754	Z14049	H1a2a	G->A	1	A
23112734	S2292	I1a2b1b∼	G->A	1	A
21571895	M2696	IJK	G->A	1	A
13597365	PF4530	J	C->T	1	T
15472863	P131	K∼	C->T	1	T
20837553	P128	K∼	C->T	1	T
23550924	M526	K2	A->C	2	C
8640485	F1369	O2a2b1a1a1a2a1a1∼	C->T	1	T
13220892	F1660	P	C->T	1	T
17559652	CTS7604	P	T->C	1	C
17726250	CTS7886	P	T->G	1	G

Table G.12 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
17954035	CTS8356	P	T->C	1	C
18647146	M1240	P	A->C	1	C
19144375	F524	P	C->T	1	T
19463273	CTS10454	P	T->C	1	C
21804435	M1261	P	C->T	1	T
21811108	L741	P	G->T	1	T
22164441	M1267	P	G->A	2	A
2686727	F4	P	G->A	2	A
28601975	CTS12524	P	G->A	1	A
28612323	L471	P	A->G	1	G
6931040	PF5858	P	C->G	1	G
8134704	F115	P	T->A	1	A
17881230	P239	P1∼	G->C	1	C
22094491	P283	P1∼	C->T	1	T
8334875	P237	P1∼	A->G	1	G
8845380	P226	P1∼	C->T	1	T
14830200	F1843	Q1a1a1	G->A	1	A
22917800	Y6852	Q1a2b1a1∼	G->A	1	A
16316263	F2125	Q1a2b2∼	G->A	1	A
15594523	F295	R	A->G	1	G
16856357	F370	R	T->C	2	C
17285993	P224	R	C->T	1	T
21219443	M760	R	A->G	1	G
21409706	P227	R	G->C	1	C
23202551	CTS11647	R	C->G	1	G
8050994	P229	R	G->C	1	C
14120054	CTS1913	R1	A->T	2	T
14556851	CTS2908	R1	C->T	1	T
16742224	L875	R1	A->G	1	G
7771131	P238	R1	G->A	1	A
15080010	CTS3625	R1b1	C->T	1	T
18914441	L278	R1b1	C->T	1	T
7081561	CTS910	R1b1	C->T	1	T
9170545	M415	R1b1	C->A	1	A

Table G.12 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
14637352	CTS3063	R1b1a	T->C	1	C
15510064	CTS4244	R1b1a	T->G	1	G
18407611	FGC35	R1b1a	C->T	2	T
22889018	L754	R1b1a	G->A	1	A
16376495	CTS5577	R1b1a1a	A->C	2	C
23452965	PF6524	R1b1a1a	T->C	1	C
18137831	CTS8665	R1b1a1a2	T->C	1	C
18180446	L500	R1b1a1a2	C->A	1	A
22739367	M269	R1b1a1a2	T->C	2	C
23124367	CTS11468	R1b1a1a2	G->T	1	T
24444622	L1351	R1b1a1a2	C->T	1	T
7220727	F69	R1b1a1a2	A->G	1	G
7762947	PF6425	R1b1a1a2	T->C	1	C
8411202	PF6434	R1b1a1a2	A->G	1	G
8826595	PF6436	R1b1a1a2	C->T	1	T
22157311	P312	R1b1a1a2a1a2	C->A	1	A
22200784	S245	R1b1a1a2a1a2c	C->G	1	G
20832490	S5191	R1b1a1a2a1a2c1a2	T->G	1	G

TABLE G.13: Yleaf output of KD043, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
22191266	L1155	A0-T	G->C	1	C
7704051	FGC25549	A00	G->A	1	A
17716223	L1005	A1	G->A	1	A
7658712	V241	A1	C->T	1	T
22111732	Z17899	A1b	T->C	1	C
23596262	Z11896	A1b	C->G	1	G
19539500	Z38160	B2b1a2a∼	G->A	1	A
7183235	Z5222	B3∼	G->A	1	A
14401169	M9098	BT	T->C	1	C
15558761	M9142	BT	G->T	1	T
15656524	M9146	BT	G->C	1	C

Table G.13 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
16808769	Z40393	BT	T->C	1	C
16913677	M9187	BT	T->C	1	C
17039032	M9192	BT	C->T	1	T
17151394	M9198	BT	C->T	2	T
17193315	M9200	BT	T->C	1	C
17510373	L962	BT	C->T	1	T
17940510	M9240	BT	C->A	1	A
18162272	M9252	BT	G->A	2	A
18589521	M9257	BT	G->A	1	A
18814791	M9263	BT	G->A	2	A
19233322	M9284	BT	T->C	1	C
21289152	M9315	BT	T->G	1	G
22178817	L440	BT	G->T	1	T
22191243	L1061	BT	G->A	1	A
22730555	M9360	BT	A->G	1	G
23201899	Z40411	BT	A->C	1	C
24506339	Z17388	BT	A->G	1	G
28787653	Z40384	BT	G->C	1	C
6691986	M8959	BT	G->A	1	A
6779907	M8967	BT	G->A	1	A
7563503	Z40386	BT	G->A	1	A
7890655	M9003	BT	C->G	1	G
7895521	M9004	BT	A->G	1	G
8511682	M9026	BT	T->C	1	C
8662142	M9034	BT	G->A	1	A
18759833	CTS9228	C1a1∼	G->A	1	A
21230995	F3805	C2b	C->T	1	T
14207088	M5656	CT	C->A	1	A
15614295	CTS4364	CT	C->T	1	T
15615637	CTS4368	CT	G->C	1	C
15775809	M9150	CT	C->T	1	T
16203547	CTS5318	CT	G->T	1	T
16346371	CTS5532	CT	C->T	1	T
16984364	M5711	CT	G->T	1	T
17104433	M5712	CT	A->C	1	C

Table G.13 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
18715075	M5751	CT	A->G	1	G
18948988	M5756	CT	T->C	1	C
19059200	CTS9760	CT	C->A	1	A
19421388	CTS10362	CT	G->A	1	A
22631570	L1492	CT	A->G	2	G
23555082	M5821	CT	T->A	1	A
23642375	Y1573	CT	C->T	1	T
7210130	M5593	CT	G->A	1	A
7256433	CTS1181	CT	A->T	1	T
7738840	M5609	CT	T->G	1	G
7611998	F3857	D1a1a2	C->T	1	T
6862512	FGC7708	E1b1b1a1a1c1a	G->A	1	A
14334396	F1767	F	G->T	1	T
15095345	CTS3654	F	A->G	1	G
16364286	F2142	F	C->A	1	A
21917313	M89	F	C->T	1	T
2895044	Z34618	H1a1a1c∼	G->A	1	A
7202703	F929	HIJK	C->T	1	T
23796594	M1728	I1a2a1a3a2a∼	G->A	1	A
14872922	S16501	I1a2b1e∼	C->T	1	T
7792789	M2684	IJK	G->A	1	A
17392059	PH3025	M1∼	G->A	1	A
14477015	Z44057	O2a2b1a1a7	G->A	1	A
14725821	M1207	P	T->A	1	A
16263188	CTS5418	P	T->C	1	C
16331558	F332	P	G->A	1	A
16535034	CTS5884	P	A->G	1	G
21139399	M1255	P	C->A	1	A
22003770	M1264	P	G->A	1	A
24443836	F680	P	G->T	1	T
2803717	CTS196	P	C->T	1	T
28612323	L471	P	A->G	1	G
6630931	F29	P	A->G	1	G
7340314	L536	P	T->C	1	C
7549445	F83	P	T->A	1	A

Table G.13 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
8516929	L721	P	G->A	1	A
9171956	PF5882	P	T->A	1	A
9885733	M1199	P	A->G	1	G
14433100	P244	P1∼	G->A	1	A
9974534	F4837	Q1a1a1a2a∼	C->T	1	T
14750668	CTS3229	R	A->C	1	C
22934109	CTS11075	R	A->G	1	G
23202551	CTS11647	R	C->G	1	G
17716251	P286	R1	C->T	1	T
23992762	PF6272	R1b1	C->A	2	A
18407611	FGC35	R1b1a	C->T	1	T
15286480	PF6459	R1b1a1a	G->C	1	C
18617596	CTS9018	R1b1a1a	C->T	1	T
15732786	CTS4608	R1b1a1a2	T->C	1	C
17813541	CTS8052	R1b1a1a2	C->T	2	T
20838224	PF6496	R1b1a1a2	A->G	2	G
21222868	PF6497	R1b1a1a2	C->G	1	G
21983827	PF6507	R1b1a1a2	C->T	1	T
28590278	CTS12478	R1b1a1a2	G->A	1	A
2897433	CTS329	R1b1a1a2	C->G	1	G
7220727	F69	R1b1a1a2	A->G	1	G
7306539	CTS1274	R1b1a1a2	C->G	1	G
8070532	PF6430	R1b1a1a2	T->A	1	A
8194310	PF6432	R1b1a1a2	C->A	1	A
17810865	Z41449	S1a1b1d	C->T	1	T

G.5 Rosemarkie Cave

TABLE G.14: Yleaf output of the Rosemarkie individual, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
13888035	L1116	A0-T	A->C	1	C
14289022	L1235	A0-T	A->T	1	T
15425676	L1123	A0-T	C->T	1	T

Table G.14 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
15426344	L1124	A0-T	T->C	1	C
16596846	L1129	A0-T	T->C	1	C
16661010	L1130	A0-T	T->G	1	G
16718811	L1132	A0-T	A->G	1	G
21593345	L1143	A0-T	A->G	1	G
21903971	L1150	A0-T	A->G	1	G
22191266	L1155	A0-T	G->C	1	C
2887280	L1089	A0-T	G->C	2	C
6753296	L1095	A0-T	A->G	1	G
16071153	FGC26596	A00	C->T	1	T
17716223	L1005	A1	G->A	1	A
2710154	P305	A1	A->G	1	G
4898665	V171	A1	C->G	1	G
7658712	V241	A1	C->T	1	T
15130659	Z11910	A1b	A->G	1	G
16243116	Z11919	A1b	C->T	1	T
18759708	L1053	A1b	C->A	1	A
19109333	Z11893	A1b	C->T	1	T
19390790	Z11917	A1b	A->C	2	A
20820626	Z17898	A1b	A->G	2	A
21217762	Z11913	A1b	T->C	1	C
21646577	L1013	A1b	C->A	1	A
22594175	Z11894	A1b	T->C	1	T
23011519	Z11902	A1b	A->G	1	G
23125972	Z11907	A1b	A->C	1	C
23264209	Z11916	A1b	A->G	1	G
23440815	Z11899	A1b	C->A	1	A
23572106	Z17900	A1b	G->A	2	A
23596262	Z11896	A1b	C->G	1	G
4198481	Z17890	A1b	C->T	1	C
7589303	V221	A1b	G->T	2	T
22107102	M6019	B2a1a1a1∼	C->T	1	T
13202478	Z40398	BT	G->A	1	A
13596520	Z40400	BT	T->A	1	A
13665248	Z40379	BT	G->A	1	A

Table G.14 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
13821858	Z40381	BT	C->G	1	G
14013228	M9075	BT	T->C	1	C
14069414	M9079	BT	C->T	2	T
14075450	M9080	BT	C->T	1	T
14096790	M9081	BT	C->A	1	A
14108214	M9083	BT	T->A	1	A
14344506	M9094	BT	C->G	1	G
14386666	M9097	BT	G->A	1	A
14401169	M9098	BT	T->C	1	C
14472914	M9100	BT	C->T	4	T
14726334	M9112	BT	A->C	2	C
14726915	M9113	BT	C->T	2	T
15041161	M9123	BT	G->T	2	T
15112480	M9128	BT	C->T	1	T
15132194	M9129	BT	T->G	1	G
15261255	Z40392	BT	C->T	1	T
15381787	M9135	BT	C->G	1	G
15451147	M9139	BT	T->A	1	A
15558761	M9142	BT	G->T	1	T
15569966	M9143	BT	T->C	1	C
15656524	M9146	BT	G->C	1	C
15731578	M9148	BT	T->C	1	C
16332337	M9166	BT	C->T	2	T
16686965	M9179	BT	A->G	1	G
16694398	M9180	BT	C->T	1	T
16808769	Z40393	BT	T->C	3	C
16946901	M9188	BT	G->A	2	A
17039032	M9192	BT	C->T	1	T
17151394	M9198	BT	C->T	1	T
17298439	M9209	BT	G->A	1	A
17372321	M9213	BT	G->A	2	A
17485537	M9217	BT	C->T	1	T
17510373	L962	BT	C->T	2	T
17511829	M9219	BT	T->C	1	C
17642965	M9225	BT	C->T	1	T

Table G.14 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
17670179	M9226	BT	C->T	1	T
17670193	M9227	BT	G->A	1	A
17753892	M9231	BT	G->A	2	A
17886479	M9237	BT	G->A	1	A
18016799	Y9422	BT	T->A	1	A
18069115	M9246	BT	G->A	1	A
18152393	M9251	BT	C->T	1	T
18179224	M9253	BT	G->A	1	A
18262817	Y8320	BT	G->A	1	A
18585407	M9255	BT	T->G	1	G
18814791	M9263	BT	G->A	2	A
19118181	M9277	BT	C->G	1	G
19140679	M9278	BT	C->T	1	T
19317082	M9289	BT	T->C	1	C
19457726	M9296	BT	G->A	1	A
19519830	M9298	BT	T->C	1	C
19555949	Z17372	BT	T->G	1	G
20817229	Y8322	BT	G->A	1	A
21092276	M9301	BT	C->T	1	T
21238206	Z40406	BT	C->T	2	T
21248962	Z40394	BT	G->A	3	A
21259569	M9311	BT	C->T	1	T
21292327	M9316	BT	C->T	1	T
21311141	M9317	BT	C->T	2	T
21486590	M9325	BT	T->C	1	C
21541028	Z40396	BT	C->T	1	T
21627877	M9334	BT	C->T	1	T
21637759	M9335	BT	A->G	1	G
21674225	L969	BT	T->C	1	C
21815245	M9341	BT	G->A	1	A
21968581	M9343	BT	G->T	2	T
22079278	M9348	BT	A->T	1	T
22129214	PF1096	BT	T->C	1	C
22178817	L440	BT	G->T	1	T
22191243	L1061	BT	G->A	1	A

Table G.14 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
22625732	M9356	BT	C->T	3	T
22730555	M9360	BT	A->G	2	G
22748506	M299	BT	T->G	2	G
22817840	M9365	BT	G->C	1	C
22881961	M9368	BT	G->C	1	C
22903616	M9369	BT	T->C	1	C
22931502	M9370	BT	C->T	1	T
23010471	M9373	BT	C->T	1	T
23237493	Z12129	BT	T->A	2	A
23244049	M9379	BT	G->C	2	C
23256074	M9380	BT	A->G	1	G
23347971	M9387	BT	C->T	1	T
23405276	M9389	BT	T->C	1	C
23558260	M9393	BT	G->A	2	A
24377975	M9406	BT	G->T	2	T
24509299	Z17389	BT	C->T	1	T
2669416	L1071	BT	A->G	2	G
2712956	M8947	BT	G->A	1	A
2750722	M8949	BT	G->C	1	C
2889399	M8954	BT	T->C	2	C
2906950	M8956	BT	C->T	1	T
6646548	M8958	BT	T->C	1	C
6859691	M8969	BT	C->T	1	T
6936412	M8971	BT	G->C	1	C
6993202	M8972	BT	C->T	2	T
7032507	Z17366	BT	C->T	2	T
7269376	M8980	BT	C->T	1	T
7302697	M8983	BT	G->A	1	A
7531097	M8988	BT	T->C	1	C
7645425	V235	BT	A->G	1	G
7711024	M8995	BT	A->G	1	G
7815370	M9000	BT	T->A	1	A
7890655	M9003	BT	C->G	1	G
7944491	M9008	BT	C->A	1	A
8021764	M9010	BT	G->A	2	A

Table G.14 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
8042040	Z40387	BT	G->A	1	A
8226191	M9017	BT	C->A	1	A
8596285	M9030	BT	G->A	1	A
8596399	M9031	BT	A->G	2	G
8613037	L1062	BT	G->A	1	A
8635293	M9032	BT	C->A	1	A
8662142	M9034	BT	G->A	1	A
8881498	M9042	BT	G->A	1	A
9827891	M9054	BT	T->C	1	C
9869354	M11760	BT	C->A	1	A
9891742	M9066	BT	C->T	2	T
16923369	Z31983	C1b2a	C->T	1	T
13213798	Z17710	CT	G->T	1	T
13594339	Y1791	CT	C->T	2	T
14289530	M5657	CT	A->C	1	C
14439817	CTS2711	CT	G->A	1	A
14917660	CTS3431	CT	C->G	1	G
15750550	M5683	CT	G->A	1	A
15792531	CTS4740	CT	G->A	1	A
16131945	Y1531	CT	C->G	1	G
16203547	CTS5318	CT	G->T	1	T
16262390	M5691	CT	T->A	1	A
16325663	M5692	CT	A->C	2	C
16822011	CTS6327	CT	A->G	2	G
16887784	M5709	CT	G->C	1	C
17104433	M5712	CT	A->C	3	C
17174326	CTS6890	CT	G->C	1	C
17186912	CTS6907	CT	C->T	1	T
17589395	M5723	CT	G->A	1	A
17630903	M5724	CT	C->T	1	T
17853378	M5734	CT	C->T	2	T
17894575	CTS8243	CT	C->T	2	T
17897543	M5737	CT	C->T	1	T
17909505	M5738	CT	G->A	1	A
18592845	CTS8980	CT	G->A	1	A

Table G.14 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
18616467	CTS9014	CT	T->A	1	A
18905190	CTS9458	CT	G->A	1	A
18940202	M5754	CT	A->G	1	G
19090637	CTS9828	CT	C->G	1	G
19169483	PF970	CT	T->C	1	C
19481868	M5771	CT	C->A	1	A
21095622	PF1016	CT	C->T	1	T
22136242	M5796	CT	C->G	2	G
22571947	M5800	CT	A->G	3	G
23081753	CTS11358	CT	A->G	2	G
23111277	M5813	CT	C->A	1	A
23126086	M5814	CT	C->T	1	T
23234852	Y1571	CT	G->A	1	A
23298645	CTS11827	CT	T->C	1	C
23567930	M5823	CT	C->T	2	T
23642375	Y1573	CT	C->T	1	T
24358251	Y1574	CT	T->C	2	C
24470911	M5826	CT	C->A	2	A
2733618	M8948	CT	C->A	1	A
2744386	CTS125	CT	G->T	2	T
28590906	Y1578	CT	A->T	1	T
28650362	CTS12633	CT	T->A	2	A
6523383	PF192	CT	T->C	1	C
6831340	CTS543	CT	C->T	2	T
6991000	M5588	CT	C->A	1	A
7008958	M5589	CT	T->C	1	C
7041647	M5590	CT	G->A	1	A
7210130	M5593	CT	G->A	1	A
7256433	CTS1181	CT	A->T	1	T
7275087	CTS1217	CT	C->T	2	T
7358315	M5599	CT	A->G	1	G
7597185	M5603	CT	G->A	2	A
7707388	M5606	CT	T->G	1	G
7707484	M5607	CT	C->A	1	A
7778691	M5611	CT	C->T	3	T

Table G.14 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
7782393	M5612	CT	C->T	1	T
7887815	M5615	CT	A->T	1	T
8066525	M5621	CT	G->A	1	A
8080397	M5622	CT	T->A	1	A
8262951	M5629	CT	T->C	1	C
8849835	M5639	CT	C->A	1	A
9758726	PF342	CT	G->A	1	A
16869670	CTS6392	D1b1c1a1	C->T	1	T
14159750	CTS2031	E1a	G->A	2	A
7855759	M4739	E1b1a1a1	C->T	1	T
17675730	Y21121	E1b1b1b2a1b1∼	G->A	1	A
14149772	F1704	F	C->T	2	T
14171665	F1714	F	C->T	1	T
14334396	F1767	F	G->T	3	T
14832620	M235	F	T->G	1	G
15234830	PF2669	F	T->A	1	T
15526751	M213	F	T->C	1	C
15651438	CTS4443	F	A->T	1	T
16242316	P163	F	A->T	1	T
17311975	P140	F	G->C	1	C
19205814	L470	F	A->T	1	T
19220444	F2985	F	A->G	1	G
19349615	P148	F	C->T	1	T
21147336	F3136	F	T->C	1	C
21618856	P135	F	C->T	2	T
21917313	M89	F	C->T	1	T
21937573	M3750	F	A->G	1	G
2722506	F719	F	G->A	1	A
28650343	CTS12632	F	C->G	1	G
7395806	P134	F	C->G	1	G
7726849	M3647	F	A->G	1	G
8131538	F3692	F	G->T	1	T
8227605	M3652	F	A->G	1	G
8602415	P146	F	C->T	1	T
8041029	Z34704	H1a1a4b2c1a2∼	G->A	1	A

Table G.14 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
17547934	Z14100	H1a2a	C->T	1	T
7206154	Z41307	H2b1∼	C->T	1	T
7202703	F929	HIJK	C->T	1	T
21155653	FGC2414	I	C->T	1	T
23197595	FGC22174	I1a2a1a4b1a3∼	C->T	1	T
2752275	Y19884	I1a3f∼	C->T	1	T
7173143	L16	IJK	G->A	1	A
7629583	F3689	IJK	A->G	2	G
6739856	P81	J2a1g	G->A	1	A
20837553	P128	K∼	C->T	1	T
10083691	Z20140	L1∼	G->A	1	A
6670461	L58	N1a1a1a1a1a1a2a1∼	T->G	1	G
13211699	PF5888	P	A->T	1	T
14117810	CTS1907	P	C->T	1	T
14933671	CTS3446	P	C->A	1	A
15172515	CTS3775	P	C->G	1	G
15480547	L779	P	C->G	1	G
15648415	M1109	P	C->T	1	T
16489668	CTS5808	P	G->T	1	T
16497020	F344	P	T->C	1	C
17341676	CTS7194	P	A->G	2	G
17726250	CTS7886	P	T->G	1	G
17954035	CTS8356	P	T->C	2	C
18647146	M1240	P	A->C	1	C
19144375	F524	P	C->T	1	T
19245417	CTS10081	P	C->A	2	A
19417241	CTS10348	P	G->A	1	A
21121986	M1254	P	A->T	1	T
21139399	M1255	P	C->A	1	A
21570974	M1260	P	G->A	1	A
21603164	M1160	P	T->G	2	G
21804435	M1261	P	C->T	1	T
21811108	L741	P	G->T	1	T
21848764	M1262	P	A->C	1	C
22040221	M1265	P	G->C	1	C

Table G.14 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
22514081	L82	P	G->T	1	T
23342709	L768	P	C->A	1	A
23431541	CTS12028	P	A->G	1	G
23993156	PF5982	P	A->C	1	C
2803717	CTS196	P	C->T	1	T
28655471	CTS12648	P	A->C	1	C
6931040	PF5858	P	C->G	2	G
7549445	F83	P	T->A	1	A
7628900	F91	P	G->A	1	A
8257661	M1190	P	A->G	1	G
8326618	M1192	P	C->T	1	T
9098922	F180	P	C->T	1	T
9171956	PF5882	P	T->A	1	A
9410828	M1197	P	C->T	1	T
9921064	PF5481	P	G->A	1	A
14433100	P244	P1∼	G->A	1	A
17881230	P239	P1∼	G->C	1	C
19198212	P235	P1∼	A->G	1	G
8334875	P237	P1∼	A->G	2	G
8845380	P226	P1∼	C->T	1	T
14300457	CTS2426	R	G->A	1	A
15078469	CTS3622	R	C->G	1	G
16615413	L747	R	G->T	1	T
17285993	P224	R	C->T	1	T
17334694	M718	R	G->T	1	T
18066156	M734	R	C->T	1	T
18611644	CTS9005	R	A->T	1	T
18744476	PF5938	R	T->C	1	C
19267344	P285	R	C->A	1	A
21263029	M764	R	G->A	1	A
22733758	L1225	R	C->G	1	G
6701239	F33	R	G->A	2	A
8027859	M628	R	G->C	1	C
14366723	CTS2565	R1	C->T	1	T
15590342	P225	R1	G->T	1	T

Table G.14 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
16742224	L875	R1	A->G	2	G
17275703	CTS7085	R1	G->A	1	A
17782178	P236	R1	C->G	1	G
17839981	CTS8116	R1	G->A	1	A
22750583	M306	R1	C->A	1	A
7647357	P242	R1	G->A	1	A
7854412	F102	R1	A->G	2	G
8633545	P245	R1	T->C	1	C
2887824	M343	R1b	C->A	1	A
14193384	CTS2134	R1b1	G->A	2	A
21995972	L506	R1b1	T->A	1	A
2686555	CTS46	R1b1	G->A	1	A
7081561	CTS910	R1b1	C->T	1	T
8110520	PF6248	R1b1	T->A	1	A
13822833	FGC36	R1b1a	G->T	1	T
14637352	CTS3063	R1b1a	T->C	2	C
17545608	CTS7585	R1b1a	G->T	1	T
21558298	L1345	R1b1a	G->T	1	T
23984056	PF6271	R1b1a	G->A	1	A
7900883	FGC41	R1b1a	C->A	1	A
8214827	PF6249	R1b1a	C->T	1	T
28733101	L389	R1b1a1	C->G	3	G
15239181	CTS3876	R1b1a1a	G->C	2	C
15286480	PF6459	R1b1a1a	G->C	1	C
16005138	CTS5082	R1b1a1a	A->C	1	C
17732408	CTS7904	R1b1a1a	T->C	1	C
19020340	L502	R1b1a1a	G->C	2	C
21447844	PF6501	R1b1a1a	A->T	1	T
10062719	PF6441	R1b1a1a2	C->G	3	G
13511147	PF6443	R1b1a1a2	C->T	1	T
13887941	L407	R1b1a1a2	G->A	1	A
13945593	M11805	R1b1a1a2	A->T	1	T
14005779	PF6448	R1b1a1a2	G->A	1	A
14136291	PF6452	R1b1a1a2	G->A	1	A
15037433	CTS3575	R1b1a1a2	C->G	1	G

Table G.14 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
18381735	PF6482	R1b1a1a2	A->G	2	G
19291359	CTS10149	R1b1a1a2	T->C	1	C
19462180	CTS10451	R1b1a1a2	C->T	1	T
20828795	PF6495	R1b1a1a2	G->A	1	A
20838224	PF6496	R1b1a1a2	A->G	1	G
21554468	PF6265	R1b1a1a2	G->A	1	A
21993844	PF6508	R1b1a1a2	G->A	1	A
22796697	CTS10834	R1b1a1a2	T->C	3	C
23124367	CTS11468	R1b1a1a2	G->T	1	T
23379254	CTS11948	R1b1a1a2	G->A	2	A
2668456	PF6399	R1b1a1a2	C->T	1	T
4446430	M520	R1b1a1a2	T->A	1	A
7863189	L482	R1b1a1a2	G->A	1	A
8296441	PF6433	R1b1a1a2	G->T	1	T
8826595	PF6436	R1b1a1a2	C->T	2	T
16492547	L151	R1b1a1a2a1a	C->T	1	T
17844018	L11	R1b1a1a2a1a	T->C	1	C
23155891	S25234	R1b1a1a2a1a1c1a2b1	G->A	1	A
14635223	CTS3057	R1b1a1a2a1a2c1a1f1	C->T	1	T

G.6 Worth Matravers

TABLE G.15: Yleaf output of KD007, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
15442970	M11301	A1b	T->C	1	T
14723798	M9111	BT	T->C	1	C
16682167	Y8489	BT	G->C	1	C
17147159	M9197	BT	C->T	1	T
17886479	M9237	BT	G->A	1	A
21311141	M9317	BT	C->T	1	T
22719912	M9357	BT	G->A	1	A
23010471	M9373	BT	C->T	1	T
23201899	Z40411	BT	A->C	1	C

Table G.15 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
23244049	M9379	BT	G->C	1	C
7221998	M8979	BT	C->A	1	A
7661695	M8994	BT	A->C	1	C
14074463	Y1525	CT	C->T	1	T
15614295	CTS4364	CT	C->T	1	T
19039750	M5762	CT	G->C	1	C
19503081	CTS10512	CT	C->A	1	A
28650362	CTS12633	CT	T->A	1	A
6991000	M5588	CT	C->A	1	A
8183439	M5627	CT	G->A	1	A
28650343	CTS12632	F	C->G	1	G
23244026	P15	G2a	C->T	1	T
17437378	M2917	H1a1	G->A	1	A
21730257	M9	K	C->G	1	G
19303108	CTS10168	P	A->C	1	C
7549445	F83	P	T->A	1	A
8845380	P226	P1∼	C->T	1	T
8568360	F4777	Q1a1a1a2a∼	C->T	1	T
14300457	CTS2426	R	G->A	1	A
14561760	CTS2913	R	A->G	1	G
16882568	CTS6417	R	T->C	1	C
17722802	CTS7876	R	G->A	1	A
22687547	CTS10663	R	A->T	1	T
14366723	CTS2565	R1	C->T	1	T
7771131	P238	R1	G->A	1	A
13816025	PF6444	R1b1a1a2	G->A	1	A
21784286	L1350	R1b1a1a2	G->A	1	A
9392948	PF6437	R1b1a1a2	C->T	1	T
8466315	Y10794	R1b1a1a2a1a2c1a2a2a1a1a	C->T	1	T

TABLE G.23: Yleaf output of KD021, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
21739790	L1145	A0-T	C->T	1	T

Table G.23 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
22191266	L1155	A0-T	G->C	1	C
10008803	L418	BT	C->G	1	G
10033862	Z40374	BT	A->G	1	G
14018605	M9076	BT	G->A	1	A
14723798	M9111	BT	T->C	1	C
15095089	M9126	BT	C->T	1	T
15218792	Z40383	BT	T->C	1	C
15451147	M9139	BT	T->A	1	A
16302343	M9165	BT	G->A	1	A
17324151	M9210	BT	A->G	1	G
17504731	M9218	BT	G->A	1	A
18162272	M9252	BT	G->A	1	A
18585407	M9255	BT	T->G	1	G
19325165	M9290	BT	C->T	1	T
19457726	M9296	BT	G->A	1	A
19519830	M9298	BT	T->C	1	C
20817229	Y8322	BT	G->A	1	A
21665439	M9338	BT	C->A	1	A
21938158	M94	BT	C->A	1	A
22191243	L1061	BT	G->A	1	A
22931502	M9370	BT	C->T	1	T
23010471	M9373	BT	C->T	1	T
23237493	Z12129	BT	T->A	1	A
23240528	Z40412	BT	C->G	1	G
2712956	M8947	BT	G->A	1	A
6779907	M8967	BT	G->A	1	A
6908261	V29	BT	A->G	1	G
6932831	L413	BT	G->A	1	A
6936412	M8971	BT	G->C	1	C
7645425	V235	BT	A->G	1	G
8180027	Z40388	BT	A->G	1	G
8596399	M9031	BT	A->G	1	G
9869354	M11760	BT	C->A	1	A
8885973	K147	C1b1a1a	G->A	1	A
14813991	M168	CT	C->T	1	T

Table G.23 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
17186912	CTS6907	CT	C->T	1	T
19261216	CTS10110	CT	C->A	1	A
23105586	M5812	CT	C->A	1	A
2733618	M8948	CT	C->A	1	A
7210130	M5593	CT	G->A	1	A
7233061	Z40571	CT	G->A	1	A
8080397	M5622	CT	T->A	2	A
9094343	M5642	CT	C->T	1	T
9158586	PF328	CT	G->A	1	A
19349615	P148	F	C->T	1	T
21618856	P135	F	C->T	1	T
2912385	M3639	F	T->C	1	C
7726849	M3647	F	A->G	1	G
8424089	P145	F	G->A	1	A
14974451	L1197	I	C->T	1	T
15389836	CTS4088	I	T->C	1	C
15595624	CTS4340	I	G->A	1	A
19233673	CTS10058	I	A->G	1	G
21067903	PF3794	I	C->T	1	T
21402723	PF3800	I	A->G	1	G
21841289	PF3815	I	G->T	1	T
7681156	PF3640	I	T->A	1	A
8382265	FI2	I	C->G	2	G
8643763	PF3665	I	A->G	2	G
14074218	CTS1802	I2a1b	A->T	1	T
22905944	CTS11030	I2a1b	G->C	1	C
2785672	CTS176	I2a1b	A->G	1	G
22513718	L161.1	I2a1b1	C->T	2	T
13676377	FGC1574	IJ	G->T	1	T
22909101	F3368	IJ	C->T	1	T
22200414	FGC23865	O2a2b1a2a1a3b2b1	T->C	1	C
9418014	PF7465	T1	G->A	2	A

G.7 West Heslerton

TABLE G.25: Yleaf output of KD031, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
15425676	L1123	A0-T	C->T	2	T
21739790	L1145	A0-T	C->T	1	T
16773792	L1004	A1	T->C	2	C
2710154	P305	A1	A->G	1	G
14774366	Z17894	A1b	C->T	1	T
15442970	M11301	A1b	T->C	2	T
18759708	L1053	A1b	C->A	1	A
21646577	L1013	A1b	C->A	1	A
13592206	Z40378	BT	C->A	2	A
14069414	M9079	BT	C->T	1	T
14075450	M9080	BT	C->T	1	T
14096790	M9081	BT	C->A	1	A
14344506	M9094	BT	C->G	1	G
14588372	M9105	BT	G->A	1	A
14723798	M9111	BT	T->C	1	C
14980170	M9121	BT	C->T	1	T
15063698	M9124	BT	C->A	1	A
15095089	M9126	BT	C->T	1	T
15112480	M9128	BT	C->T	6	T
15256839	M9133	BT	A->G	1	G
15261255	Z40392	BT	C->T	1	T
15382324	M9136	BT	C->G	1	G
15514552	M9140	BT	T->C	1	C
16542464	M9174	BT	A->C	1	C
16959883	M9189	BT	A->G	3	G
17041558	M9193	BT	G->A	1	A
17063750	M9195	BT	T->A	1	A
17261013	M9202	BT	G->A	1	A
17298439	M9209	BT	G->A	2	A
17353147	L1220	BT	A->G	1	G
17372321	M9213	BT	G->A	1	A
17504731	M9218	BT	G->A	1	A
17510373	L962	BT	C->T	3	T

Table G.25 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
17886479	M9237	BT	G->A	2	A
17940510	M9240	BT	C->A	1	A
17966904	M9242	BT	C->A	1	A
18067088	M9245	BT	T->A	2	A
18628317	M9260	BT	G->A	1	A
18814791	M9263	BT	G->A	4	A
18917853	M9267	BT	C->T	1	T
19325165	M9290	BT	C->T	1	T
19337259	M9291	BT	T->C	1	C
19416000	M9293	BT	C->G	1	G
19457726	M9296	BT	G->A	1	A
19555949	Z17372	BT	T->G	1	G
21092276	M9301	BT	C->T	1	T
21259569	M9311	BT	C->T	1	T
21268114	M9312	BT	C->G	2	G
21289152	M9315	BT	T->G	1	G
21326695	M9319	BT	A->G	2	G
21567686	M9331	BT	C->G	1	G
21646196	M9336	BT	G->A	2	A
22719912	M9357	BT	G->A	4	A
22727889	M9359	BT	C->T	1	T
22730777	M9361	BT	T->C	1	C
22817840	M9365	BT	G->C	1	C
23201899	Z40411	BT	A->C	2	C
23237493	Z12129	BT	T->A	1	A
23244049	M9379	BT	G->C	1	C
23405276	M9389	BT	T->C	1	C
23565114	M9394	BT	T->A	2	A
24509299	Z17389	BT	C->T	1	T
2712956	M8947	BT	G->A	2	A
28746408	Z17390	BT	G->C	2	C
28787653	Z40384	BT	G->C	2	C
2889399	M8954	BT	T->C	1	C
2899711	M8955	BT	A->G	1	G
6779907	M8967	BT	G->A	1	A

Table G.25 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
6859691	M8969	BT	C->T	1	T
6932831	L413	BT	G->A	1	A
7021041	Z17365	BT	G->T	1	T
7171532	M8977	BT	G->T	1	T
7269376	M8980	BT	C->T	1	T
7531097	M8988	BT	T->C	2	C
7621889	M8993	BT	C->T	1	T
7645425	V235	BT	A->G	1	G
7904101	M9005	BT	G->A	1	A
7907933	M9006	BT	G->A	1	A
8180027	Z40388	BT	A->G	1	G
8446082	M9025	BT	A->T	1	T
8594070	Z40389	BT	G->A	2	A
8596399	M9031	BT	A->G	1	G
8635293	M9032	BT	C->A	2	A
8881498	M9042	BT	G->A	1	A
9106630	M9046	BT	A->G	2	G
9869354	M11760	BT	C->A	1	A
7545688	F988	C1b1a2b	C->T	1	T
23301174	CTS11833	C2c1	C->T	1	T
15203676	CTS3818	CF	A->G	1	G
10049308	Z17708	CT	A->G	1	G
13652845	Z17714	CT	A->T	1	T
14079528	L957	CT	C->T	1	T
14149520	M5652	CT	T->A	1	A
14207088	M5656	CT	C->A	1	A
14439817	CTS2711	CT	G->A	1	A
15097073	CTS3662	CT	G->A	1	A
16465771	CTS5746	CT	T->G	1	G
16844116	M5706	CT	G->A	1	A
17104433	M5712	CT	A->C	3	C
17301733	M5716	CT	T->C	1	C
17515027	M5721	CT	G->A	2	A
17630903	M5724	CT	C->T	1	T
17742131	CTS7922	CT	C->T	1	T

Table G.25 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
17894575	CTS8243	CT	C->T	4	T
17975155	FGC33852	CT	G->T	2	T
18616467	CTS9014	CT	T->A	1	A
18715075	M5751	CT	A->G	1	G
18974195	M5760	CT	C->T	1	T
19167672	CTS9948	CT	G->C	1	C
19169483	PF970	CT	T->C	1	C
19261216	CTS10110	CT	C->A	2	A
19503081	CTS10512	CT	C->A	2	A
21281606	M5778	CT	G->C	1	C
22744945	M294	CT	C->T	1	T
23105586	M5812	CT	C->A	2	A
23234852	Y1571	CT	G->A	2	A
23567930	M5823	CT	C->T	4	T
24443898	M5825	CT	C->T	4	T
28790215	PF1337	CT	G->C	2	C
6831340	CTS543	CT	C->T	1	T
6864267	M5585	CT	A->G	1	G
7041647	M5590	CT	G->A	1	A
7256433	CTS1181	CT	A->T	3	T
7275087	CTS1217	CT	C->T	2	T
7358315	M5599	CT	A->G	2	G
7707484	M5607	CT	C->A	1	A
7796647	M5613	CT	C->T	1	T
8080397	M5622	CT	T->A	2	A
8183439	M5627	CT	G->A	1	A
8526565	M5632	CT	G->A	1	A
8543804	M5633	CT	G->A	1	A
9094343	M5642	CT	C->T	1	T
9158586	PF328	CT	G->A	2	A
7013032	Z31583	D1a1a1a2	C->T	1	T
15651438	CTS4443	F	A->T	1	T
16242316	P163	F	A->T	1	T
17256018	P166	F	C->T	1	T
17493513	P158	F	C->T	4	T

Table G.25 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
21618856	P135	F	C->T	2	T
22673903	F3335	F	T->G	1	G
2756471	M3637	F	A->T	1	T
8442341	F1209	F	A->G	1	G
8572150	F1320	F	G->A	1	A
8602415	P146	F	C->T	3	T
14237670	CTS2254	GHIJK	C->T	1	T
7202703	F929	HIJK	C->T	2	T
22228628	L672	I2a1a1a	T->A	3	A
6753519	L15	IJK	A->G	1	G
7173143	L16	IJK	G->A	1	A
7702973	M2683	IJK	T->A	1	A
7792789	M2684	IJK	G->A	1	A
23550924	M526	K2	A->C	4	C
8609424	SK1417	L1a∼	C->T	1	T
24465292	M2128	N1a1a1a1a	G->A	1	A
10035664	FGC285	P	G->C	1	C
14679362	CTS3135	P	C->A	1	A
15111148	CTS3697	P	G->C	1	C
17341676	CTS7194	P	A->G	1	G
17502468	CTS7481	P	C->T	1	T
19245417	CTS10081	P	C->A	1	A
21298239	M1256	P	C->G	1	G
21570974	M1260	P	G->A	1	A
22003770	M1264	P	G->A	1	A
2686727	F4	P	G->A	1	A
28612323	L471	P	A->G	1	G
28655471	CTS12648	P	A->C	1	C
7549445	F83	P	T->A	1	A
7848043	M1186	P	C->G	2	G
7996768	PF5867	P	G->A	1	A
8096754	M1188	P	T->C	1	C
8257661	M1190	P	A->G	1	G
9098922	F180	P	C->T	1	T
9171956	PF5882	P	T->A	1	A

Table G.25 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
17881230	P239	P1∼	G->C	2	C
8334875	P237	P1∼	A->G	1	G
8845380	P226	P1∼	C->T	1	T
14561760	CTS2913	R	A->G	1	G
17723850	CTS7880	R	C->T	2	T
21263029	M764	R	G->A	1	A
22934109	CTS11075	R	A->G	1	G
2810583	CTS207	R	A->G	1	G
9889199	M651	R	G->A	1	A
14366723	CTS2565	R1	C->T	2	T
14829196	CTS3321	R1	C->T	1	T
16394489	CTS5611	R1	T->G	1	G
7570822	P294	R1	G->C	1	C
7771131	P238	R1	G->A	1	A
2887824	M343	R1b	C->A	2	A
14193384	CTS2134	R1b1	G->A	1	A
14226692	CTS2229	R1b1	T->A	2	A
22722580	L1349	R1b1	T->C	1	C
10038192	A702	R1b1a	G->A	1	A
18407611	FGC35	R1b1a	C->T	1	T
7900883	FGC41	R1b1a	C->A	2	A
15239181	CTS3876	R1b1a1a	G->C	2	C
17986687	PF6475	R1b1a1a	C->A	1	A
18617596	CTS9018	R1b1a1a	C->T	2	T
19020340	L502	R1b1a1a	G->C	2	C
13657777	L777	R1b1a1a2	T->C	1	C
15037433	CTS3575	R1b1a1a2	C->G	1	G
17594966	CTS7659	R1b1a1a2	C->G	1	G
17813541	CTS8052	R1b1a1a2	C->T	1	T
18095336	CTS8591	R1b1a1a2	A->C	1	C
19054889	L757	R1b1a1a2	C->T	2	T
19417394	CTS10349	R1b1a1a2	A->C	1	C
20811307	PF6494	R1b1a1a2	G->A	1	A
22796697	CTS10834	R1b1a1a2	T->C	1	C
23124367	CTS11468	R1b1a1a2	G->T	1	T

Table G.25 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
23242935	L1348	R1b1a1a2	G->C	1	C
2668456	PF6399	R1b1a1a2	C->T	1	T
28771116	CTS12972	R1b1a1a2	C->G	1	G
7220727	F69	R1b1a1a2	A->G	1	G
9464078	PF6438	R1b1a1a2	C->T	1	T
6753511	L23	R1b1a1a2a	G->A	1	A
8502236	L51	R1b1a1a2a1	G->A	2	A
16492547	L151	R1b1a1a2a1a	C->T	1	T
8796078	M405	R1b1a1a2a1a1	C->T	1	T
18945748	S21514	R1b1a1a2a1a1g	C->T	2	T
19305653	A225	R1b1a1a2a1a2c1a1a1a1a1∼	G->A	1	A

TABLE G.26: Yleaf output of KD032, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
15425676	L1123	A0-T	C->T	1	T
16428812	L1128	A0-T	C->T	1	T
16596846	L1129	A0-T	T->C	1	C
16661010	L1130	A0-T	T->G	1	G
6753296	L1095	A0-T	A->G	1	G
7590048	L1105	A0-T	C->T	1	T
22957136	L1084	A1	G->A	1	A
14774366	Z17894	A1b	C->T	1	T
13596520	Z40400	BT	T->A	1	A
13665248	Z40379	BT	G->A	1	A
14018605	M9076	BT	G->A	1	A
14588372	M9105	BT	G->A	1	A
15112480	M9128	BT	C->T	1	T
15382324	M9136	BT	C->G	1	G
15905094	M9152	BT	G->T	2	T
16542464	M9174	BT	A->C	1	C
16676762	M9178	BT	T->C	2	C
16913677	M9187	BT	T->C	1	C
17063750	M9195	BT	T->A	1	A

Table G.26 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
17270785	M9204	BT	C->A	1	A
17353147	L1220	BT	A->G	1	G
17388578	M9214	BT	A->G	1	G
17510373	L962	BT	C->T	1	T
17670179	M9226	BT	C->T	2	T
17670193	M9227	BT	G->A	2	A
17886479	M9237	BT	G->A	1	A
17940510	M9240	BT	C->A	1	A
18053867	M9244	BT	T->A	1	A
18814791	M9263	BT	G->A	2	A
19149592	M9280	BT	C->A	2	A
19445252	M9295	BT	A->C	1	C
20817229	Y8322	BT	G->A	1	A
21122071	M9303	BT	A->G	1	G
21326695	M9319	BT	A->G	1	G
21567686	M9331	BT	C->G	1	G
21627877	M9334	BT	C->T	6	T
21665439	M9338	BT	C->A	1	A
22719912	M9357	BT	G->A	2	A
23005081	M9372	BT	G->T	1	T
23010471	M9373	BT	C->T	1	T
23070865	M9376	BT	A->C	1	C
23199069	Z40409	BT	G->T	2	T
23244049	M9379	BT	G->C	2	C
23405276	M9389	BT	T->C	1	C
23565114	M9394	BT	T->A	1	A
6859691	M8969	BT	C->T	1	T
6932831	L413	BT	G->A	1	A
6993202	M8972	BT	C->T	1	T
7032507	Z17366	BT	C->T	1	T
7269376	M8980	BT	C->T	2	T
7302697	M8983	BT	G->A	1	A
7621889	M8993	BT	C->T	2	T
7815370	M9000	BT	T->A	1	A
8226191	M9017	BT	C->A	1	A

Table G.26 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
8540845	M9028	BT	A->G	2	G
8596285	M9030	BT	G->A	1	A
9381562	Y8311	BT	G->A	1	A
9869354	M11760	BT	C->A	1	A
6913949	CTS626	C2c1	C->T	1	T
14197867	P143	CF	G->A	2	A
16863259	CTS6376	CF	C->G	1	G
13213798	Z17710	CT	G->T	1	T
13852599	PF629	CT	A->G	1	G
14074463	Y1525	CT	C->T	2	T
15615637	CTS4368	CT	G->C	1	C
16131945	Y1531	CT	C->G	2	G
17186912	CTS6907	CT	C->T	1	T
17502617	CTS7482	CT	A->G	2	G
17589395	M5723	CT	G->A	2	A
17897543	M5737	CT	C->T	1	T
18715075	M5751	CT	A->G	2	G
19039750	M5762	CT	G->C	1	C
19090637	CTS9828	CT	C->G	1	G
19132472	M5765	CT	G->A	1	A
19169483	PF970	CT	T->C	1	C
19503081	CTS10512	CT	C->A	1	A
21281606	M5778	CT	G->C	3	C
21775402	M5790	CT	G->A	1	A
22529475	Y1567	CT	C->G	2	G
22639802	PF1205	CT	C->T	1	T
24470911	M5826	CT	C->A	1	A
28789895	Y1580	CT	A->T	1	T
28790215	PF1337	CT	G->C	1	C
6619341	M5578	CT	C->A	2	A
6864267	M5585	CT	A->G	1	G
7210130	M5593	CT	G->A	1	A
7256433	CTS1181	CT	A->T	1	T
7533511	M5601	CT	G->A	1	A
7782393	M5612	CT	C->T	1	T

Table G.26 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
8183439	M5627	CT	G->A	1	A
8396636	M5631	CT	G->A	1	A
8526565	M5632	CT	G->A	2	A
8543804	M5633	CT	G->A	1	A
9758726	PF342	CT	G->A	1	A
14832620	M235	F	T->G	1	G
15323154	CTS3996	F	A->G	2	G
15651438	CTS4443	F	A->T	1	T
16364286	F2142	F	C->A	1	A
16401339	F2155	F	C->T	1	T
21618856	P135	F	C->T	1	T
6941218	P141	F	G->A	1	A
7395806	P134	F	C->G	1	G
8442341	F1209	F	A->G	1	G
8680661	P151	F	T->C	1	C
8589031	F1329	GHIJK	C->T	1	T
7997053	Z14319	H1a2b1a	C->T	1	T
13849996	Z13526	H3	C->T	1	T
7202703	F929	HIJK	C->T	1	T
21180695	Y3984	I1a3a1a2a2∼	C->T	1	T
6753519	L15	IJK	A->G	1	G
7173143	L16	IJK	G->A	2	A
7629583	F3689	IJK	A->G	1	G
7792789	M2684	IJK	G->A	1	A
23550924	M526	K2	A->C	1	C
10077460	PF5887	P	A->G	1	G
15172515	CTS3775	P	C->G	1	G
17341676	CTS7194	P	A->G	1	G
19245417	CTS10081	P	C->A	1	A
19303108	CTS10168	P	A->C	1	C
21298239	M1256	P	C->G	2	G
22514081	L82	P	G->T	1	T
7340314	L536	P	T->C	1	C
7996768	PF5867	P	G->A	1	A
8096754	M1188	P	T->C	4	C

Table G.26 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
9885733	M1199	P	A->G	1	G
22094491	P283	P1∼	C->T	1	T
8334875	P237	P1∼	A->G	1	G
15667208	FGC1168	R	G->C	2	C
16491135	CTS5815	R	C->T	1	T
18017528	F459	R	G->T	1	T
21263029	M764	R	G->A	1	A
21843090	P280	R	C->G	1	G
22934109	CTS11075	R	A->G	1	G
23202551	CTS11647	R	C->G	1	G
24360964	F765	R	G->A	1	A
9889199	M651	R	G->A	1	A
16394489	CTS5611	R1	T->G	2	G
16742224	L875	R1	A->G	2	G
7771131	P238	R1	G->A	1	A
2887824	M343	R1b	C->A	2	A
14193384	CTS2134	R1b1	G->A	1	A
7960019	L822	R1b1	G->A	2	A
15510064	CTS4244	R1b1a	T->G	1	G
18026855	CTS8436	R1b1a	G->A	1	A
22889018	L754	R1b1a	G->A	1	A
15239181	CTS3876	R1b1a1a	G->C	1	C
18617596	CTS9018	R1b1a1a	C->T	1	T
19020340	L502	R1b1a1a	G->C	1	C
21249870	Y13208	R1b1a1a1	T->C	1	C
13657777	L777	R1b1a1a2	T->C	1	C
17594966	CTS7659	R1b1a1a2	C->G	1	G
17932763	CTS8321	R1b1a1a2	C->T	1	T
18381735	PF6482	R1b1a1a2	A->G	1	G
18865298	L753	R1b1a1a2	C->T	1	T
23242935	L1348	R1b1a1a2	G->C	1	C
28590278	CTS12478	R1b1a1a2	G->A	2	A
28771116	CTS12972	R1b1a1a2	C->G	1	G
7073423	CTS894	R1b1a1a2	G->A	1	A
8070532	PF6430	R1b1a1a2	T->A	2	A

Table G.26 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
9464078	PF6438	R1b1a1a2	C->T	3	T
8796078	M405	R1b1a1a2a1a1	C->T	2	T

TABLE G.27: Yleaf output of KD034, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
14231291	L1118	A0-T	T->C	2	C
16428812	L1128	A0-T	C->T	1	T
16661010	L1130	A0-T	T->G	1	G
21593345	L1143	A0-T	A->G	1	G
21739790	L1145	A0-T	C->T	2	T
22191266	L1155	A0-T	G->C	2	C
7590048	L1105	A0-T	C->T	2	T
15442970	M11301	A1b	T->C	1	T
16243116	Z11919	A1b	C->T	1	T
17472765	CTS7423	B2a1a1a1∼	G->A	1	A
10008803	L418	BT	C->G	1	G
10068588	Z40376	BT	A->G	1	G
13596520	Z40400	BT	T->A	2	A
13665248	Z40379	BT	G->A	2	A
13840089	Z40402	BT	T->C	1	C
13841677	Y8316	BT	A->G	1	G
14096790	M9081	BT	C->A	2	A
14108214	M9083	BT	T->A	1	A
14123744	M9087	BT	C->T	1	T
14472914	M9100	BT	C->T	1	T
14839376	M9117	BT	T->C	1	C
14886273	P97	BT	G->T	1	T
15068525	M9125	BT	A->T	1	T
15095089	M9126	BT	C->T	1	T
15112480	M9128	BT	C->T	1	T
15542606	M9141	BT	A->T	2	T
15905094	M9152	BT	G->T	1	T
16302343	M9165	BT	G->A	1	A

Table G.27 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
16542464	M9174	BT	A->C	2	C
16959883	M9189	BT	A->G	1	G
17041558	M9193	BT	G->A	1	A
17063750	M9195	BT	T->A	2	A
17298439	M9209	BT	G->A	1	A
17510373	L962	BT	C->T	1	T
17511829	M9219	BT	T->C	2	C
18958429	M9269	BT	C->T	1	T
19149592	M9280	BT	C->A	1	A
19306455	M9288	BT	A->C	1	C
19325165	M9290	BT	C->T	2	T
19416000	M9293	BT	C->G	2	G
20817229	Y8322	BT	G->A	1	A
21268114	M9312	BT	C->G	2	G
21289152	M9315	BT	T->G	1	G
21567686	M9331	BT	C->G	1	G
21646196	M9336	BT	G->A	1	A
22191243	L1061	BT	G->A	2	A
22625732	M9356	BT	C->T	1	T
22730777	M9361	BT	T->C	1	C
22817840	M9365	BT	G->C	2	C
23113295	M9378	BT	T->C	1	C
23237493	Z12129	BT	T->A	2	A
23405276	M9389	BT	T->C	1	C
23565114	M9394	BT	T->A	2	A
23757465	M9397	BT	G->C	1	C
2712956	M8947	BT	G->A	1	A
2842830	M8951	BT	A->G	1	G
28746408	Z17390	BT	G->C	2	C
28787653	Z40384	BT	G->C	2	C
2889399	M8954	BT	T->C	1	C
2899711	M8955	BT	A->G	1	G
6932831	L413	BT	G->A	1	A
7171532	M8977	BT	G->T	1	T
7221998	M8979	BT	C->A	2	A

Table G.27 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
7302697	M8983	BT	G->A	1	A
7531097	M8988	BT	T->C	1	C
7890655	M9003	BT	C->G	1	G
8021764	M9010	BT	G->A	1	A
8226191	M9017	BT	C->A	1	A
8446082	M9025	BT	A->T	1	T
8511682	M9026	BT	T->C	1	C
8520679	M9027	BT	C->A	1	A
8540845	M9028	BT	A->G	1	G
8594070	Z40389	BT	G->A	1	A
8635293	M9032	BT	C->A	1	A
8662142	M9034	BT	G->A	1	A
8677316	M9036	BT	C->G	1	G
8881498	M9042	BT	G->A	1	A
9106630	M9046	BT	A->G	1	G
9930675	Z40372	BT	C->T	1	T
13594339	Y1791	CT	C->T	1	T
13852599	PF629	CT	A->G	1	G
14439817	CTS2711	CT	G->A	1	A
14516781	M5661	CT	T->C	1	C
14813991	M168	CT	C->T	1	T
16465771	CTS5746	CT	T->G	1	G
16844116	M5706	CT	G->A	1	A
17301733	M5716	CT	T->C	1	C
17379868	CTS7257	CT	C->T	1	T
17750457	CTS7933	CT	C->T	2	T
17909505	M5738	CT	G->A	1	A
18179043	M5747	CT	C->A	1	A
18715075	M5751	CT	A->G	1	G
18974195	M5760	CT	C->T	1	T
19039750	M5762	CT	G->C	3	C
19132472	M5765	CT	G->A	1	A
21250559	M5775	CT	C->G	2	G
21267676	M5777	CT	C->T	1	T
21281606	M5778	CT	G->C	2	C

Table G.27 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
22631570	L1492	CT	A->G	1	G
22744945	M294	CT	C->T	1	T
23081753	CTS11358	CT	A->G	1	G
23105586	M5812	CT	C->A	1	A
24443898	M5825	CT	C->T	1	T
28650362	CTS12633	CT	T->A	1	A
6864267	M5585	CT	A->G	1	G
6975295	L1462	CT	C->G	2	G
7233061	Z40571	CT	G->A	1	A
7275087	CTS1217	CT	C->T	2	T
7358315	M5599	CT	A->G	1	G
7597185	M5603	CT	G->A	1	A
7646619	M5605	CT	G->C	1	C
7887815	M5615	CT	A->T	2	T
8183439	M5627	CT	G->A	1	A
8262951	M5629	CT	T->C	1	C
8396636	M5631	CT	G->A	1	A
8543804	M5633	CT	G->A	2	A
8867489	M5640	CT	T->C	2	C
9758726	PF342	CT	G->A	1	A
19066988	CTS9776	E1b1a1a1a1c1a1a3d3∼	G->A	1	A
19194067	PF2450	E1b1b1b1b1b1∼	G->A	1	A
14334396	F1767	F	G->T	1	T
14832620	M235	F	T->G	2	G
15323154	CTS3996	F	A->G	2	G
16364286	F2142	F	C->A	1	A
19205814	L470	F	A->T	4	T
28650343	CTS12632	F	C->G	1	G
6941218	P141	F	G->A	1	A
7726849	M3647	F	A->G	1	G
8474189	P160	F	A->C	2	C
6758698	Z30714	G2a2b2a1a1a2a2b1∼	C->T	1	T
8589031	F1329	GHIJK	C->T	1	T
18887101	M2974	H1a1	C->T	1	T
7202703	F929	HIJK	C->T	2	T

Table G.27 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
6753519	L15	IJK	A->G	1	G
7702973	M2683	IJK	T->A	2	A
21730257	M9	K	C->G	1	G
13840745	Y5523	L1∼	G->A	1	A
17643584	CTS7747	NO1	C->T	1	T
8474878	F1234	O2a2a1a2a1a2	C->T	1	T
10035664	FGC285	P	G->C	2	C
14679362	CTS3135	P	C->A	1	A
14933671	CTS3446	P	C->A	3	A
15648415	M1109	P	C->T	1	T
17374007	CTS7244	P	G->T	2	T
18720028	CTS9162	P	C->T	2	T
20812228	F556	P	T->C	1	C
21139399	M1255	P	C->A	1	A
21456883	M1257	P	G->A	2	A
2803717	CTS196	P	C->T	1	T
2817429	CTS216	P	C->G	1	G
6931040	PF5858	P	C->G	1	G
7996768	PF5867	P	G->A	1	A
8096754	M1188	P	T->C	2	C
14433100	P244	P1∼	G->A	2	A
17470112	P230	P1∼	G->A	1	A
17881230	P239	P1∼	G->C	1	C
8334875	P237	P1∼	A->G	1	G
13802432	YP800	Q1a1b1a∼	C->T	1	T
7778164	Y2048	Q1b1a	G->A	1	A
14300457	CTS2426	R	G->A	1	A
15667208	FGC1168	R	G->C	1	C
17334694	M718	R	G->T	1	T
17722802	CTS7876	R	G->A	1	A
17723850	CTS7880	R	C->T	1	T
18017528	F459	R	G->T	1	T
18066156	M734	R	C->T	1	T
22733758	L1225	R	C->G	1	G
22818334	L1347	R	C->T	1	T

Table G.27 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
24360964	F765	R	G->A	1	A
7177189	F63	R	G->A	1	A
7548900	F82	R	G->A	1	A
14366723	CTS2565	R1	C->T	2	T
16394489	CTS5611	R1	T->G	2	G
16742224	L875	R1	A->G	2	G
22750583	M306	R1	C->A	1	A
7132713	CTS997	R1	G->A	1	A
2887824	M343	R1b	C->A	2	A
14193384	CTS2134	R1b1	G->A	2	A
14226692	CTS2229	R1b1	T->A	3	A
15510064	CTS4244	R1b1a	T->G	1	G
17545608	CTS7585	R1b1a	G->T	1	T
18026855	CTS8436	R1b1a	G->A	1	A
22889018	L754	R1b1a	G->A	2	A
8214827	PF6249	R1b1a	C->T	2	T
16183412	PF6463	R1b1a1a	C->A	2	A
10008791	L150.1	R1b1a1a2	C->T	1	T
13657777	L777	R1b1a1a2	T->C	2	C
13816025	PF6444	R1b1a1a2	G->A	1	A
14136291	PF6452	R1b1a1a2	G->A	1	A
14317555	CTS2466	R1b1a1a2	G->A	2	A
15037433	CTS3575	R1b1a1a2	C->G	1	G
15740440	PF6462	R1b1a1a2	A->G	1	G
16971648	CTS6532	R1b1a1a2	T->G	3	G
17594966	CTS7659	R1b1a1a2	C->G	1	G
17932763	CTS8321	R1b1a1a2	C->T	1	T
18117193	CTS8627	R1b1a1a2	C->T	1	T
18381735	PF6482	R1b1a1a2	A->G	1	G
18865298	L753	R1b1a1a2	C->T	2	T
19054889	L757	R1b1a1a2	C->T	2	T
19417394	CTS10349	R1b1a1a2	A->C	3	C
20838224	PF6496	R1b1a1a2	A->G	1	G
21784286	L1350	R1b1a1a2	G->A	1	A
22190371	PF6509	R1b1a1a2	A->G	1	G

Table G.27 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
22796697	CTS10834	R1b1a1a2	T->C	1	C
7220727	F69	R1b1a1a2	A->G	3	G
7891188	PF6428	R1b1a1a2	T->C	1	C
6753511	L23	R1b1a1a2a	G->A	2	A
22200784	S245	R1b1a1a2a1a2c	C->G	1	G
24411932	Z260	R1b1a1a2a1a2c	G->T	1	T
22471235	S471	R1b1a1a2a1a2c1∼	A->T	1	T
15339857	S883	R1b1a1a2a1a2c1a4b2c	A->G	1	G
22732394	SK2055	R1b1b2a	G->A	1	A
18768123	Z41872	S1a3a	C->T	1	T

TABLE G.28: Yleaf output of KD036, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
6753296	L1095	A0-T	A->G	1	G
17716223	L1005	A1	G->A	1	A
14774366	Z17894	A1b	C->T	1	T
18759708	L1053	A1b	C->A	1	A
13202478	Z40398	BT	G->A	1	A
13596520	Z40400	BT	T->A	1	A
13806787	Z17371	BT	G->A	1	A
13841677	Y8316	BT	A->G	1	G
14013228	M9075	BT	T->C	1	C
14541174	M9103	BT	T->G	1	G
14588372	M9105	BT	G->A	1	A
14839376	M9117	BT	T->C	2	C
14980170	M9121	BT	C->T	1	T
16538095	M9173	BT	C->T	1	T
16542464	M9174	BT	A->C	2	C
17298439	M9209	BT	G->A	1	A
17511829	M9219	BT	T->C	1	C
17886479	M9237	BT	G->A	1	A
18113973	M9248	BT	G->A	1	A
18814791	M9263	BT	G->A	2	A

Table G.28 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
18917853	M9267	BT	C->T	1	T
19207843	M9282	BT	C->A	1	A
19235631	M9285	BT	C->T	2	T
19306455	M9288	BT	A->C	1	C
19325165	M9290	BT	C->T	1	T
19445252	M9295	BT	A->C	1	C
21069381	M9300	BT	G->A	1	A
21486590	M9325	BT	T->C	1	C
21490191	M9326	BT	A->G	1	G
21637759	M9335	BT	A->G	1	G
21646196	M9336	BT	G->A	1	A
22719912	M9357	BT	G->A	1	A
22730555	M9360	BT	A->G	1	G
22748506	M299	BT	T->G	1	G
28746408	Z17390	BT	G->C	1	C
28787653	Z40384	BT	G->C	2	C
7171532	M8977	BT	G->T	2	T
7221998	M8979	BT	C->A	1	A
7531097	M8988	BT	T->C	1	C
8635293	M9032	BT	C->A	2	A
9106630	M9046	BT	A->G	1	G
14149520	M5652	CT	T->A	1	A
15097073	CTS3662	CT	G->A	2	A
16465771	CTS5746	CT	T->G	1	G
17630903	M5724	CT	C->T	1	T
17742131	CTS7922	CT	C->T	1	T
17909505	M5738	CT	G->A	1	A
22639802	PF1205	CT	C->T	1	T
23090404	M5809	CT	G->A	1	A
23234852	Y1571	CT	G->A	1	A
23567930	M5823	CT	C->T	1	T
24443898	M5825	CT	C->T	1	T
7597185	M5603	CT	G->A	1	A
7707388	M5606	CT	T->G	1	G
7740081	M5610	CT	G->A	2	A

Table G.28 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
8526565	M5632	CT	G->A	1	A
8635526	M5636	CT	C->A	1	A
9894285	Y18536	E1b1a1a1a2a1a3b1a2a4∼	A->G	1	G
14996960	CTS3536	F	A->G	1	G
15323154	CTS3996	F	A->G	1	G
15651438	CTS4443	F	A->T	1	T
16242316	P163	F	A->T	1	T
21147336	F3136	F	T->C	1	C
8589031	F1329	GHIJK	C->T	1	T
13804066	Z16985	I	G->C	1	C
16567253	CTS5946	I	A->G	1	G
16780748	CTS6265	I	C->G	1	G
17924382	CTS8300	I	T->A	1	A
17949402	CTS8345	I	C->G	1	G
19233673	CTS10058	I	A->G	1	G
21130059	PF3797	I	A->G	1	G
21402723	PF3800	I	A->G	1	G
21452125	PF3803	I	A->G	1	G
21839183	PF3814	I	A->G	1	G
8484606	PF3661	I	C->A	1	A
8643763	PF3665	I	A->G	1	G
13822499	Z2747	I1	A->G	1	G
14047137	CTS1748	I1	A->G	1	G
14071004	L841	I1	T->A	1	A
14402997	CTS2644	I1	G->A	1	A
14782128	CTS3268	I1	T->A	1	A
15754964	CTS4656	I1	T->C	1	C
18759669	L125	I1	T->C	1	C
19235426	CTS10062	I1	G->A	1	A
21288153	Z2844	I1	C->G	1	G
21402851	Z2848	I1	C->T	1	T
23153786	CTS11534	I1	A->G	1	G
2753908	CTS136	I1	T->G	1	G
6677619	FGC2426	I1	T->G	1	G
8554437	Z2727	I1	C->T	1	T

Table G.28 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
13676296	FGC7747	I1∼	T->G	3	G
9847423	S244	I1a2	T->C	1	C
14144593	P129	IJ	A->G	1	G
16244914	CTS5387	IJ	G->A	1	A
18633084	F2794	IJ	C->A	1	A
19324937	CTS10200	IJ	G->A	1	A
21389837	Z40732	IJ	A->G	1	G
21461202	PF3562	IJ	G->A	1	A
22909101	F3368	IJ	C->T	1	T
7702973	M2683	IJK	T->A	1	A
22928068	F632	O2a1c1a1a1	G->A	1	A
8510540	M971	Q1a2a1b	C->T	1	T
16938335	Y1371	R2a2b1b2b3b1	G->A	1	A

TABLE G.29: Yleaf output of KD038, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
16661010	L1130	A0-T	T->G	1	G
18147303	L1135	A0-T	C->A	1	A
21593345	L1143	A0-T	A->G	1	G
21739790	L1145	A0-T	C->T	1	T
2887280	L1089	A0-T	G->C	3	C
16773792	L1004	A1	T->C	1	C
17716223	L1005	A1	G->A	2	A
14774366	Z17894	A1b	C->T	2	T
18180294	L1009	A1b	A->G	2	G
18759708	L1053	A1b	C->A	1	A
10008803	L418	BT	C->G	1	G
13596520	Z40400	BT	T->A	1	A
13668726	Z40380	BT	T->C	2	C
13806787	Z17371	BT	G->A	1	A
13821858	Z40381	BT	C->G	1	G
13861285	Z40404	BT	G->A	1	A
14013228	M9075	BT	T->C	1	C

Table G.29 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
14075450	M9080	BT	C->T	3	T
14096790	M9081	BT	C->A	1	A
14108214	M9083	BT	T->A	1	A
14123744	M9087	BT	C->T	1	T
14344506	M9094	BT	C->G	1	G
14461702	M9099	BT	G->A	3	A
14588372	M9105	BT	G->A	2	A
14723798	M9111	BT	T->C	1	C
14980170	M9121	BT	C->T	3	T
15068525	M9125	BT	A->T	1	T
15112480	M9128	BT	C->T	2	T
15132194	M9129	BT	T->G	3	G
15381787	M9135	BT	C->G	2	G
15450357	M9138	BT	C->A	1	A
15451147	M9139	BT	T->A	1	A
16542464	M9174	BT	A->C	3	C
16946901	M9188	BT	G->A	1	A
16959883	M9189	BT	A->G	2	G
17041558	M9193	BT	G->A	1	A
17063750	M9195	BT	T->A	2	A
17147159	M9197	BT	C->T	2	T
17192307	M9199	BT	G->C	1	C
17270785	M9204	BT	C->A	4	A
17324151	M9210	BT	A->G	4	G
17353147	L1220	BT	A->G	3	G
17372321	M9213	BT	G->A	3	A
17388578	M9214	BT	A->G	3	G
17510373	L962	BT	C->T	3	T
17511829	M9219	BT	T->C	1	C
17670179	M9226	BT	C->T	1	T
17670193	M9227	BT	G->A	1	A
17728330	M9230	BT	T->C	1	C
17886479	M9237	BT	G->A	2	A
17940510	M9240	BT	C->A	1	A
17966904	M9242	BT	C->A	4	A

Table G.29 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
18016799	Y9422	BT	T->A	2	A
18053867	M9244	BT	T->A	3	A
18067088	M9245	BT	T->A	1	A
18262817	Y8320	BT	G->A	2	A
18628317	M9260	BT	G->A	3	A
18814791	M9263	BT	G->A	3	A
18917853	M9267	BT	C->T	1	T
18958429	M9269	BT	C->T	1	T
19140679	M9278	BT	C->T	1	T
19149592	M9280	BT	C->A	2	A
19207843	M9282	BT	C->A	1	A
19233322	M9284	BT	T->C	1	C
19249565	M9286	BT	C->G	3	G
19306455	M9288	BT	A->C	1	C
19325165	M9290	BT	C->T	2	T
19337259	M9291	BT	T->C	2	C
19445252	M9295	BT	A->C	2	C
19457726	M9296	BT	G->A	2	A
21069381	M9300	BT	G->A	1	A
21122071	M9303	BT	A->G	2	G
21268114	M9312	BT	C->G	1	G
21311141	M9317	BT	C->T	2	T
21326695	M9319	BT	A->G	2	G
21391240	M9321	BT	G->A	2	A
21567686	M9331	BT	C->G	2	G
21627877	M9334	BT	C->T	2	T
21637759	M9335	BT	A->G	1	G
21665439	M9338	BT	C->A	1	A
21698424	M9340	BT	G->C	1	C
21938158	M94	BT	C->A	1	A
22062727	M9347	BT	C->T	2	T
22191243	L1061	BT	G->A	1	A
22625732	M9356	BT	C->T	1	T
22817840	M9365	BT	G->C	4	C
22931502	M9370	BT	C->T	4	T

Table G.29 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
23010471	M9373	BT	C->T	1	T
23070865	M9376	BT	A->C	2	C
23190598	Y8325	BT	A->T	2	T
23199069	Z40409	BT	G->T	1	T
23201899	Z40411	BT	A->C	2	C
23237493	Z12129	BT	T->A	1	A
23240528	Z40412	BT	C->G	4	G
23565114	M9394	BT	T->A	2	A
23991003	M9405	BT	G->A	2	A
2669416	L1071	BT	A->G	1	G
2712956	M8947	BT	G->A	3	A
2790584	V202	BT	G->C	1	C
28746408	Z17390	BT	G->C	3	C
28787653	Z40384	BT	G->C	2	C
2899711	M8955	BT	A->G	1	G
6859691	M8969	BT	C->T	3	T
6908261	V29	BT	A->G	2	G
6993202	M8972	BT	C->T	1	T
7032507	Z17366	BT	C->T	1	T
7126782	Y8310	BT	G->A	2	A
7302697	M8983	BT	G->A	1	A
7531097	M8988	BT	T->C	1	C
7645425	V235	BT	A->G	1	G
7661695	M8994	BT	A->C	2	C
7890655	M9003	BT	C->G	1	G
7904101	M9005	BT	G->A	1	A
7907933	M9006	BT	G->A	1	A
8021764	M9010	BT	G->A	1	A
8180027	Z40388	BT	A->G	2	G
8330559	M9019	BT	T->C	1	C
8337827	M9020	BT	T->C	1	C
8511682	M9026	BT	T->C	1	C
8540845	M9028	BT	A->G	3	G
8594070	Z40389	BT	G->A	2	A
8635293	M9032	BT	C->A	3	A

Table G.29 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
9869354	M11760	BT	C->A	1	A
9891742	M9066	BT	C->T	1	T
23629400	Z7989	C1a1a2	T->C	1	C
15203676	CTS3818	CF	A->G	5	G
13213798	Z17710	CT	G->T	1	T
14074463	Y1525	CT	C->T	1	T
14079528	L957	CT	C->T	1	T
14149010	CTS1996	CT	G->A	1	A
14149520	M5652	CT	T->A	1	A
14207088	M5656	CT	C->A	1	A
14439817	CTS2711	CT	G->A	3	A
15097073	CTS3662	CT	G->A	1	A
15614295	CTS4364	CT	C->T	1	T
16346371	CTS5532	CT	C->T	1	T
16465771	CTS5746	CT	T->G	4	G
16844116	M5706	CT	G->A	1	A
16887784	M5709	CT	G->C	2	C
17104433	M5712	CT	A->C	2	C
17301733	M5716	CT	T->C	2	C
17515027	M5721	CT	G->A	1	A
17670046	M5726	CT	C->A	1	A
17750457	CTS7933	CT	C->T	5	T
17753650	M5730	CT	A->G	1	G
17853378	M5734	CT	C->T	1	T
17894575	CTS8243	CT	C->T	1	T
17909505	M5738	CT	G->A	1	A
17975155	FGC33852	CT	G->T	1	T
18077583	CTS8542	CT	T->C	1	C
18179043	M5747	CT	C->A	1	A
18616467	CTS9014	CT	T->A	1	A
18715075	M5751	CT	A->G	2	G
18974195	M5760	CT	C->T	1	T
19212465	L1480	CT	A->G	1	G
19261216	CTS10110	CT	C->A	1	A
19503081	CTS10512	CT	C->A	1	A

Table G.29 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
21250559	M5775	CT	C->G	1	G
21267676	M5777	CT	C->T	1	T
21281606	M5778	CT	G->C	1	C
21429988	M5783	CT	A->G	1	G
22529475	Y1567	CT	C->G	1	G
23081753	CTS11358	CT	A->G	1	G
23090404	M5809	CT	G->A	2	A
23234852	Y1571	CT	G->A	4	A
23561669	M5822	CT	A->T	1	T
23567930	M5823	CT	C->T	4	T
24443898	M5825	CT	C->T	1	T
28789895	Y1580	CT	A->T	1	T
28790215	PF1337	CT	G->C	1	C
6831340	CTS543	CT	C->T	2	T
6864267	M5585	CT	A->G	2	G
6975295	L1462	CT	C->G	1	G
6991000	M5588	CT	C->A	1	A
7041647	M5590	CT	G->A	1	A
7256433	CTS1181	CT	A->T	1	T
7275087	CTS1217	CT	C->T	1	T
7358315	M5599	CT	A->G	1	G
7416169	M5600	CT	T->C	1	C
7597185	M5603	CT	G->A	1	A
7707388	M5606	CT	T->G	1	G
7738840	M5609	CT	T->G	1	G
7778691	M5611	CT	C->T	1	T
7796647	M5613	CT	C->T	4	T
7887815	M5615	CT	A->T	1	T
8080397	M5622	CT	T->A	1	A
8183439	M5627	CT	G->A	1	A
8526565	M5632	CT	G->A	4	A
9094343	M5642	CT	C->T	1	T
9758726	PF342	CT	G->A	2	A
15353877	Z15227	E1a2a2	C->T	1	T
14448405	CTS2727	E1b1b1∼	G->A	1	A

Table G.29 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
14334396	F1767	F	G->T	1	T
14832620	M235	F	T->G	1	G
15323154	CTS3996	F	A->G	1	G
15651438	CTS4443	F	A->T	1	T
16401339	F2155	F	C->T	1	T
17493513	P158	F	C->T	5	T
17762668	F2587	F	T->C	1	C
19205814	L470	F	A->T	3	T
19220444	F2985	F	A->G	1	G
21618856	P135	F	C->T	2	T
2722506	F719	F	G->A	2	A
2756471	M3637	F	A->T	1	T
8131538	F3692	F	G->T	7	T
8572150	F1320	F	G->A	3	A
8602415	P146	F	C->T	1	T
9108252	P187	F	G->T	3	T
8682185	Z40765	G2a2b2a1a1b1a1a2a1b∼	G->A	1	A
14237670	CTS2254	GHIJK	C->T	1	T
6804093	Z34878	H1a1a4b2c4∼	G->A	1	A
13678448	Z12773	H3a2	G->A	1	A
15984996	Z13917	H3b	C->T	1	T
7202703	F929	HIJK	C->T	1	T
13835003	FGC2415	I	T->C	3	C
14073053	CTS1800	I	G->A	1	A
14286853	CTS2387	I	T->C	3	C
14352669	CTS2536	I	G->A	1	A
14484379	P38	I	A->C	1	C
14884646	CTS3383	I	C->T	1	T
14884659	CTS3384	I	A->C	1	C
14974451	L1197	I	C->T	1	T
15479899	CTS4209	I	T->A	1	A
15536870	CTS4273	I	C->T	1	T
17497181	CTS7469	I	C->A	1	A
17511797	CTS7502	I	A->G	1	G
17692855	CTS7831	I	T->A	2	A

Table G.29 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
17940414	CTS8333	I	G->A	1	A
17949402	CTS8345	I	C->G	1	G
18018313	CTS8420	I	C->A	1	A
18257568	CTS8876	I	G->A	1	A
18394743	L751	I	A->G	1	G
18582617	CTS8963	I	C->T	2	T
18992894	CTS9618	I	T->C	2	C
19104986	CTS9860	I	G->A	1	A
19233673	CTS10058	I	A->G	1	G
21689728	FGC2412	I	A->G	2	G
21839183	PF3814	I	A->G	1	G
21841289	PF3815	I	G->T	1	T
21939618	PF3817	I	G->A	2	A
23401471	CTS11979	I	C->T	1	T
6926038	CTS646	I	T->A	2	A
7321418	CTS1301	I	C->T	1	T
7642823	FGC2416	I	G->T	2	T
7681156	PF3640	I	T->A	1	A
8262092	FGC2413	I	C->T	1	T
8484606	PF3661	I	C->A	2	A
8485677	FI3	I	C->A	1	A
8536868	L758	I	C->G	1	G
8643763	PF3665	I	A->G	3	G
8728974	PF3666	I	T->G	1	G
8873160	FI4	I	G->T	3	T
19077754	L181	I2a2	G->T	2	T
22725379	L35	I2a2	C->A	1	A
16199051	L39	I2a2b	T->C	2	C
3436239	L272.3	I2a2b	A->T	1	T
14144593	P129	IJ	A->G	1	G
17200869	CTS6932	IJ	G->A	1	A
19166861	F4281	IJ	T->C	1	C
21225770	FGC1566	IJ	C->G	1	G
21321273	FGC1571	IJ	A->T	3	T
21389837	Z40732	IJ	A->G	3	G

Table G.29 continued from previous page

Position	Marker	Haplogroup	Mutation	Reads	Called base
22909101	F3368	IJ	C->T	2	T
7694266	FGC1564	IJ	G->A	1	A
8888300	F1450	IJ	G->A	3	A
21571895	M2696	IJK	G->A	2	A
7173143	L16	IJK	G->A	1	A
7702973	M2683	IJK	T->A	1	A
15963515	Y1079	Q1b	C->T	1	T
19068749	BZ314	Q1b1a1c2∼	G->T	1	T
21583368	Z41479	S1a1b1d	C->T	1	T

TABLE G.6: Yleaf output of KD049, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
15426344	L1124	A0-T	T->C	1	C
6859819	L1101	A0-T	T->C	1	C
21219262	FGC27383	A00	G->A	1	A
6788449	Z11900	A1b	T->C	1	C
17063750	M9195	BT	T->A	1	A
17151394	M9198	BT	C->T	1	T
17623760	M9223	BT	G->A	1	A
23201899	Z40411	BT	A->C	1	C
23244049	M9379	BT	G->C	1	C
6908261	V29	BT	A->G	1	G
7661695	M8994	BT	A->C	2	C
8337827	M9020	BT	T->C	1	C
8613037	L1062	BT	G->A	1	A
14671991	CTS3120	CT	C->A	1	A
21683734	M5788	CT	C->T	1	T
23090404	M5809	CT	G->A	1	A
2733618	M8948	CT	C->A	1	A
6975295	L1462	CT	C->G	1	G
7275087	CTS1217	CT	C->T	1	T
8396636	M5631	CT	G->A	1	A
17145596	Z17453	E1a1	C->T	1	T
18178091	P161	F	G->A	1	A
21618856	P135	F	C->T	1	T
20813298	GG297	G1a2b	C->T	1	T
14976485	Z13538	H3	G->A	1	A
17424807	CTS7329	I	C->T	1	T
21452125	PF3803	I	A->G	1	G
8267857	L578	I	G->A	1	A
8888300	F1450	IJ	G->A	1	A
14887209	F4178	O2b2	G->A	1	A

TABLE G.7: Yleaf output of KD057, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
2756519	V187	BT	T->C	1	C
14124435	M5650	CT	G->A	1	A
18940202	M5754	CT	A->G	1	G
2744386	CTS125	CT	G->T	1	T
2756471	M3637	F	A->T	1	T
18257568	CTS8876	I	G->A	1	A
8382265	FI2	I	C->G	1	G

TABLE G.8: Yleaf output of KD059, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
18147303	L1135	A0-T	C->A	1	A
6847637	L1098	A0-T	C->A	1	A
13213798	Z17710	CT	G->T	1	T
8867489	M5640	CT	T->C	1	C
16751000	CTS6231	I	C->T	1	T
19048724	L226	R1b1a1a2a1a2c1a4b2a	C->T	1	T

TABLE G.16: Yleaf output of KD008, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
14096790	M9081	BT	C->A	2	A
19235631	M9285	BT	C->T	1	T
22817840	M9365	BT	G->C	2	C
23190598	Y8325	BT	A->T	2	T
2669416	L1071	BT	A->G	1	G
6932831	L413	BT	G->A	1	A
8635293	M9032	BT	C->A	1	A
14603298	F1804	C	C->T	1	T
13852599	PF629	CT	A->G	1	G
17301733	M5716	CT	T->C	1	C
15124129	CTS3716	E1a	C->T	1	T
8442341	F1209	F	A->G	1	G
21406999	L764	I1	C->T	2	T
21562105	A9128	I1a2a1a2∼	T->A	1	A

TABLE G.17: Yleaf output of KD010, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
16773792	L1004	A1	T->C	1	C
13668726	Z40380	BT	T->C	1	C
14069414	M9079	BT	C->T	1	T
14075450	M9080	BT	C->T	1	T
16808769	Z40393	BT	T->C	1	C
19457726	M9296	BT	G->A	1	A
21268114	M9312	BT	C->G	2	G
6932831	L413	BT	G->A	1	A
8881498	M9042	BT	G->A	1	A
18905190	CTS9458	CT	G->A	1	A
18940202	M5754	CT	A->G	1	G
23298645	CTS11827	CT	T->C	1	C
22688731	L507	E	G->C	1	C
7340243	L535	E1b1∼	C->A	1	A
22770232	CTS10775	E1b1a1∼	G->T	1	T
14696327	CTS3171	E1b1b1a	C->A	1	A

TABLE G.18: Yleaf output of KD013, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
15256839	M9133	BT	A->G	1	G
15261255	Z40392	BT	C->T	1	T
15542606	M9141	BT	A->T	1	T
16403627	M9169	BT	G->A	1	A
16913677	M9187	BT	T->C	1	C
16959883	M9189	BT	A->G	1	G
18585407	M9255	BT	T->G	1	G
19233322	M9284	BT	T->C	1	C
21122071	M9303	BT	A->G	1	G
6779907	M8967	BT	G->A	1	A
7021041	Z17365	BT	G->T	1	T
8635293	M9032	BT	C->A	1	A
16903417	SK1061	C2b1a2a1∼	G->A	1	A
17909505	M5738	CT	G->A	1	A
18948988	M5756	CT	T->C	1	C
7233061	Z40571	CT	G->A	1	A
14334396	F1767	F	G->T	1	T
8227605	M3652	F	A->G	1	G
8442341	F1209	F	A->G	1	G
22729194	CTS10721	G	C->T	1	T
14829196	CTS3321	R1	C->T	1	T
7900883	FGC41	R1b1a	C->A	1	A
17594966	CTS7659	R1b1a1a2	C->G	1	G
17813541	CTS8052	R1b1a1a2	C->T	1	T
9084870	PF6540	R1b1a1a2a1a	G->T	1	T

TABLE G.19: Yleaf output of KD014, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
13668726	Z40380	BT	T->C	1	C
14713690	M9109	BT	T->A	2	A
17670179	M9226	BT	C->T	1	T
17670193	M9227	BT	G->A	1	A
18053867	M9244	BT	T->A	1	A
21665439	M9338	BT	C->A	2	A
24377975	M9406	BT	G->T	1	T
8226191	M9017	BT	C->A	1	A
8635293	M9032	BT	C->A	1	A
9381562	Y8311	BT	G->A	1	A
14197867	P143	CF	G->A	1	A
24443898	M5825	CT	C->T	1	T
6864267	M5585	CT	A->G	1	G
7041647	M5590	CT	G->A	1	A
7778691	M5611	CT	C->T	1	T
21782675	BY8693	E1b1b1a1a2a3∼	C->T	1	T
8131538	F3692	F	G->T	1	T
21139399	M1255	P	C->A	1	A
24443836	F680	P	G->T	1	T
8096754	M1188	P	T->C	1	C
9823794	F4832	Q1a1b1∼	C->T	1	T
18066156	M734	R	C->T	1	T
7132713	CTS997	R1	G->A	1	A
7073423	CTS894	R1b1a1a2	G->A	1	A
18248698	P311	R1b1a1a2a1a	A->G	1	G

TABLE G.20: Yleaf output of KD017, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
14096790	M9081	BT	C->A	1	A
15063698	M9124	BT	C->A	1	A
15381787	M9135	BT	C->G	1	G
21259569	M9311	BT	C->T	1	T
21541028	Z40396	BT	C->T	1	T
22025458	M9346	BT	C->T	1	T
7171532	M8977	BT	G->T	1	T
9930675	Z40372	BT	C->T	1	T
10049308	Z17708	CT	A->G	1	G
14014443	M5647	CT	C->T	1	T
14149520	M5652	CT	T->A	1	A
19261216	CTS10110	CT	C->A	1	A
22631570	L1492	CT	A->G	1	G
28650362	CTS12633	CT	T->A	1	A
7256433	CTS1181	CT	A->T	1	T
7597185	M5603	CT	G->A	1	A
8183439	M5627	CT	G->A	1	A
9758726	PF342	CT	G->A	1	A
17762668	F2587	F	T->C	1	C
28650343	CTS12632	F	C->G	1	G
18720028	CTS9162	P	C->T	1	T
21298239	M1256	P	C->G	1	G
8290951	M1191	P	A->T	1	T
17285993	P224	R	C->T	1	T
17930099	CTS8311	R	C->A	1	A
21263029	M764	R	G->A	1	A
22818334	L1347	R	C->T	1	T
15510064	CTS4244	R1b1a	T->G	1	G
17461478	CTS7400	R1b1a1a2	T->C	1	C
18095336	CTS8591	R1b1a1a2	A->C	1	C
14641193	L52	R1b1a1a2a1a	C->T	1	T
17589518	CTS7650	R1b1a1a2a1a	C->T	1	T
9084870	PF6540	R1b1a1a2a1a	G->T	1	T

TABLE G.21: Yleaf output of KD018, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
15068525	M9125	BT	A->T	1	T
15095089	M9126	BT	C->T	1	T
15218792	Z40383	BT	T->C	1	C
21490191	M9326	BT	A->G	1	G
23240528	Z40412	BT	C->G	2	G
8446082	M9025	BT	A->T	1	T
8869241	M9041	BT	G->A	1	A
22639802	PF1205	CT	C->T	1	T
15323154	CTS3996	F	A->G	1	G
8442341	F1209	F	A->G	1	G
17974176	Z13562	H3	G->A	1	A
7173143	L16	IJK	G->A	2	A
7702973	M2683	IJK	T->A	1	A
16713287	F359	P	G->A	1	A
2817429	CTS216	P	C->G	1	G
21263029	M764	R	G->A	1	A
2887824	M343	R1b	C->A	1	A
21312064	PF6498	R1b1a1a	C->A	1	A
24444622	L1351	R1b1a1a2	C->T	1	T
14641193	L52	R1b1a1a2a1a	C->T	1	T

TABLE G.22: Yleaf output of KD020, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
16676762	M9178	BT	T->C	1	C
17372321	M9213	BT	G->A	1	A
21665439	M9338	BT	C->A	2	A
15792531	CTS4740	CT	G->A	2	A
16542862	M5698	CT	A->T	1	T
18905190	CTS9458	CT	G->A	1	A
22067044	M5794	CT	A->T	1	T
21139399	M1255	P	C->A	1	A
9885733	M1199	P	A->G	1	G
22750583	M306	R1	C->A	1	A
18248698	P311	R1b1a1a2a1a	A->G	1	G
2836431	CTS241	R1b1a1a2a1a2c1a	A->C	1	C

TABLE G.24: Yleaf output of KD022, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
15218792	Z40383	BT	T->C	1	C
17372321	M9213	BT	G->A	1	A
18900525	M9266	BT	C->T	1	T
7621889	M8993	BT	C->T	1	T
7907933	M9006	BT	G->A	1	A
17104433	M5712	CT	A->C	1	C
17630903	M5724	CT	C->T	1	T
19407727	M5769	CT	C->G	1	G
15286480	PF6459	R1b1a1a	G->C	1	C
21312064	PF6498	R1b1a1a	C->A	1	A
23973367	Z326	R1b1a1a2a1a1c2b2b1a	C->T	1	T

TABLE G.30: Yleaf output of KD040, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
16332337	M9166	BT	C->T	1	T
16694398	M9180	BT	C->T	1	T
7661695	M8994	BT	A->C	1	C
9930675	Z40372	BT	C->T	2	T
14043535	M5648	CT	A->C	1	C
2756471	M3637	F	A->T	1	T
7726849	M3647	F	A->G	1	G
15472863	P131	K∼	C->T	1	T
16535034	CTS5884	P	A->G	1	G
8667179	PF6435	R1b1a1a2	A->G	1	G
3274923	L478	R1b1a1a2a	A->C	1	C

TABLE G.31: Yleaf output of KD041, filtered for derived alleles with call percentage 100, excluding positions with low coverage and at the last three bases of reads.

Position	Marker	Haplogroup	Mutation	Reads	Called base
17393971	M9215	BT	C->T	1	T
7531097	M8988	BT	T->C	1	C
8009052	M9009	BT	T->C	1	C
7358315	M5599	CT	A->G	1	G
8572150	F1320	F	G->A	1	A
21571895	M2696	IJK	G->A	1	A
20837553	P128	K∼	C->T	1	T
10077460	PF5887	P	A->G	1	G
18028661	P282	P1∼	A->G	1	G
6868118	F47	R	G->A	1	A
18180446	L500	R1b1a1a2	C->A	1	A
18719565	PF6485	R1b1a1a2	T->C	1	C

Appendix H

ADMIXTURE – Supplementary Material

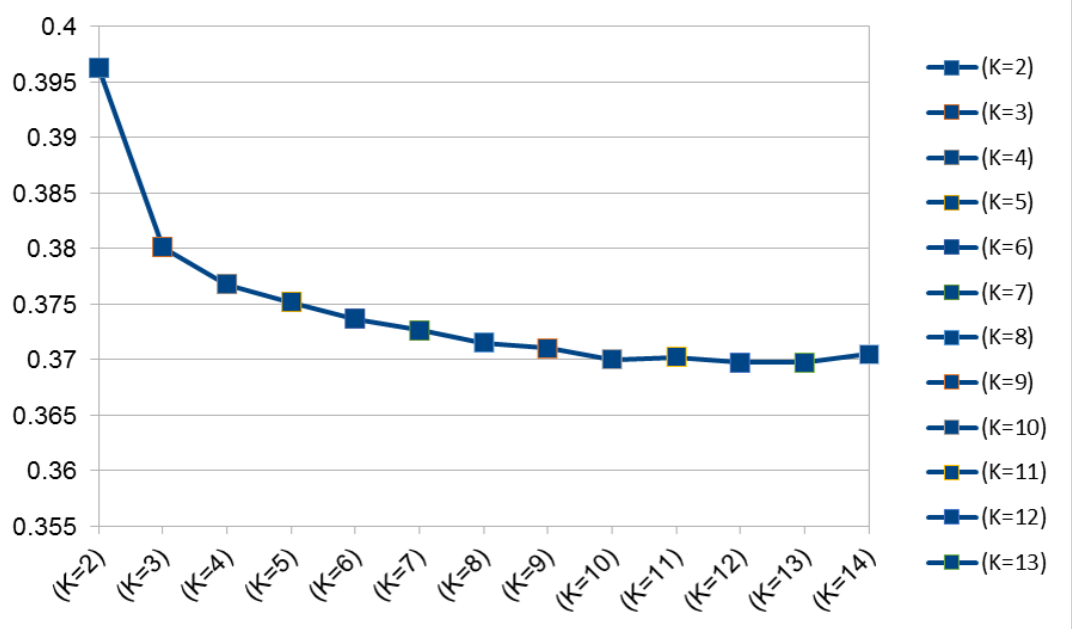

FIGURE H.1: CV-error plot for ADMIXTURE analysis including all samples analysed, ancient British data from Olalde et al. (2018), and the entire dataset used in Lazaridis et al. (2016), run for *K*2 to *K*14.

TABLE H.1: CV-errors for ADMIXTURE analysis including all samples analysed, ancient British data from Olalde et al. (2018), and the entire dataset used in Lazaridis et al. (2016), run for *K*2 to *K*14.

K	CV-Error
(K=2)	0.39627
(K=3)	0.38012
(K=4)	0.37676
(K=5)	0.37515
(K=6)	0.37368
(K=7)	0.37263
(K=8)	0.37151
(K=9)	0.37102
(K=10)	0.37000
(K=11)	0.37023
(K=12)	**0.36976**
(K=13)	**0.36976**
(K=14)	0.37047

Appendix I

Additional data

I.1 Carsington Pasture Cave

Carsington Pasture Cave (CPC), Derbyshire, was excavated by A.T. Chamberlain in 1998-1999. It contained 34 human burials, consisting of 16 adults and 18 juveniles, of which four mandible/maxilla fragments were given to Dr. C.J. Edwards for aDNA analysis.

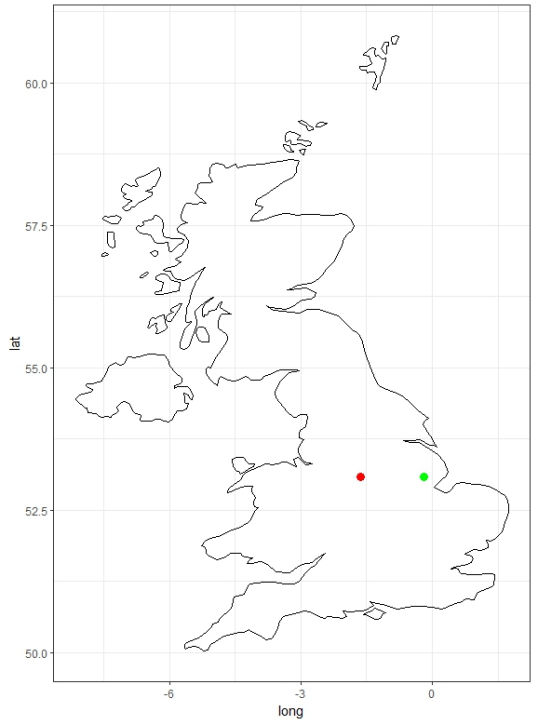

FIGURE I.1: The location of the Carsington Pasture Cave (red) and Beeston Tor (green) within the British Isles.

16 samples were later given to Dr. Tom Booth at the Natural History Museum, London. The human remains found in the cave dated to the Neolithic and Iron Age, and three of the four mandible/maxilla samples were analysed in this study.

TABLE I.1: Archaeological information of the Carsington Pasture Cave and Beeston Tor individuals analysed.

DNA code	DNA sample	Site information	Sample code	Age	Sex	Skeletal element(s)	Era	Date (uncal. bp)/ (Dating lab #)
CE001	LRM2	Carsington Pasture Cave, Derbyshire	CPC98-312	>45 years	M	mandible (with LM3, RM2 and RM3)	early/middle Neolithic	4795 ± 29 (OxA-28867)
CE003	UM2	Carsington Pasture Cave, Derbyshire	CPC02-Y-062	17-25 years	M	maxilla with teeth	early Iron Age	2460 ± 30 (OxA-29233)
CE004	LRM2	Carsington Pasture Cave, Derbyshire	CPC98-002	35-45 years	M	mandible with teeth	middle Iron Age	2240 ± 24 OxA-28865
CE002	ULP2	Beeston Tor, Staffordshire	BCX09-136	17-25 years	x	left maxilla (with P1, P2, M1, M2 and M3)	Bronze Age/ Iron Age	x

The Neolithic individual (CE001) was determined as genetically male, with T2c1d1 as his mitochondrial haplogroup. This is the same mtDNA haplogroup as determined for a Neolithic sample from Carsington Pasture Cave reported in the major Bell Beaker study this year (Olalde et al., 2018), and it is likely that the two samples belonged to the same individual. Haplogroup T2c1 comprises of the majority of T2c, and originated in the Near East ∼18.5 kya and dispersed into Europe ∼10 kya (Pala et al., 2012). It is most common in the Near East, specifically in the Gulf region, and is found in the Levant and Mediterranean Europe, however, it has a low frequency in the rest of Europe (Pala et al., 2012). Other reported Neolithic individuals belonging to this haplogroup were found in Holm of Papa, Westray North, Orkney, Scotland (Olalde et al., 2018), France (Olalde et al., 2018), Germany (Lazaridis et al., 2014; Haak et al., 2015), Hungary (Lipson et al., 2017), and Iran (Broushaki et al., 2016). The Early Iron Age CPC individual (CE003) was determined as female and belonged to mitochondrial haplogroup X2b4a1. Although ancient individuals reported to belong to X2b4 were found in Bell Beakers from France (Olalde et al., 2018), Late Neolithic and Alemanni Germany (Haak et al., 2015; O'Sullivan et al., 2018), and Bell Beakers in the Netherlands (Olalde et al., 2018), the mitochondrial haplogroup X2b4a has not been reported from the British Isles previous to this study. Alongside this Early Iron Age individual, one of the Anglo-Saxons from West Heslerton had haplotype X2b4a, and so these two people are the first evidence for this haplogroup in ancient periods in Britain. One-third of the European sequences within X2 are encompassed by subclades X2b and X2c, with X2b also including some North African sequences (Reidla et al., 2003). The diversity of subclade X2b in West Eurasia and North Africa is consistent with a post-glacial population expansion (Reidla et al., 2003). The Middle Iron Age individual (CE004) was a male and belonged to mtDNA haplogroup H10b, which has not yet been reported in ancient individuals. Ancient individuals belonging to H10 or other subclades of it have been reported from Bell Beakers in the Czech Republic (Olalde et al., 2018), Neolithic Croatia (Mathieson et al., 2018), Neolithic, Bell Beaker and Alemanni Germany (Brotherton et al., 2013; Olalde et al., 2018; O'Sullivan et al., 2018), and Late Bronze Age Latvia (Mittnik et al., 2018).
Although the Y-chromosome coverage was very low and only few Y-chromosome haplogroup predictive markers were found, it was possible to type the Neolithic individual as haplogroup I, whereas the Middle Iron Age individual was most likely R1b, possibly subgroup R1b1a1a2a1a. When comparing the Neolithic with the published data from the same individual (Olalde et al.,

2018), the predicted haplogroup was I2a2a1a1a.

I.2 Beeston Tor

An individual (CE002) from Beeston Tor, Staffordshire, England, from the Bronze Age/Iron Age was sequenced to a very low coverage of only 0.002x, which made sex identification inconclusive, and genome-wide analysis impossible (due to the low number of overlapping SNPs <10,000). However, the mitochondrial genome was sequenced to 1.2568x coverage, which was enough to assign the individual to mitochondrial haplogroup U8b. Mitochondrial haplogroup U8b is an eastern European haplogroup, found in the Neolithic in Croatia (Mathieson et al., 2018), the Czech Republic (Olalde et al., 2018), Hungary (Lipson et al., 2017), Scotland (Olalde et al., 2018), and the Ukraine (Mathieson et al., 2018). It has also been found in Neolithic Jordan (Lazaridis et al., 2016) and Turkey (Mathieson et al., 2015). In the Chalcolithic/Bronze Age, individuals belonging to subclades of U8b have been found in Armenia, the Republic of Artsakh (Margaryan et al., 2017), as well as in the Czech Republic (Lipson et al., 2017) and Hungary (Olalde et al., 2018). Furthermore, it has been reported in an individual from the Ptolemaic period in Egypt (Schuenemann et al., 2017).

Appendix J

Chapter 5 – Supplementary Information

TABLE J.1: Rationale behind every disregarded mutation of the foetal haplotype. Indicators for rationale for discounting differences from rCRS: 1 – low coverage, but rCRS base covered at least once; 2 – most frequent base same as rCRS; 3 – deamination; 4 – private mutation?; 5 – Transversion < Transition.

Position	633	911	5239	6272	6359	6842?	7022	7025?	8576	10674
rCRS	A	T	T	A	A	T	T	A	T	T
total # reads	2	2	8	6	2	3	5	6	6	7
base calls	1A/ 1C	1G/ 1T	2G/1A/ 4T/1C	3G/ 3A	1A/ 1C	1T/ 2C	2G/2T/ 1C	4G/ 2A	2G/ 4T	3G/3T/ 1C
Final base called	A/C	G/T	T	R	A/C	C	G/T	G	T	G/T
	A	T	T	R	A	C	T	G	T	T
Indicator	5	5			5		5			5
	1	1	2	1,2	1	2,3	1	4	2	1,2

Position	10750	11629	12308	12766	12822	12864	12873	13207	14864	15601
rCRS	A	A	A	T	A	T	T	T	T	T
total # reads	2	2	2	5	8	3	3	10	7	2
base calls	1G/1A	1G/1A	1G/1A	3T/2C	2G/5A/1C	2G/1T	1G/1A/1T	2G/2A/6T	2G/1A/4T	1G/1T
Final base called	R	R	R	T	A	G	G/A/T	T	T	G/T
	R	R	R	T	A	T	T	T	T	T
Indicator						5	5			5
	1	1	1	2	2	1	1	2	2	1

Appendix K

Milla Skerra - Supplementary Information

K.1 Introduction

The Glasgow University Archaeological Research Division (GUARD) excavated an eroding coastal site between 2004 and 2007 at Sandwick on Unst in the Shetland Isles (Figure K.1).
During this excavation the remains of a partly truncated building of later prehistoric date were revealed. While excavating the building, an inhumation burial was also discovered, which had been cut through windblown sand that sealed the building after its abandonment. The burial was radiocarbon dated to AD 130-390 (SUERC-10745), which falls within the Scottish Iron Age. The individual was identified as a 50-60 years old male at time of death. Although there was slight surface erosion of the bone, the preservation of the skeleton was good. The individual was buried with small grave goods such as a polished disc of cordierite talc schist, as well as yellow beads on a copper-alloy spiral ring.
The burial being close to an abandoned settlement appears to be a widespread practice during the later Iron Age of placing articulated bodies into or around ruins of buildings across the North and West of Scotland. This practice points to a shared concept that indicates connections between the dead and ruined buildings, joining the people to their lineages. Occasional burials along the foreshore of Sandwick are also known. On the same beach, other remains were found in short cists. Another burial was found slightly further north in a kerbed cairn dating to 370-520 cal AD (GU-1291).
Viking longships from Norway arrived on Unst in the 8th century AD. Including Scandinavia, Unst has the highest density of early rural longhouses known in the Viking world. As the obliteration of indigenous toponymy in the Northern Isles indicates that the Viking settlement involved the gradual

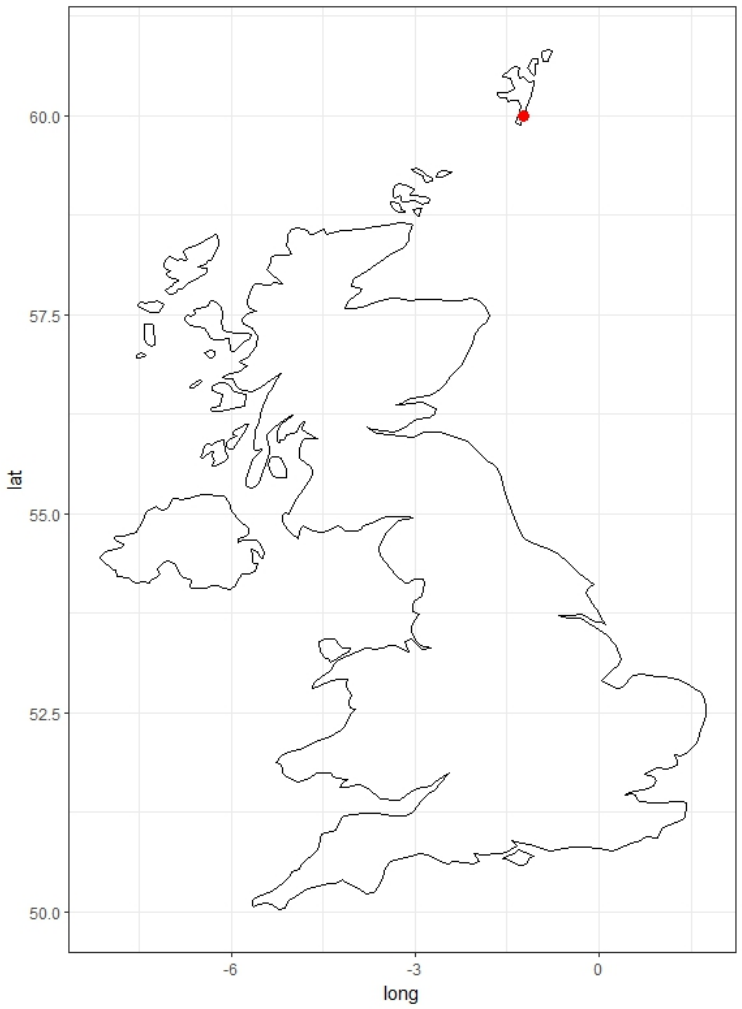

FIGURE K.1: Location of Milla Skerra, Sandwick, Shetland, within the British Isles.

destruction of the indigenous population, rather than intermarrige or cultural assimilation (Smith 2003), the genetic analysis of the male skeleton from Milla Skerra might shed light on the population dynamics in Shetland during the first millenium AD.

K.2 Materials and Methods

The petrous bone was sampled for DNA extraction. The sampling procedure and NGS library preparation followed as described in Chapter 2. The sample was single-indexed for 100bp paired-end sequencing on a HiSeq4000 (Macrogen, South Korea). Bioinformatic processing followed all steps as described in Chapter 2.

K.3 Results

The endogenous content of the sequenced sample was 6.76% and the overall genome coverage was 0.0131x and for the mitochondrial genome 3.1477x. Sequencing results are displayed in Table K.1.

TABLE K.1: Sequencing results of the Milla Skerra individual.

ID	KD073
Endogenous content in %	6.76
Duplication rate in %	0.1384
Genome coverage	0.0131
Std. genome coverage	0.1339
Mitogenome coverage	3.1477
Std. mitogenome coverage	3.6466

K.3.1 Dietary stable isotope analysis

In order to identify the type of diet the Milla Skerra individual consumed over its lifetime, dietary stable isotopes were analysed (Figure K.2).

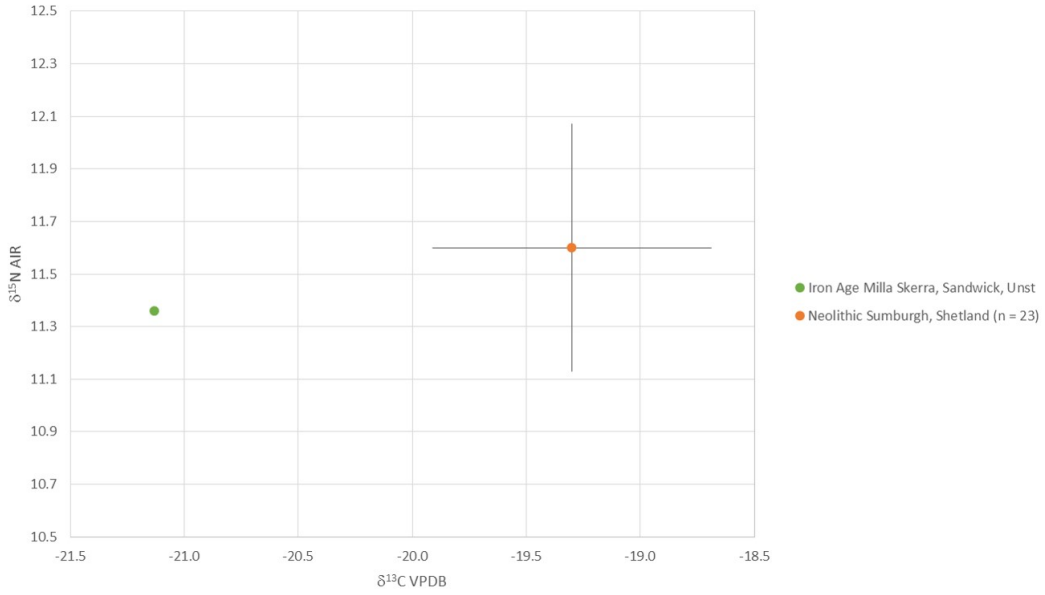

FIGURE K.2: Dietary stable isotope data of the Iron Age individual from Milla Skerra.

K.3.2 Genetic sex determination

The genetic sex could not be assigned using the Skoglund script, therefore a karyotype plot was created to see if an indication for the genetic sex could be

made based on the chromosome length and the number or reads per chromosome. Additionally, the percentage of reads of chromosome X was calculated as 70.3%.

K.3.3 Uniparental marker haplogroup assignment

The individual was found to have all the defining mutations for haplogroup J1b1a1.

TABLE K.2: Mitochondrial DNA haplogroup of the Milla Skerra individual and mutations identified.

Sample	Haplogroup	Average coverage	Polymorphisms
KD073	J1b1a1	3.1477x	73G 146C 152C 263G 279C 709A 750G 1438G 1888A 2706G 4216C 4769G 4917G 5187T 6261A 7028T 7873T 8697A 8860G 10463C 10822T 11251G 11719A 11812G 11914A 13368A 14233G 14766T 14905A 15326G 15452A 15607G 15928A 16126C 16292T

The Y-chromosome haplogroup prediction resulted in only two predictive markers, namely M9370 and Z40372, were found.

K.4 Discussion

The shot-gun sequencing of the Milla Skerra sample resulted in a very low coverage genome of 0.0131x. After pileup calling of variants against the Lazaridis reference dataset, only ∼8000 SNPs were called. This is too low for any genome-wide analysis, therefore, no principal component or ADMIXTURE analysis could be performed. Furthermore, the number of reads was too low to securely determine the genetic sex, and none of the sexing methods gave any conclusive result.

As the osteological sex determination indicated a male individual, Yleaf was run to see if predictive SNPs could be found. Only two predictive markers were found in this sample, M9370 and Z40372. In order to analyse the Y-chromosome for the paternal haplogroup and the whole genome for genetic admixture, further sequencing of this sample is required. A better coverage of the genome would allow to analyse this sample alongside other individuals from the Viking era and compare it to Anglo-Saxon and Romano-British

genomes.

The Milla Skerra individual was assigned to belong to mitochondrial haplogroup J1b1a1. The clade of J1b1a1 encompasses the clade of J1b1a1a, which was determined to be the mitochondrial haplogroup of the previously discussed Rosemarkie individual (see Chapter 6. Haplogroup J1b1a is the only European clade within J1b (Pala et al., 2012).

The Iron Age Milla Skerra individual had a different diet from the Neolithic individuals from Sumburgh in Shetland, which is interesting as both sites are close to the coast. The trophic level of both the Milla Skerra individual and the individuals from Sumburgh were very similar, suggesting a similar amount of protein intake into their diets. However, the $\delta^{13}C$ values are very different, most likely because the Neolithic individuals from Sumburgh were eating a diet that relied more heavily upon fish.

Appendix L

Frequency tables of modern data

TABLE L.1: Haplogroup frequencies for each country of the British Isles.

Haplogroup	Countries of the British Isles			
	England	Ireland	Scotland	Wales
A4	0.13%	0	0.05%	0
B4′5	0.04%	0	0	0
C5	0	0.27%	0	0
H	1.82%	1.07%	1.14%	0.95%
H1	17.24%	13.64%	13.24%	12.96%
H10	0.43%	0.27%	0.31%	0.82%
H11	1.26%	1.87%	1.30%	1.50%
H13	0.87%	0.27%	1.19%	0.41%
H14	0	0.27%	0.36%	0
H15	0	0	0.16%	0.14%
H16	0.39%	0	0.57%	0
H17	0.22%	0.27%	0.31%	1.09%
H18	0.30%	0	0.16%	0.27%
H2	2.39%	3.48%	2.80%	3.96%
H23	0.17%	0.27%	0.52%	0.14%
H24	0.43%	0.80%	0.16%	0.27%
H26	0.17%	0	0.05%	0.14%
H27	0.39%	0	0.36%	0.27%
H28	0.04%	0	0	0.14%
H29	0	0	0.05%	0
H3	4.56%	7.49%	4.41%	5.59%
H30	0.09%	0	0	0
H31	0.39%	0	0.52%	0.14%
H34	0.04%	0	0.05%	0

Table L.1 continued from previous page

Haplogroup	Countries of the British Isles			
	England	Ireland	Scotland	Wales
H35	0	0	0.05%	0
H36	0.04%	0	0.05%	0.14%
H39	0.17%	0	0.21%	0.14%
H4	2.00%	2.41%	1.82%	2.05%
H41	0.09%	0	0	0
H42	0.04%	0	0	0
H43	0.04%	0	0.21%	0
H45	0.22%	0.27%	0.26%	0
H46	0.04%	0	0	0
H47	0.04%	0	0.16%	0
H48	0.09%	0	0	0.55%
H49	0.09%	0	0.10%	0
H5	3.82%	2.41%	2.80%	3.82%
H51	0.04%	0	0.42%	0
H52	0.17%	0	0.36%	0
H5'36	0.04%	0	0	0.14%
H55	0.09%	0	0	0
H56	0.30%	0.53%	0.52%	0
H58	0.04%	0	0.10%	0
H59	0	0.27%	0	0
H6	2.56%	1.87%	2.18%	4.91%
H60	0.04%	0	0	0
H61	0.04%	0	0	0
H63	0.04%	0	0.05%	0.14%
H66	0.22%	0.27%	0.26%	0.14%
H67	0.04%	0	0.05%	0
H69	0	0.80%	0.05%	0
H7	1.35%	1.87%	1.92%	1.36%
H73	0	0	0.05%	0
H76	0	0	0	0.14%
H78	0	0	0	0.14%
H8	0.17%	0	0.83%	0.95%
H80	0	0	0	0.14%
H82	0	0	0.10%	0

Table L.1 continued from previous page

Haplogroup	Countries of the British Isles			
	England	Ireland	Scotland	Wales
H84	0.04%	0	0	0
H86	0.09%	0	0.10%	0
H9	0.04%	0	0	0.27%
H91	0	0	0.05%	0.14%
HV	0.13%	0.27%	0.31%	0
HV0	0.65%	2.41%	0.52%	1.36%
HV1	0.17%	0.27%	0	0.27%
HV15	0.09%	0	0.05%	0.14%
HV16	0.04%	0	0.26%	0
HV17	0.09%	0	0	0
HV4	0.04%	0	0	0
HV6	0.30%	0.53%	0	0
HV9	0.13%	0	0.16%	0.14%
I1	0.74%	1.07%	1.77%	2.86%
I2	1.26%	0.53%	2.13%	1.50%
I3	0.26%	0	0.05%	0
I4	0.74%	1.34%	0.67%	0.82%
J1	8.55%	10.70%	12.98%	12.14%
J2	1.48%	0.53%	2.44%	0.95%
K1	7.21%	8.29%	7.48%	8.59%
K2	1.61%	0.80%	1.40%	3.14%
L1	0.26%	0	0.16%	0
L2	0.22%	0	0.21%	0
L3	0.09%	0	0.05%	0
M1	0.09%	0	0	0
M18	0.04%	0	0	0
N1	0.26%	0.27%	0.26%	0
N3	0	0	0.10%	0
R0	0.13%	0	0	0
R1	0.26%	0	0	0
R2	0.04%	0	0	0
T1	1.82%	1.34%	1.92%	2.32%
T2	7.69%	9.09%	6.65%	5.59%
U1	0.13%	0.27%	0.10%	0

Table L.1 continued from previous page

Haplogroup	Countries of the British Isles			
	England	Ireland	Scotland	Wales
U2	0.61%	2.14%	1.14%	1.09%
U3	1.26%	0.27%	0.67%	1.23%
U4	2.39%	1.87%	2.28%	0.68%
U5	11.12%	9.36%	8.10%	6.82%
U6	0.09%	0.27%	0.05%	0
U7	0.17%	0	0.10%	0
U8	0.30%	0.53%	0.16%	0.14%
V	1.39%	1.07%	1.14%	3.14%
V1	0.22%	0	0	0.14%
V10	0.17%	1.34%	0.47%	0.14%
V11	0.04%	0	0.05%	0
V15	0.26%	0.53%	0.52%	0
V17	0.04%	0	0	0
V18	0.04%	0	0.05%	0
V19	0	0.53%	0.05%	0
V2	0.22%	0	0.42%	0.14%
V3	0.17%	0.27%	0.21%	0
V4	0.04%	0	0	0
V7	0.09%	0	0.21%	0
V8	0	0	0.42%	0
V9	0.22%	0	0.16%	0.14%
W	0.04%	0	0	0
W1	0.35%	1.34%	0.67%	0.27%
W3	0.30%	0	0.05%	0.27%
W4	0.22%	0	0.67%	0.27%
W5	0.35%	0.80%	0.62%	0.27%
W6	0.13%	0	0	0
X2	1.82%	1.34%	1.66%	1.50%

TABLE L.2: Haplogroup frequencies of the different regions of England, the Channel Isles and the Isle of Man.

Haplogroup	Regions							
	England C	Channel Is	England E	England N	England SE	England SW	England York	Isle of Man
A4	0	0	0.47%	0	0	0.52%	0.34%	0
B4'5	0	0	0	0	0	0	0.34%	0
H	1.44%	0	3.26%	1.69%	2.11%	1.04%	1.71%	0
H1	17.75%	14.29%	16.74%	16.74%	17.07%	16.58%	18.49%	42.86%
H10	0.72%	0	0.47%	0.21%	0.16%	1.04%	0.34%	0
H11	1.20%	0	0.93%	1.48%	1.30%	0.52%	1.71%	0
H13	0.96%	0	0.93%	1.69%	0.65%	0	0.68%	0
H14	0	0	0	0	0	0	0	0
H15	0	0	0	0	0	0	0	0
H16	0.24%	0	0	0	0.16%	0.52%	1.71%	0
H17	0	0	0.47%	0	0.33%	1.04%	0	0
H18	0.48%	0	0.93%	0.42%	0	0	0.34%	0
H2	3.12%	14.29%	3.26%	2.54%	1.63%	2.07%	2.05%	0
H23	0.24%	0	0	0.21%	0.33%	0	0	0
H24	0.72%	0	0.47%	0	0.98%	0	0	0
H26	0	0	0.47%	0	0.49%	0	0	0
H27	0.48%	0	0	0.64%	0.49%	0	0.34%	0
H28	0	0	0	0.21%	0	0	0	0

Table L.2 continued from previous page

Haplogroup	Regions							
	England C	Channel Is	England E	England N	England SE	England SW	England York	Isle of Man
H29	0	0	0	0	0	0	0	0
H3	3.84%	14.29%	2.79%	4.87%	4.88%	5.70%	3.77%	0
H30	0	0	0.47%	0	0.16%	0	0	0
H31	0	0	0	0.85%	0.49%	0	0.34%	0
H34	0.24%	0	0	0	0	0	0	0
H35	0	0	0	0	0	0	0	0
H36	0	0	0	0	0.16%	0	0	0
H39	0	0	0	0	0.65%	0	0	0
H4	1.44%	0	2.33%	2.75%	1.95%	2.07%	1.37%	14.29%
H41	0	0	0	0.42%	0	0	0	0
H42	0.24%	0	0	0	0	0	0	0
H43	0	0	0	0	0	0	0.34%	0
H45	0.72%	0	0.47%	0	0.16%	0	0	0
H46	0	0	0	0	0.16%	0	0	0
H47	0	0	0	0	0	0	0.34%	0
H48	0	0	0	0.21%	0	0	0	0
H49	0	0	0	0.42%	0	0	0	0
H5	4.80%	0	4.65%	2.54%	2.76%	6.22%	4.45%	0
H51	0	0	0	0	0.16%	0	0	0

Table L.2 continued from previous page

Haplogroup	Regions							
	England C	Channel Is	England E	England N	England SE	England SW	England York	Isle of Man
H52	0	0	0.93%	0.21%	0	0	0.34%	0
H5′36	0	0	0.47%	0	0	0	0	0
H55	0.24%	0	0.47%	0	0	0	0	0
H56	0.48%	0	0.47%	0.42%	0.16%	0.52%	0	0
H58	0	0	0	0.21%	0	0	0	0
H6	1.92%	0	2.33%	3.18%	1.79%	4.66%	2.74%	14.29%
H60	0	0	0.47%	0	0	0	0	0
H61	0	0	0	0	0.16%	0	0	0
H63	0	0	0	0.21%	0	0	0	0
H66	0.48%	0	0	0	0.33%	0.52%	0	0
H67	0	0	0	0.21%	0	0	0	0
H69	0	0	0	0	0	0	0	0
H7	1.68%	0	2.79%	0.64%	1.46%	1.04%	0.68%	0
H73	0	0	0	0	0	0	0	0
H76	0	0	0	0	0	0	0	0
H78	0	0	0	0	0	0	0	0
H8	0.48%	0	0	0.42%	0	0	0	0
H80	0	0	0	0	0	0	0	0
H82	0	0	0	0	0	0	0	0

Table L.2 continued from previous page

Haplogroup	Regions							
	England C	Channel Is	England E	England N	England SE	England SW	England York	Isle of Man
H84	0	0	0	0.21%	0	0	0	0
H86	0	0	0	0	0.33%	0	0	0
H9	0.24%	0	0	0	0	0	0	0
H91	0	0	0	0	0	0	0	0
HV	0.24%	0	0	0.21%	0.16%	0	0	0
HV0	0.72%	0	0.93%	0.64%	0.33%	0.52%	1.03%	0
HV1	0.24%	0	0.47%	0.42%	0	0	0	0
HV15	0.24%	0	0	0.21%	0	0	0	0
HV16	0	0	0	0.21%	0	0	0	0
HV17	0.24%	0	0	0.21%	0	0	0	0
HV4	0	0	0	0	0.16%	0	0	0
HV6	0.48%	0	0.47%	0.21%	0.16%	1.04%	0	0
HV9	0	0	0.47%	0	0.16%	0	0.34%	0
I1	0.48%	0	1.40%	1.06%	0.33%	0	1.03%	0
I2	1.68%	0	1.40%	1.69%	0.98%	1.04%	1.03%	0
I3	0	0	0.47%	0.21%	0.49%	0	0.34%	0
I4	0.96%	0	0	0.85%	0.65%	1.55%	0.68%	0
J1	8.15%	0	9.77%	7.20%	9.92%	8.29%	7.19%	0
J2	1.20%	0	2.33%	1.48%	1.63%	1.04%	0.68%	0

Table L.2 continued from previous page

Haplogroup	Regions							
	England C	Channel Is	England E	England N	England SE	England SW	England York	Isle of Man
K1	7.67%	0	4.65%	5.93%	7.64%	9.33%	8.90%	0
K2	2.64%	0	0.93%	1.06%	1.30%	2.59%	1.37%	0
L1	0.24%	0	0	0.42%	0.49%	0	0	0
L2	0	0	0.47%	0	0.49%	0	0.34%	0
L3	0	0	0	0	0.16%	0	0.34%	0
M1	0	0	0	0	0.33%	0	0	0
M18	0.24%	0	0	0	0	0	0	0
N1	0	0	0	0.42%	0.16%	0	0.68%	0
N3	0	0	0	0	0	0	0	0
R0	0	0	0.47%	0	0.33%	0	0	0
R1	0.24%	0	0.47%	0.42%	0	0.52%	0.34%	0
R2	0	0	0	0	0	0	0.34%	0
T1	1.44%	0	0	2.54%	2.44%	1.04%	1.71%	0
T2	9.83%	14.29%	6.98%	6.78%	7.97%	8.29%	6.51%	0
U1	0.24%	0	0	0.21%	0.16%	0	0	0
U2	0.48%	0	0.93%	0.42%	0.65%	0.52%	0.68%	0
U3	1.44%	0	0.47%	2.12%	1.14%	1.55%	0.34%	0
U4	2.16%	0	2.33%	2.75%	2.93%	2.59%	1.37%	0
U5	8.39%	14.29%	10.70%	11.23%	10.08%	10.36%	17.12%	28.57%

Table L.2 continued from previous page

Haplogroup	Regions							
	England C	Channel Is	England E	England N	England SE	England SW	England York	Isle of Man
U6	0	0	0	0	0.16%	0.52%	0	0
U7	0	0	0	0.21%	0.16%	0	0.68%	0
U8	0.72%	0	0	0	0.33%	0	0.68%	0
V	0.48%	0	0.93%	1.06%	2.76%	1.04%	1.03%	0
V1	0.72%	0	0	0.21%	0	0.52%	0	0
V10	0.24%	0	0.47%	0.42%	0	0	0	0
V11	0	0	0	0	0	0	0	0
V15	0.24%	0	0	0.21%	0.16%	0.52%	0.68%	0
V17	0	0	0	0.21%	0	0	0	0
V18	0.24%	0	0	0	0	0	0	0
V19	0	0	0	0	0	0	0	0
V2	0	0	0	0.42%	0.33%	0	0.34%	0
V3	0	0	0	0	0.33%	0.52%	0	0
V4	0	0	0	0.21%	0	0	0	0
V7	0	0	0	0.21%	0	0.52%	0	0
V8	0	0	0	0	0	0	0	0
V9	0	0	0	0.42%	0.16%	1.04%	0	0
W	0	0	0.47%	0	0	0	0	0
W1	0.48%	0	1.40%	0.42%	0	0	0.34%	0

Table L.2 continued from previous page

Haplogroup	Regions							
	England C	Channel Is	England E	England N	England SE	England SW	England York	Isle of Man
W3	0.24%	0	0	0.21%	0.81%	0	0	0
W4	0	0	0	0.64%	0.33%	0	0	0
W5	0.48%	0	0.47%	0.42%	0.16%	0	0.68%	0
W6	0	0	0	0.42%	0.16%	0	0	0
X2	2.40%	28.57%	3.26%	1.69%	1.46%	1.04%	0.68%	0

TABLE L.3: Haplogroup frequencies of the different regions of Ireland and Scotland divided by regions.

Haplogroup	Regions								
	Ire. Connacht	Ire. Leinster	Ire. Munster	Ire. Ulster	Scot. C	Scot. N&W	Scot. NE	Scot. SE	Scot. SW
A4	0	0	0	0	0	0	0	0	0
B4′5	0	0	0	0	0	0	0	0	0
H	0	2.33%	0	0	2.02%	0	0.37%	1.44%	1.08%
H1	13.64%	10.47%	10.67%	19.53%	14.16%	15.25%	13.33%	11.87%	12.67%
H10	2.27%	0	0	0	0	1.69%	0.37%	0.72%	0
H11	0	1.16%	2.67%	1.56%	1.73%	0.56%	1.11%	1.44%	1.48%
H13	0	0	0	0.78%	0.29%	0.56%	1.11%	2.16%	1.48%
H14	0	0	0	0.78%	1.45%	0	0	0.36%	0.13%
H15	0	0	0	0	0.29%	0	0.37%	0	0.13%
H16	0	0	0	0	0.29%	2.26%	1.48%	0	0.13%
H17	0	1.16%	0	0	0.29%	0	0.74%	0	0.40%
H18	0	0	0	0	0	0.56%	0.37%	0	0.13%
H2	2.27%	4.65%	2.67%	2.34%	2.60%	2.26%	3.70%	3.60%	2.70%
H23	2.27%	0	0	0	0	0.56%	0	0.72%	0.94%
H24	2.27%	1.16%	1.33%	0	0	0	0	0	0.40%
H26	0	0	0	0	0	0	0	0	0.13%
H27	0	0	0	0	0.87%	0	0.74%	0	0.27%
H28	0	0	0	0	0	0	0	0	0

Table L.3 continued from previous page

Haplogroup	Regions								
	Ire. Connacht	Ire. Leinster	Ire. Munster	Ire. Ulster	Scot. C	Scot. N&W	Scot. NE	Scot. SE	Scot. SW
H29	0	0	0	0	0.29%	0	0	0	0
H3	6.82%	10.47%	6.67%	5.47%	4.91%	1.69%	5.56%	2.52%	5.12%
H30	0	0	0	0	0	0	0	0	0
H31	0	0	0	0	0.29%	0.56%	0.74%	0.72%	0.40%
H34	0	0	0	0	0	0	0	0	0.13%
H35	0	0	0	0	0	0	0	0	0.13%
H36	0	0	0	0	0	0	0	0.36%	0
H39	0	0	0	0	0	0	0.74%	0.36%	0.13%
H4	2.27%	4.65%	1.33%	1.56%	0.29%	4.52%	2.59%	1.08%	1.89%
H41	0	0	0	0	0	0	0	0	0
H42	0	0	0	0	0	0	0	0	0
H43	0	0	0	0	0	0	0.74%	0.36%	0.13%
H45	0	1.16%	0	0	0.58%	0	0.37%	0.72%	0
H46	0	0	0	0	0	0	0	0	0
H47	0	0	0	0	0.58%	0.56%	0	0	0
H48	0	0	0	0	0	0	0	0	0
H49	0	0	0	0	0.29%	0	0.37%	0	0
H5	0	1.16%	2.67%	3.13%	2.31%	2.82%	3.33%	3.24%	2.96%
H51	0	0	0	0	0.29%	0	0.37%	0.72%	0.54%

Table L.3 continued from previous page

Haplogroup	Regions								
	Ire. Connacht	Ire. Leinster	Ire. Munster	Ire. Ulster	Scot. C	Scot. N&W	Scot. NE	Scot. SE	Scot. SW
H52	0	0	0	0	0	0	1.11%	0.36%	0.27%
H5'36	0	0	0	0	0	0	0	0	0
H55	0	0	0	0	0	0	0	0	0
H56	2.27%	1.16%	0	0	0.29%	0.56%	0.74%	0.72%	0.54%
H58	0	0	0	0	0	0.56%	0	0	0.13%
H6	2.27%	2.33%	1.33%	0	3.18%	2.26%	2.22%	2.16%	1.75%
H60	0	0	0	0	0	0	0	0	0
H61	0	0	0	0	0	0	0	0	0
H63	0	0	0	0	0	0	0.37%	0	0
H66	0	0	0	0	0.29%	0	0	0	0.54%
H67	0	0	0	0	0	0.56%	0	0	0
H69	2.27%	1.16%	0	0.78%	0	0	0	0	0
H7	0	2.33%	4.00%	0.78%	2.02%	0.56%	1.85%	2.16%	2.29%
H73	0	0	0	0	0.29%	0	0	0	0
H76	0	0	0	0	0	0	0	0	0
H78	0	0	0	0	0	0	0	0	0
H8	0	0	0	0	2.31%	0	0	0.36%	0.67%
H80	0	0	0	0	0	0	0	0	0
H82	0	0	0	0	0	0	0	0	0.27%

Table L.3 continued from previous page

Haplogroup	Regions								
	Ire. Connacht	Ire. Leinster	Ire. Munster	Ire. Ulster	Scot. C	Scot. N&W	Scot. NE	Scot. SE	Scot. SW
H84	0	0	0	0	0	0	0	0	0
H86	0	0	0	0	0	0	0	0.36%	0
H9	0	0	0	0	0	0	0	0	0
H91	0	0	0	0	0	0	0	0	0.13%
HV	2.27%	0	0	0	0.29%	0	0.74%	0.72%	0.13%
HV0	2.27%	2.33%	2.67%	1.56%	0.29%	0.56%	0.37%	1.08%	0.54%
HV1	0	0	0	0	0	0	0	0	0
HV15	0	0	0	0	0	0	0.37%	0	0
HV16	0	0	0	0	0.58%	0	0.37%	0	0.13%
HV17	0	0	0	0	0	0	0	0	0
HV4	0	0	0	0	0	0	0	0	0
HV6	0	0	0	1.56%	0	0	0	0	0
HV9	0	0	0	0	0.29%	0.56%	0.37%	0	0
I1	0	1.16%	0	2.34%	2.31%	1.13%	1.11%	2.16%	1.48%
I2	0	0	1.33%	0.78%	2.31%	2.82%	2.22%	2.88%	1.48%
I3	0	0	0	0	0	0	0	0	0.13%
I4	0	1.16%	1.33%	1.56%	0.87%	0	1.11%	0	0.40%
J1	4.55%	10.47%	9.33%	15.63%	11.85%	12.99%	12.59%	15.47%	11.46%
J2	0	0	0	1.56%	1.45%	1.13%	4.07%	2.88%	2.29%

Table L.3 continued from previous page

Haplogroup	Regions								
	Ire. Connacht	Ire. Leinster	Ire. Munster	Ire. Ulster	Scot. C	Scot. N&W	Scot. NE	Scot. SE	Scot. SW
K1	6.82%	6.98%	14.67%	7.03%	8.38%	6.78%	7.04%	4.68%	8.76%
K2	0	2.33%	0	0.78%	0.29%	1.13%	1.11%	0.36%	2.43%
L1	0	0	0	0	0	0.56%	0.74%	0	0
L2	0	0	0	0	0	0.56%	0	0.72%	0.13%
L3	0	0	0	0	0.29%	0	0	0	0
M1	0	0	0	0	0	0	0	0	0
M18	0	0	0	0	0	0	0	0	0
N1	0	0	1.33%	0	0	0	0.37%	0.72%	0.27%
N3	0	0	0	0	0.29%	0	0	0	0.13%
R0	0	0	0	0	0	0	0	0	0
R1	0	0	0	0	0	0	0	0	0
R2	0	0	0	0	0	0	0	0	0
T1	6.82%	0	1.33%	0.78%	2.02%	0.56%	2.22%	2.52%	2.16%
T2	11.36%	9.30%	12.00%	7.81%	7.51%	9.04%	2.96%	6.47%	7.14%
U1	0	0	0	0.78%	0.29%	0.56%	0	0	0
U2	2.27%	2.33%	0	3.91%	0.87%	0	1.48%	2.16%	1.21%
U3	0	1.16%	0	0	0.58%	1.69%	0.74%	0.72%	0.40%
U4	4.55%	2.33%	0	2.34%	2.02%	2.26%	2.59%	2.16%	2.56%
U5	13.64%	11.63%	12.00%	4.69%	9.25%	11.30%	6.67%	6.47%	8.09%

Table L.3 continued from previous page

Haplogroup	Regions								
	Ire. Connacht	Ire. Leinster	Ire. Munster	Ire. Ulster	Scot. C	Scot. N&W	Scot. NE	Scot. SE	Scot. SW
U6	0	1.16%	0	0	0.29%	0	0	0	0
U7	0	0	0	0	0	0	0	0	0.27%
U8	2.27%	0	1.33%	0	0	0	0	0.36%	0.27%
V	0	0	1.33%	2.34%	0.58%	1.69%	0.74%	0.72%	1.62%
V1	0	0	0	0	0	0	0	0	0
V10	0	1.16%	2.67%	0.78%	0.29%	0.56%	0.37%	1.44%	0.27%
V11	0	0	0	0	0	0	0	0	0.13%
V15	0	0	2.67%	0	0	0	1.11%	0.36%	0.54%
V17	0	0	0	0	0	0	0	0	0
V18	0	0	0	0	0	0	0	0	0
V19	0	0	0	1.56%	0	0	0	0	0.13%
V2	0	0	0	0	0.58%	0	0	0.72%	0.54%
V3	0	0	0	0	0	0	0	0.72%	0.27%
V4	0	0	0	0	0	0	0	0	0
V7	0	0	0	0	0.58%	0	0	0	0.27%
V8	0	0	0	0	0.29%	0	0	0.36%	0.67%
V9	0	0	0	0	0	0	0	0	0.40%
W	0	0	0	0	0	0	0	0	0
W1	0	0	2.67%	2.34%	0.29%	1.69%	0.74%	0.36%	0.81%

Table L.3 continued from previous page

Haplogroup	Regions								
	Ire. Connacht	Ire. Leinster	Ire. Munster	Ire. Ulster	Scot. C	Scot. N&W	Scot. NE	Scot. SE	Scot. SW
W3	0	0	0	0	0	0	0	0.36%	0
W4	0	0	0	0	1.16%	0.56%	0.37%	0.72%	0.67%
W5	2.27%	1.16%	0	0.78%	0.87%	0.56%	0	0.72%	0.54%
W6	0	0	0	0	0	0	0	0	0
X2	2.27%	0	0	2.34%	0.87%	2.82%	2.59%	2.52%	1.08%

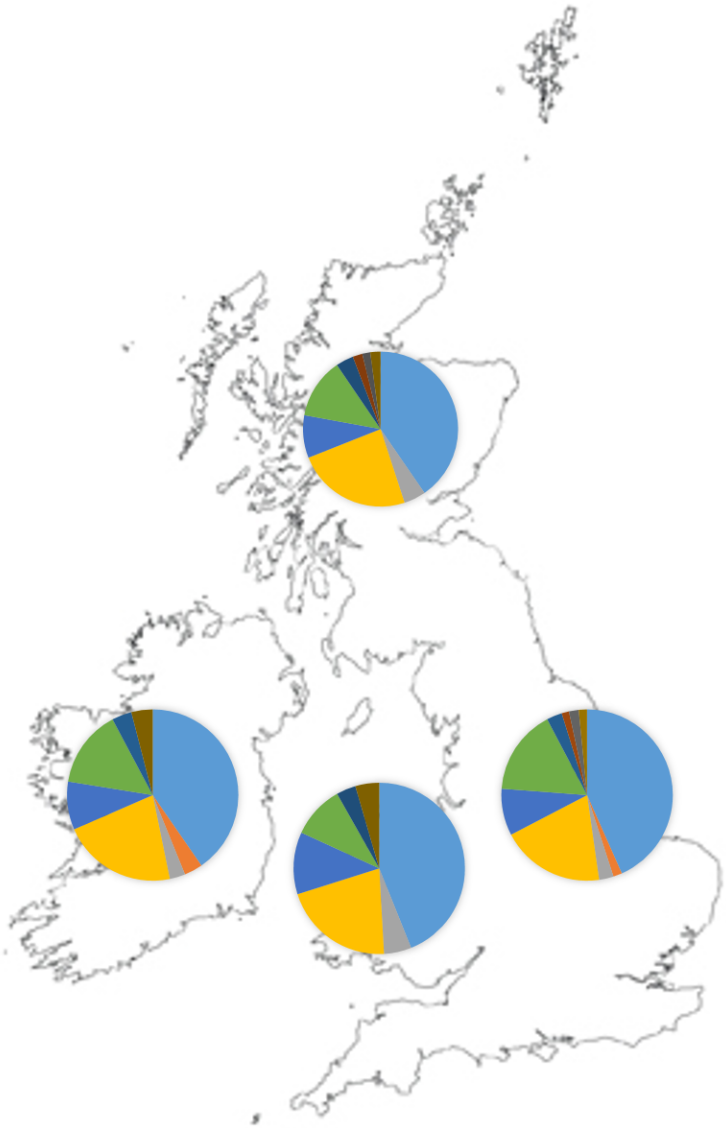

FIGURE L.1: Haplogroup compositions in England, Ireland, Scotland and Wales. Light Blue = hg H, Orange = hg HV, Grey = hg I, Yellow = hg JT, Blue = hg K, Green = hg U, Dark Blue = hg V, Brown = hg W, Dark Grey = hg X, Light Brown = all other hgs present in the British SNP dataset.

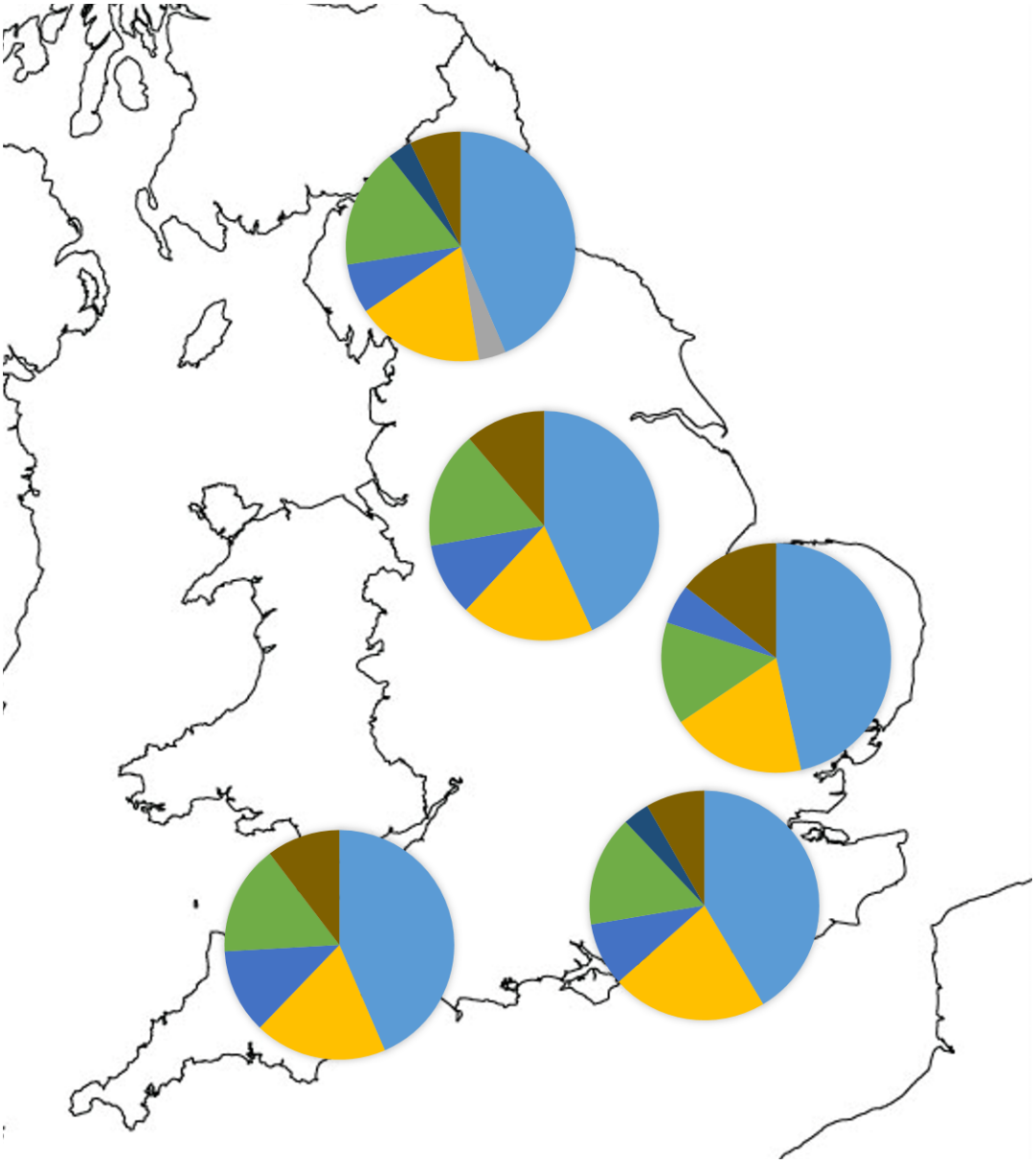

FIGURE L.2: Haplogroup composition of northern, central, eastern, south-western and south-eastern England. Light Blue = hg H, Grey = hg I, Yellow = hg JT, Blue = hg K, Green = hg U, Dark Blue = hg V, Light Brown = all other hgs present in the British SNP dataset.

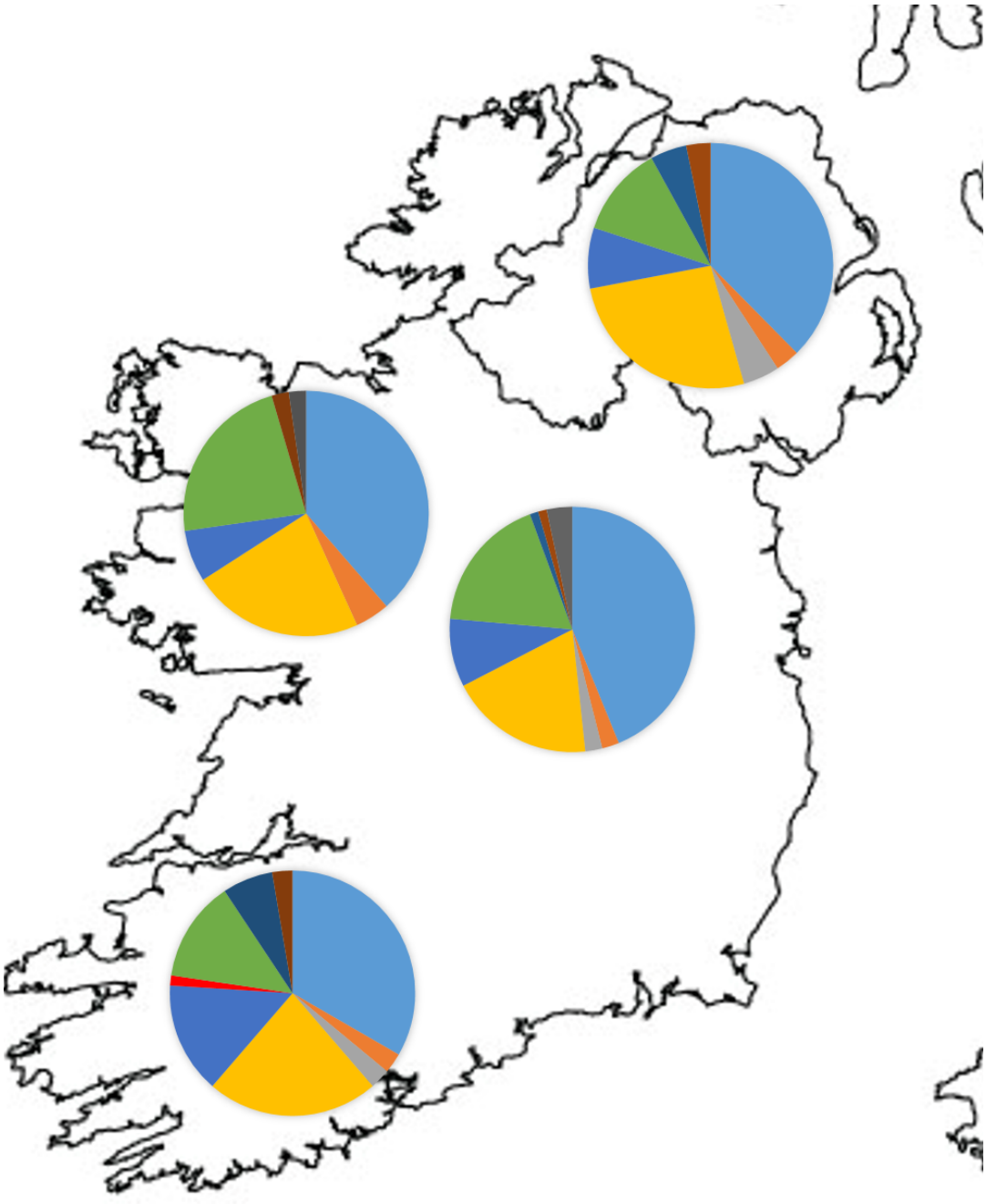

FIGURE L.3: Haplogroup composition of Munster, Ulster, Connacht and Leinster in Ireland. Light Blue = hg H, Orange = hg HV, Grey = hg I, Yellow = hg JT, Blue = hg K, Red = hg N, Green = hg U, Dark Blue = hg V, Dark Grey = hg X, Light Brown = all other hgs present in the British SNP dataset.

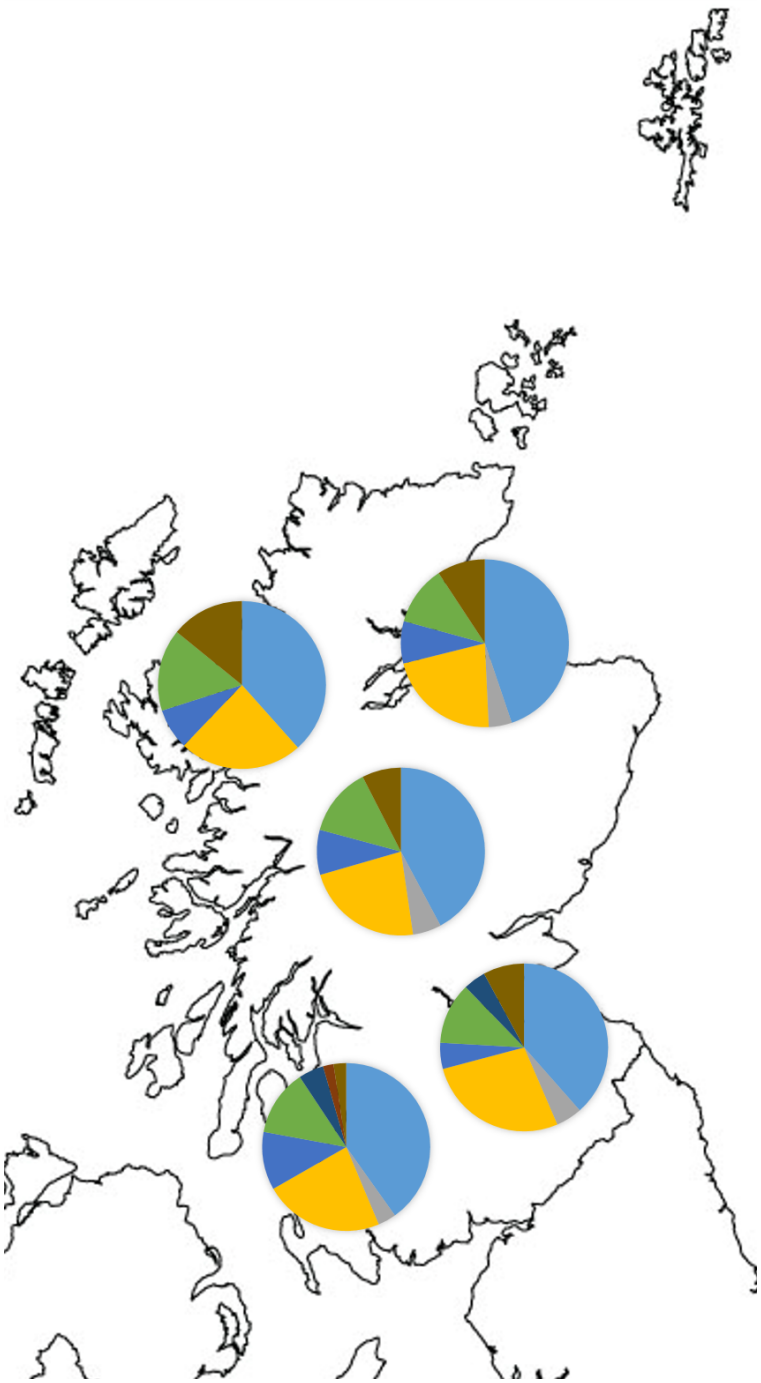

FIGURE L.4: Haplogroup composition of northwestern, northeastern, central, southwestern and southeastern Scotland. Light Blue = hg H, Grey = hg I, Yellow = hg JT, Blue = hg K, Green = hg U, Dark Blue = hg V, Brown = hg W, Light Brown = all other hgs present in the British SNP dataset.

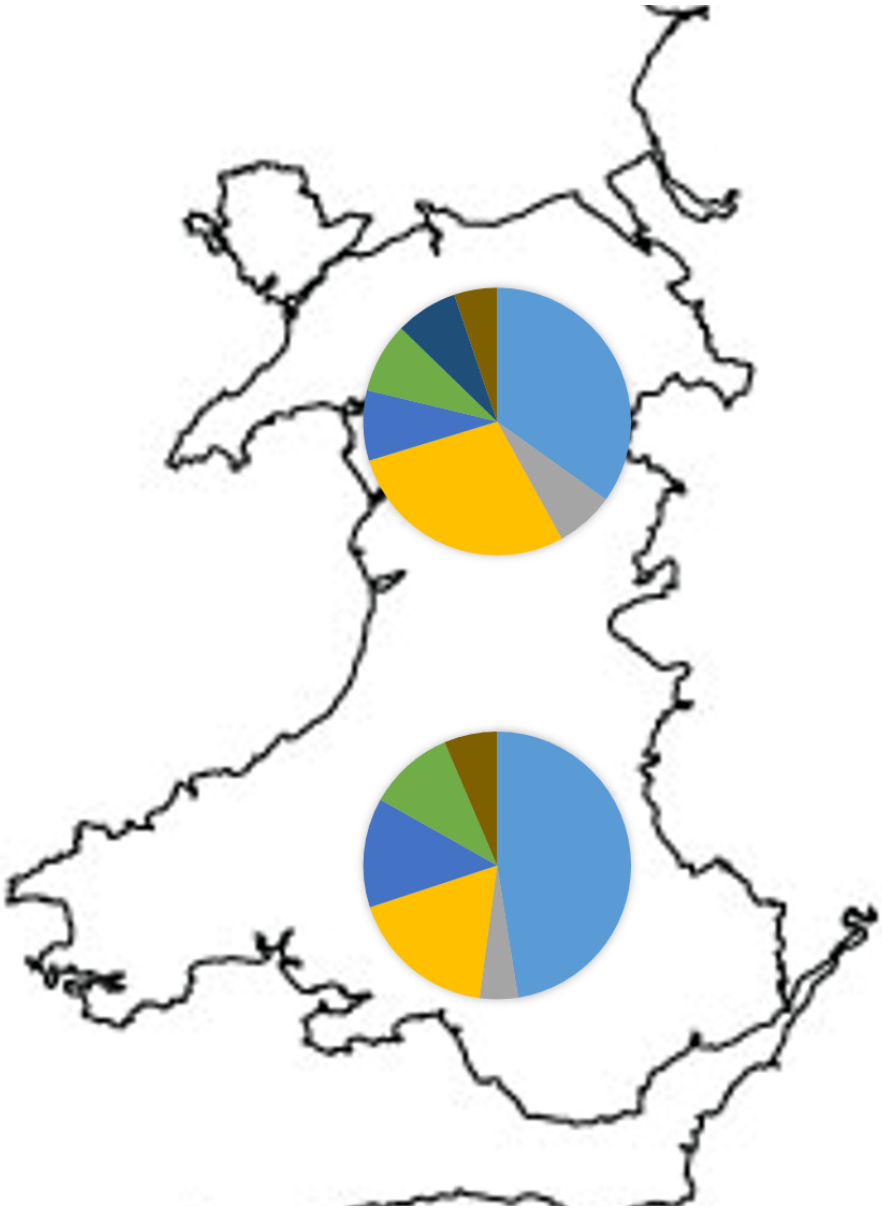

FIGURE L.5: Haplogroup composition of northern and southern Wales. Light Blue = hg H, Grey = hg I, Yellow = hg JT, Blue = hg K, Green = hg U, Dark Blue = hg V, Light Brown = all other hgs present in the British SNP dataset.