REASSESSING DNA PRESERVATION OF ANCIENT MATERIAL FROM POULTON, UK AND TESTING AND OPTIMIZING OF CATS: A NOVEL SINGLE-STRAND DNA LIBRARY PREPARATION TECHNIQUE

ASHLEY LYNN MAY

A thesis submitted in partial fulfillment of the requirements of Liverpool John Moores University for the degree of Master of Philosophy

November 2017

ABSTRACT

Sequencing of ancient DNA with low preservation is still a major challenge, despite the immense progress that the field of ancient genetic studies has made since its inception over thirty years ago. The main aim of this study was to reassess the DNA preservation of ancient medieval material from Poulton, Cheshire, UK, using standard dsDNA library preparation methods. We targeted cementum of teeth as the optimal substrate for DNA extraction, along with long bones for comparison. A second aim, using a newer single-stranded DNA (ssDNA) library method, namely Capture and Amplification by Tailing and Switching (CATS) was also tested for the potential of isolating higher levels of endogenous DNA from Poulton material. Ten libraries built by the double-stranded DNA (dsDNA) method were sequenced successfully, displaying well-preserved DNA with good potential for high-coverage genome studies. Principal Component Analysis (PCA) and haplogroup data suggests that the Poulton samples group within modern European variation, though with some affinity to Southeastern Europe relative to modern British people. Sex determination of the Poulton samples presented four males and three females. The ssDNA CATS method holds potential for further development, but in our study did not perform better than the standard dsDNA library preparation method. We also compared the sequence quality of samples that were washed upon excavation with samples that were not washed. These results highlighted the sensitivity of ancient samples to washing and importance of following up-to-date guidelines for optimal protection of DNA in excavated ancient material.

STATEMENT

I confirm that the research undertaken and written thesis is my own work. Some research techniques were not undertaken by myself and are clearly stated in the methods section. All other work is my own.

TABLE OF CONTENTS

Chapter 1: Introduction and Background 1

1.1 Poulton Excavation Site 1

1.1.1 Poulton through the eras 2

1.2 Genetic History of the British Isles 4

1.3 Evolution of Ancient DNA Techniques 6

1.3.1 Background 6

1.3.2 Damage and Contamination 9

1.4 The move to Next Generation Sequencing (NGS) 12

1.4.1 Background 12

1.4.2 DNA sequences and Data Analysis 14

1.5 Ancient DNA samples and Preservation 15

1.6 DNA Library Preparation Techniques 16

1.6.1 Double strand DNA Library Techniques 16

1.6.2 Single strand DNA Library Techniques 17

1.7 Aims and Objectives 20

Chapter 2: Methods 21

2.1 Sampling and Powdering 21

2.2 Ancient DNA Authentication 21

2.3 DNA Extractions 22

2.4 DNA Library Preparations 22

2.4.1 Meyer Library Preparation 23

2.4.2 Capture and Amplification by Tailing and Switching 24

2.5 Sequencing 28

2.5.1 Sequencing Process and Alignment 29

2.5.2 Mitochondrial DNA Haplogroups 29

2.5.3 Mitochondrial DNA Authentication 29

2.5.4 Biological Sexing 30

2.6 Population Genetics 30

2.6.1 Principal Component Analysis (PCA) 30

Chapter 3: Results 31

3.1 Poulton samples and DNA Extractions 31
3.2 Sequencing and Authenticity 32
3.3 Washed upon excavation versus not washed 34
3.4 Phylogenetic Analysis 34
3.5 Library Preparation Techniques 38
3.5.1 Meyer 38
3.5.2 CATS 39
3.5.3 Comparison of CATS and Meyer 39

Chapter 4: Discussion 41

4.1 Assessment of Poulton DNA Preservation 41
4.2 Analysis of Poulton sequences 42
4.3 Population Genetics 44
4.4 Capture and Amplification by Tailing and Switching 47
4.5 Cementum versus Long Bone 48
4.6 Conclusions 49
4.7 Future Work 49

Reference List 51

Appendix 1: Images of ten Poulton tooth samples that were sequenced 67
Appendix 2: Primers and Indexing Oligos 72
Appendix 3: Mitochondrial DNA 75
Appendix 4: Quantitative PCR Results 95
Appendix 5: Mitochondrial DNA damage plots of Poulton 101

Chapter 1

1.1 Poulton Excavation Site

The Poulton Research Project is a multi-period archaeological excavation site located in Poulton, Cheshire, which is south of Chester, England and west of the River Dee (Figure 1). The project began in 1995 with the hope of discovering a settlement archaeologically rich over several time periods in the Chester hinterland (Poulton Research Project, 2014). The chapel site at Poulton provided more than this. Knowledge of the history and origin of the chapel was scarce, yet it was known that it had a close connection with the last Cistercian Abbey of Poulton. The abbey was removed in the 13th century, but the Cistercians continued to dominate the landscape around Poulton until 1534. The remains of the chapel are located at the southern part at the far end of a pastoral field, which slopes into Old Pulford Brook (Burrell & Carpenter, 2014).

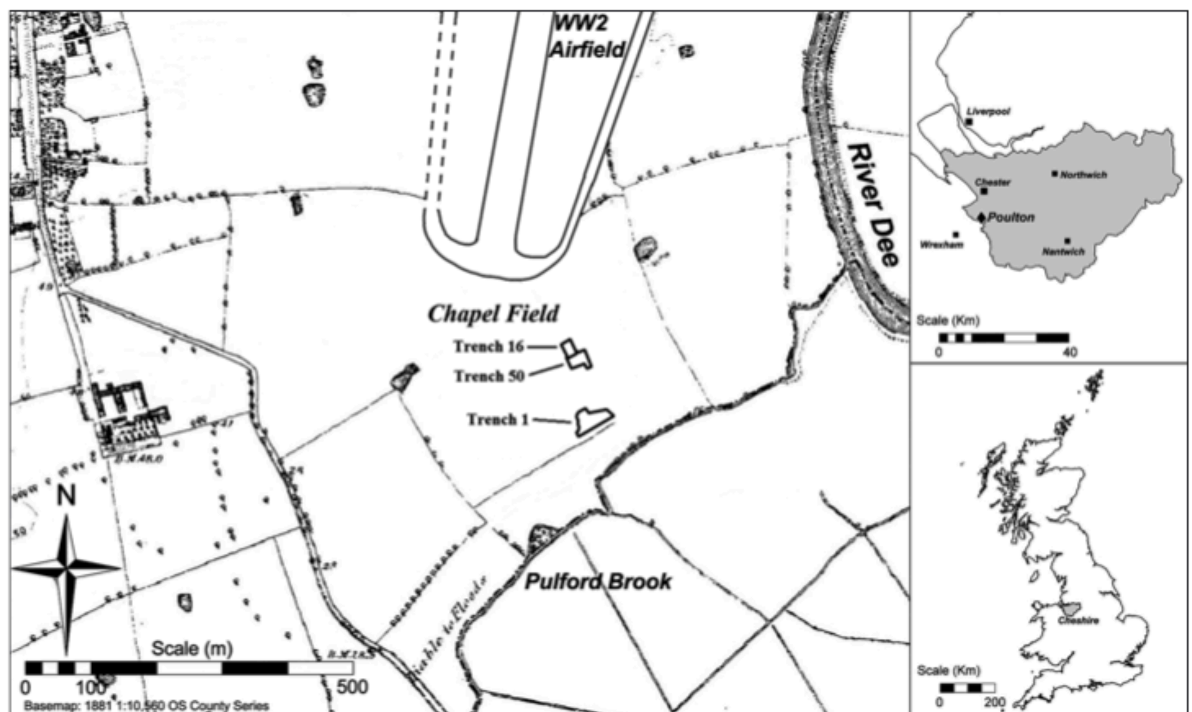


Figure 1: Location map of the Poulton site showing trenches (Scale 1:1000)

In addition to medieval archaeology, evidence for Roman and prehistoric occupation was revealed in 1995. It has since been confirmed that at least two Roman buildings existed. The four excavation seasons from 1995 to 1998 revealed the chapel's upper foundations and a large part of the cemetery. Artifacts excavated include one of the richest collections of medieval findings from Cheshire, including building material, domestic equipment and

personal belongings (Poulton Research Project, 2014). It is thought that the first early settlers to Cheshire arrived in the late upper Paleolithic period, around 11,000 BC. The traditional views of prehistoric Cheshire were ones of a virtually unpopulated region, but fieldwork has since revealed evidence for a number of possible sites in the area by the discovery of worked flint and other artifacts. The idea of prehistoric occupation at Poulton is supported by these discoveries, along with the presence of documented prehistoric activity in the nearby Welsh countryside (Emery, 2000).

The site spans an incredible 9000 years; the finds ranging from carefully worked pre-historic flints to 17th century artifacts from the Civil War. As of 2013, 697 human skeletons have been excavated, along with substantial amounts of disarticulated bone (Burrell & Carpenter, 2014). Over the years, LJMU has become increasingly more involved with the excavation, with students now completing course modules at the site, as well as extensive post-graduate research. Over 400 whole skeletons are stored at the University (LJMU Blog, 2013). Previous research attempted to extract DNA from Poulton material but was unsuccessful (Town, 2015) and it was concluded that the DNA is poorly preserved.

1.1.1 Poulton through the eras

Poulton during the Mesolithic Era, Neolithic Period and Iron Age

The initial survey of Poulton in 1995 revealed the surprising discovery of prehistoric artifacts. Poulton would have been ideal for transient hunter-gatherers as it is located on higher ground looking over part of the River Dee where they would have been able to camp and exploit the available food sources. Several hundred flints were found, ranging from tools to flakes mostly made from small boulder clay pebbles (Emery, 2000). The Neolithic period brought with it the arrival of new techniques in agriculture such as the plough and domestication of animals. The traditional nomadic lifestyle transformed into a settled existence, which increased population size and brought about important social and environmental changes. Trees were cut down to build structures with the timber as well as to have more land area for cultivation. A stone axe, one of the most important Neolithic artifacts to be found throughout Britain was found roughly a mile away from the Poulton site along with a polished axe head (Emery, 2000). The one era that appears to be absent from Poulton archaeological records is the Iron Age, although other sites nearby in North Wales as well as in Cheshire have identified Iron Age presence (Emery, 2000).

Medieval Poulton

The Poulton Project (Poulton Research Project, 2014) began in aim to find the long-lost Cistercian Abbey, known to have existed in the twelfth and thirteenth centuries. The Poulton chapel was focused on in hope that it may give clues or indications as to where the abbey may be. The earliest known mention of Poulton is in the Domesday Book from 1086AD, compiled by the orders of William the Conqueror (Emery, 2000). This document has detailed records of the land ownership in medieval England. The sixty years following this have no surviving documentation for Poulton. The original charter for the Poulton Abbey foundation has survived and is the earliest of its kind in Cheshire, initially endowed by Robert Botelier, who was a member of the Earl of Chester's household. Money and land were granted to the monks of Combermere to establish a monastery at Poulton for the specific purpose of prayer and deliverance of the earl and his future well-being. Self-sufficiency was vital to the Cistercian way of life without dependency or interference from neighbours (Tobin, 1996). Despite this, the monks gained wealth quickly. Being innovative hydraulic engineers they transformed the landscape, reclaiming marshlands and wastelands, and established themselves predominantly as sheep farmers. Their wealth and success was also due to the Grange System; whereby uneducated men, known as the 'conversii', who followed the Cistercian Rule, were recruited for manual work. The Poulton Abbey land spanned over 3000 acres of which 1900 acres were arable. The monks would have consisted of only ~20 men and thus the conversii system was necessary for land advancements, making previously underdeveloped areas available. The Cistercians had a ruthless approach, making them the most successful farmers and businessmen in the Cheshire area. Traditions of land division and village life were often disregarded to make way for Cistercian farming land (Tobin, 1996).

Construction of the stone monastery was a lengthy undertaking, taking several years to complete. The church was always the first to be built, generally occupying the higher ground and one side of the quadrangle or cloister was situated to the northeast of the rectangular arrangement of buildings (Tobin, 1996). This, being a universal practice of the Cistercian order, led the initial investigation of the 'chapel' to dismiss any theory of it being physically part of the abbey (Emery, 2000).

Initial surveys raised some questions about the church, such as its location in the southeast of the monastic complex as opposed to the traditional northeast, its relatively small size, 26 m x 5 m, and its location on a gentle slope rather than on the highest point (Emery, 2000). Several phases of construction were revealed, showing a complex history. The first phase, a single-cell sandstone structure, 9 m x 4 m, is thought to be of Cistercian origin. It is possible that the

monks used stonework from an extant building immediately to the north, explaining the odd position of the church due to space limitations (Emery, 2000). Excavation of the surrounding graveyard has indicated quality burial space with bodies often stacked on top of one another. It has been estimated that 1200 burials exist, of which lie the remains of those who worked the surrounding land and their families. The accompanying materials, which overlap all historical periods, emphasizes the longevity of both the chapel and graveyard (Adkins & Adkins, 1982).

Disappearance of the conversii system and frequent warfare in the fifteenth century forced the Cistercian monks to lease out the Poulton estates to the Manleys, an influential Cheshire family. The chapel, which consisted of one room, was upgraded to a three-celled structure. The addition of a tower and chancel shows a change in status and function for the chapel with the Manleys household using it themselves as a private chapel. A slate roof replaced the original ceramic tiles. During the English Civil War (1642-1649) the chapel was used as stables and a lookout point. The chapel was visited in 1672 by a local antiquarian who declared it to be 'in great decay' until finally, in 1719, a church commissioner closed the chapel and its history with the words, '*The chapel is no more, there is nothing to be seen'* (Emery, 2000).

1.2 Genetic History of British Isles

The history of Britain's population is shaped by a number of immigrations (Schiffels *et al.*, 2016). This leaves an open question of how these have influenced the genetic composition of the current British population as well as the Poulton population. Multiple well-documented immigrations to the British Isles are recorded from the last 2000 years alone. These include military invasions and settlement by the Romans in the first century AD, Anglo-Saxons from the North Sea coast of Europe between ~400 and 650 AD, Scandinavians during the late Saxon 'Viking period' 800-1000 AD and Normans in 1066 AD (Schiffels *et al.*, 2016). Prior to this was the occurrence of depopulation during the last glacial maximum and subsequent resettlement by hunter gatherers (~ 7000 BC.), the appearance of sedentary agricultural communities (the Neolithic transition) (~ 4000 BC), the arrival and spread of Late Bronze-Iron Age and Celtic material culture (~ 1000-100 BC) followed by the previously mentioned Roman occupation and influence (43-410 AD) (Weale *et al.*, 2002).

These immigration events coupled with other previous population movements have shaped a complex ancestry of the current British population (Schiffels *et al.*, 2016). Although only a slight genetic difference between north and south exists (O'Dushlaine *et al.*, 2010), recent

studies have discovered fine-scale genetic structure in the Northern and Western parts of Great Britain, alongside noticeable homogeneity in Southern and Eastern England (Leslie *et al.*, 2015) in the regions where archaeologists identify early Anglo-Saxon artifacts, cemeteries and communities. Numerous estimates of the fraction of Anglo-Saxon genetic ancestry in England have been made (Capelli *et al.*, 2003, Thomas *et al.*, 2006, Töpf *et al.*, 2006), with the more recent fine structure analysis suggesting most likely 10-40 % (Leslie *et al.*, 2015). A study using new methodology presented whole-genome sequences of 10 ancient samples from archaeological excavations in East England dating to the late Iron Age and early and middle Anglo-Saxon periods. This study estimates that the modern-day East English population derives 38 % of its ancestry, on average, from Anglo-Saxon migrants, giving evidence for mixing of migrants and natives in the early Anglo-Saxon period, as well as that the Anglo-Saxon samples have close ancestry with modern-day Dutch and Danish populations (Schiffels *et al.*, 2016).

Another study reported ancient genomes of seven Northern British individuals from a Roman era York cemetery and analysed these in comparison to genomes from an earlier Iron Age burial, and a later Anglo Saxon burial. Six of the Roman-era genomes showed affinity with modern British Celtic populations, in particular Welsh, but were significantly diverged from populations from Yorkshire and other eastern English samples. Population continuity was suggested by observation of similarity with the earlier Iron Age genome, while differences from the later Anglo Saxon genome suggest an impact of migrations during the Anglo Saxon era. One Roman skeleton was of Middle East origin, confirming the diversity of the Empire (Martiniano *et al*., 2016).

Genomic data from 69 European individuals (Haak *et al.,* 2015) supports a view of European pre-history shaped by two major migrations. The first was the arrival of the first farmers from the Near East during the Early Neolithic and the second was the arrival of Yamnaya pastoralists from the steppe during the Late Neolithic. They further showed that these migrations were followed by resurgences of the previous inhabitants, where, during the Middle Neolithic, hunter-gatherer ancestry rose again as well as between the Late Neolithic and the present, with the reappearance of farmer and hunter-gatherer ancestry. Yamnaya ancestry is today higher in northern Europe while lower in southern Europe, and all European populations can be described as a mixture of western Europe hunter gatherer, Early Neolithic and Yamnaya. Some outlier populations show additional admixture with populations from Siberia and the Near East (Haal (Haak *et al.*, 2015).

1.3 Evolution of Ancient DNA Techniques

1.3.1 Background

Thirty-three years ago saw the beginning of the field of ancient DNA (aDNA) studies with the extraction and sequencing of DNA from the quagga, an equid (*Equus quagga quagga*) from South Africa that went extinct in the 19th century (Higuchi *et al*., 1984) and from an Egyptian mummy (Pääbo, 1985). The techniques used by these studies differ greatly from the new and optimised techniques frequently used today. These studies amplified small DNA sequences from skin fragments of the specimens using bacterial cloning and showed that the origin of the majority of the extracted DNA was microbial or fungal. They found the endogenous DNA to be very low concentrations of damaged, short fragments of multi-copy loci, such as mitochondrial DNA (mtDNA) (Rizzi *et al*., 2012).

The polymerase chain reaction (PCR) was developed and introduced a few years later, allowing for routine amplification and studies of surviving aDNA molecules, even if only in a single copy. This began a rapid increase and diversification into the studies and research of aDNA (Pääbo *et al.*, 1989). Due to the powerful ability of PCR to amplify even a few copies of DNA sequences, contamination by modern DNA has become a crucial problem. This issue has disputed and even disregarded many studies on aDNA, especially the more extravagant reports such as claims of DNA sequences surviving for millions of years in plants (Golenberg *et* al., 1990) and dinosaur bones (Woodward *et al*., 1994).

The introduction of new sequencing technologies over the last few years has brought the field of aDNA studies into a new era where what was once impossible is now possible. Examples of this include; drafting sequences of extinct specimens such as *Homo neanderthalensis* (Green *et* al., 2010) and differentiating endogenous from contaminant DNA in archaic *Homo sapiens* specimens (Krause *et al*., 2010), while more recent examples include; complete mitochondrial genomes of ancient canids (Thalmann *et al*., 2013) and a high-quality genome sequence of a Neanderthal woman from Siberia as well as a low coverage genome of a Neanderthal from the Caucasus (Prufer *et al*., 2014).

The methodological aspect of aDNA studies has an extensive and profound history, evolving from the 'classical methodology' of PCR amplification, cloning and Sanger sequencing to more recent next generation sequencing (NGS) technologies which are revolutionizing the field (Rizzi *et al*., 2012). The pioneering studies in the 1980s have been developed and

constantly improved by focusing on eliminating the two main limitations; the poor preservation of endogenous DNA and the presence of contaminating exogenous DNA.

The classical methodology follows three main steps: 1) amplification of several short and overlapping target fragments (60-200 bp long) by PCR to recover larger regions, 2) amplified fragments subject to the production and sequencing of several clones, 3) reconstruction of the final consensus sequence of the entire region of interest by aligning and comparing sequences from different clones and different overlapping fragments. This protocol was used to produce the first reconstruction of a DNA sequence from an extinct hominin, *Homo sapiens neanderthlensis* (Krings *et al.,* 1997). The results were later corroborated by additional mtDNA sequences from other Neanderthal specimens distributed all over Europe (Orlando *et al.,* 2006). Classical methods using PCR were applied to many studies, such as ancient human brain tissue, maize remains, human archaeological bones, dry skins of the extinct marsupial wolf and kangaroo rats, New Zealand moas and fossilized remains of plants and insects aged millions of years (Hagelberg *et al.*, 2014).

Almost all the genetic studies performed on ancient specimens targeted mtDNA regions, until more recently. mtDNA is maternally inherited and not subject to recombination, there are 100 to 10,000 copies of mtDNA in each cell and mtDNA mutations can be used to trace maternal lineages. These mtDNA characteristics have allowed for several successful studies to reconstruct the phylogenetic relationships between extant and extinct species such as Australian marsupial wolves (Thomas *et al*., 1989), New Zealand moas (Cooper *et al*., 1992), American ground sloths (Greenwood *et al*., 2001) endemic Hawaiian goose (Paxinos *et al*., 2002), cave bears (Hänni *et al*., 1994), Balearic Islands cave goats (Ramírez *et al*., 2009), giant lemurs (Orlando *et al*., 2008) and Caspian tigers (Driscoll *et al*., 2009).

Human mtDNA haplogroup nomenclature was introduced in the mid 1990s, with variation in Asian and American lineages assigned A-G labels, Europe labelled H-K, and L being assigned to the variation observed in Africa (Kivisild, 2015).

In comparison to our nuclear genes, mtDNA does not have introns or much non-coding sequences around them; the whole genome is packed densely (93%) with protein coding, ribosomal and transport RNA genes. Human mtDNA has also been described as not been protected by histones and hence is more vulnerable to damage than nuclear DNA (Kang & Hamsaki, 2005).

Tens of thousands of publicly available whole mitochondrial genome sequences exist, covering virtually all the extant populations of the world. This has been vital in addressing questions about demographic history of populations, natural selection, the extent of admixture and many more with regards to human populations. mtDNA sequences have been particularly important in the study of human evolutionary genetics. aDNA analysis combined with data from extant populations allows for increased understanding of the progressive change in genetic diversity in regions such as Europe (Brandt *et al.*, 2013).

It has always been said that aDNA samples typically contain more copies of mtDNA than nuclear DNA but the mtDNA that is sufficiently preserved for analysis is very low. Thus there is a greater chance of recovering longer intact strands of nuclear DNA than mtDNA as nuclear DNA is less prone to degradation and damage over time and that DNA damage events are also a less frequent occurrence in nuclear DNA than in mtDNA (Rizzi *et al*., 2012). These have been challenged by newer studies. Allentoft et al (2012) studying the half-life of DNA in fossil bones produced data suggesting that mtDNA degrades at a slower rate than nuclear DNA, which is consistent with another study by Schwarz et al (2012). This is possibly explained by mtDNA's circular structure making it less exposed to exonuclease activity. This emphasizes the importance of, if possible, focusing ancient DNA studies on both mitochondrial and nuclear genomes for optimal results.

Access to nuclear DNA allows for investigation into ancient human and animal phenotypes such as skin colour and behavioural traits. Krause et al (2016) addressed the issue of whether Neanderthals could speak, and if so, how, by analyzing the *FOXP2* gene, which is connected to humans' ability to speak. They found that Neanderthals share two evolutionary changes in *FOXP2* with modern humans, which differ in all other mammals, suggesting that the Neanderthals probably had the ability to speak like modern humans (Krause *et al*., 2016).

In the same period, a fragment of the *melancortin 1 receptor* (*MC1R*) gene from two Neanderthal remains was amplified and sequenced (Lalueza-Fox *et al*., 2007). In humans and other mammals, the *MC1R* gene regulates pigmentation. *MC1R* variants having reduced function are particularly associated with pale skin colour and red hair in European originating humans. The study revealed that both Neanderthal specimens carry a mutation in the *MC1R* gene that is completely absent from 3700 samples analysed from modern humans. Functional analyses also showed that the *MC1R* variant in Neanderthals decreases *MC1R* function to the equivalent level that alters hair and skin pigmentation in modern humans. These findings point to the fact that Neanderthals would have varied in pigmentation levels in a similar scale

to modern humans as well as that *MC1R* variants have evolved independently in Neanderthals and modern humans.

These are just two examples of the type of information nuclear DNA can provide. Other studies, more closely related to our study, access nuclear DNA for authenticating ancient human DNA sequences such as the study of the Neolithic transition in Scandinavia (Malmström *et al.*, 2014)) and the migrations of Norwegian Vikings (Krzewińska *et al.*, 2014), proving that next-generation sequencing promises to revolutionize the genetic study of the Neanderthals and other potential human ancestors (Hagelberg, 2014).

1.3.2 Damage and Contamination

Enzymatic repair processes continuously maintain the integrity of DNA molecules within living cells (Lindahl, 1993). When an organism dies, the cellular components that normally remove catabolic enzymes break down and consequently DNA is rapidly degraded. Bacteria, fungi and insects that feed on macromolecules also degrade DNA molecules (Eglinton & Logan, 1991). DNA may escape enzymatic and microbial degradation under rare conditions, such as if tissue becomes rapidly desiccated after death or the DNA becomes adsorbed to a mineral matrix (Pääbo *et al.*, 2004).

The most prevalent type of DNA damage to sub-fossil and fossil remains is degradation to a small size found to be around 100 to 500 bp (Pääbo *et al.*, 2004). In 2013, Dabney et al (2013) presented an improved silica-based extraction protocol enabling efficient retrieval of fragments shorter than 50 bp. Reduction in size of DNA after death is due to the enzymatic action as well as non-enzymatic hydrolytic cleavage of phosphodiester bonds in the phosphate sugar backbone that create single-stranded nicks. Glycosidic bonds that are present between the sugar backbone and nitrous bases face hydrolytic cleavage, resulting in abasic sites (Pääbo *et al.*, 2004). Upon release of a nucleotide, the abasic site can undergo a chemical rearrangement promoting the occurrence of a strand breakage at a rate slightly slower than or similar to base loss (Friedberg, 2003). The degree of degradation from these processes depends on the features of preservation and will differ greatly among samples, even among museum specimens of the same age. Fragments as long as a few hundred base pairs (Cooper *et al.*, 2001) or even more than 1 kb (Lambert *et al.*, 2002) can sometimes be amplified.

The length able to be amplified is limited not only by degradation and strand breaks but also by lesions that cause blocks to amplification by the *Taq* polymerase. These lesions are

induced by free radicals such as hydrogen peroxide (H_2O_2), peroxide radicals (O_2) and hydroxyl radicals (OH). Sites undergoing major oxidative attack are the double bonds of pyrimidines and purines, which lead to ring fragmentation. Other sites susceptible to oxidation are the chemical bonds of the deoxyribose residues which results in fragmentation of the sugar ring (Friedberg, 2003).

Damage can also occur in the form of cross-links that also block the DNA polymerase. In addition to these there are both known and unknown types of damages that are common in aDNA which are problematic, not because they prevent amplification, but because they cause incorrect bases to be incorporated during PCR, i.e. miscoding lesions. In 2001, it was shown for the first time that most of these damage-derived errors are caused by hydrolytic deamination of cytosine (C) into uracil (U), leading to apparent substitutions of C to tyrosine (T) or guanine (G) to adenine (A) in DNA templates sequenced after DNA amplification (Hofreiter *et al.*, 2001). aDNA samples were subject to pre-treatment with uracil N-glycosylase to remove uracil residues and leave abasic sites that prevent replication by the *Taq* polymerase during PCR. This method was not widely adopted due to uracil N-glycosylase possibly destroying all amplifiable templates of small samples. Another approach was reported in 2007; Single Primer Extension (SPEX), an amplification technique which, unlike PCR, uses single biotinylated primers to specifically target one strand of the aDNA template at a locus of interest without a predefined target length (Brotherton *et al.*, 2007). Although results comparing SPEX to standard amplification procedures showed that SPEX can produce sequence data of unprecedented accuracy, it has not been widely applied due to its intrinsic laborious protocol compare to other innovative approaches (NGS).

Even well preserved DNA from ideal conditions will degrade over time. Classical methods of amplification, cloning and sequencing is usually inhibited by the short length of the recovered DNA fragments as well as the small amount of template DNA from the specimen. Longer sequences usually provide more information, but when using classical methodology, only shorter overlapping fragments, meticulously assembled together, are available. This meant that it was not possible to sequence DNA longer than 1000 bp. In 2005 a breakthrough in ancient DNA from a study by Krause et al., using an innovative multiplexing strategy to reconstruct long DNA sequences from several shorter fragments, reported the complete sequence of the 16,770 bp mitochondrial genome of the Pleistocene woolly mammoth (*Mammuthus primigenius*) starting from 200 mg of bone (Krause *et al.*, 2006). The multiplexing method consists of a two-stage PCR: 1) multiple primer pairs target sub-sequences within the complete DNA sequence and 2) the amplified product is split into

aliquots of the same number of primer pairs which are then templates for secondary multiplex PCR by individual primer pairs.

In another attempt to improve classical methodology, blocking oligonucleotides that preferentially bind modern contaminant DNA are used to prevent amplification of the contaminating molecules (Vestheim & Jarman, 2008). This is used in conjunction with standard primers that are specific for the specimen of interest. This novel method has been tested on four Neanderthal samples of different contamination levels and taphonomic conditions (Gigli *et al*., 2008), significantly increasing the Neanderthal DNA sequences yield in all four samples from 25.23 % up to 90.18 %.

A serious problem for aDNA studies is contamination, especially when using human aDNA in comparison to animal or plant aDNA. It is now known that high temperature climates do not help DNA preservation. For this reason early results from human remains collected from regions of hot climates such as Florida (Hauswirth, 1994) and Egypt (Pääbo, 1985) are now considered to probably come from modern human contaminants. Several papers have shown that even with the use of thorough and rigorous protocols (Hofreiter *et al*., 2001), modern human DNA contamination is still present in the amplified products of ancient specimens (Krings *et al*., 1997, Kolman & Tuross, 2000).

It has also been shown that even with extensive treatments of UV and bleaching, it is impossible to remove modern human DNA from ancient bones and teeth (Gilbert *et al*., 2005). This is probably due to the porosity of bone and tooth dentine, being the main entry routes for DNA from sweat, skin fragments and exhaled cells. Well-preserved teeth directly removed from the jaw or maxilla appear less prone to contamination than bone fragments (Rizzi *et al.*, 2012). It is thus essential to collect skeletal material that has only been handled with gloves and facemasks during excavation and selecting remains with well-known taphonomic history, when using the classical methodology for aDNA analysis.

Many excavated specimens are found to contain DNA from numerous individuals (Gilbert *et al*., 2003), which raises the issue of authenticating ancient human DNA sequences, such as Neanderthals or distinct modern human groups like the Andaman Islanders (Endicott *et al.*, 2003), when they do not differ from potential contaminants (modern human).

1.4 The move to Next Generation Sequencing (NGS)

1.4.1 Background

The main tools used during the classical methodology era for DNA sample analysis were PCR and Sanger sequencing. With the development of miniaturized gel electrophoresis (capillary electrophoresis) and the automation of reactions, gel loading and signal detection, the Sanger method became the gold standard for DNA sequencing. Although having these notable features, Sanger sequencing has a low throughput and is thus expensive for large-scale sequencing. As well as amplification, library preparation and colony preparation are time consuming with a low efficiency. These are critical drawbacks for aDNA studies and hence the development of Next Generation Sequencing (NGS) was vital, opening up new possibilities, extending the field of applications (Millar *et al.*, 2008).

NGS has made it possible to increase the number of bases sequenced per run, whilst decreasing sequencing costs by improving the technology used. Single molecule sequencing is a technology able to read through DNA templates in real time without amplification, providing accurate sequencing data and potentially long-reads. NGS has allowed for this single molecule sequencing and efforts have focused in this direction (Zhang *et al*., 2011).

The most important NGS platforms for aDNA analyses during early development were the 454/Roche FLX and the Illumina Genome Analyzer (Shendure *et al.*, 2004), whilst now only Illumina is relevant. Both platforms follow the same basis for sequence production but differ in amplification and sequencing chemistry resulting in different throughputs. NGS sequencers, despite their high sensitivity and productivity, have detection systems not sensitive enough to measure the sequencing signal originating from a single molecule. The detection systems for both NGS platforms are able to identify a signal only if is generated by millions of DNA molecules and it is therefore necessary to amplify the sequencing library. The three key steps for generating reads are: a) library preparation, b) library amplification and c) sequencing.

a) Library preparation

The 454/Roche and Illumina platforms require the preparation of a library of the aDNA fragments ligated at both ends to specific DNA adapters for DNA amplification (Rizzi *et al.*, 2012). DNA is isolated from the ancient specimen and double strand DNA (dsDNA) is polished at the 5' and 3' ends and converted to blunt end DNA. To achieve the polishing step,

DNA polymerase and polynucleotide kinase are used simultaneously to catalyze phosphorylation at the 5' fragment end. Adaptors are then ligated to the polished and phosphorylated ends. Adaptors are short oligonucleotides of known sequence, allowing for the design of complementary primers for library amplification and sequencing downstream.

b) Library amplification

The 454/Roche platform amplifies the DNA library in a water-in-oil emulsion PCR. Each DNA molecule in the library is bound to a bead and amplified in an aqueous droplet. The Illumina platform differs, utilizing an isothermal bridge amplification process run on a glass slide where molecules are amplified independently, which creates spatially distinct "clusters". Library quantification is a vital step for both platforms in order to obtain high quality sequences

c) Sequencing approaches

Depending on the available ancient material as well as the research goal, different types of DNA target can be sequenced;

Shotgun sequencing

This is performed when the extracted DNA is sequenced without any prior inferred selection. This approach has the capacity to identify all the known species when total DNA has been isolated from bone, teeth or shaft specimen. With these type of samples, the amount of endogenous target DNA can be very low due to contamination by bacteria and fungi. The shotgun approach is also used for metagenomic studies, when the sequencing goal is to identify all possible known organisms present in an isolated specimen. Poinar and Miller utilized this approach for the sequencing of mammoth DNA whereby they confirmed the presence of large amounts of exogenous DNA (Poinar *et al*., 2006, Miller *et al*., 2008). Sequences obtained using this approach are usually identified with a 'blasting' protocol (BLAST) whereby sequences are matched to sequence databases. Parameters are carefully selected so as to avoid non-specific or incorrect results. Methods are taken to decrease the amount of microbial contamination that is high when using this sequencing approach. Green et al (2010) proposed an enzymatic digestion of the sequencing library with restriction enzymes that degrade DNA fragments with a GC composition similar to bacterial genomes.

Amplicon sequencing

This strategy uses PCR, specifically when the target is well known and the goal is to detect SNPs or small variants used as markers for haplotyping. This approach has been used to

successfully identify nucleotide variants in short sequences of specific genes such as blood group determiners, taste perception and brain development in Neanderthal samples (Lari *et al*., 2015).

Sequence capture

This approach uses specifically designed probes to recognize and capture a target DNA, allowing enrichment of the sample as well as information recovery on DNA misincorporations at the 3' and 5' ends. The primer extension capture (PEC) was the first capture strategy, which uses biotinylated primers exclusively designed for identification of particular regions and to allow extension to continue until the end of the DNA fragments. PEC was used to capture and sequence the entire Neanderthal mitochondrial genome from whole DNA isolated from an ancient specimen and converted into a labeled NGS library (Briggs *et al*., 2009). This capture method has a very high specificity and improved the capacity for complete mtDNA recovery of complex samples, yet the step of synthesizing biotinylated primers is very expensive. Probe enrichment methods are available for numerous specimens for the study of specific regions of the genome or whole genome analysis (Choi *et al*., 2009), in either solid- or liquid-phase forms. The solid-phase method immobilizes probes on an array surface, the liquid phase is similar yet the supporting material for capture consists of beads suspended in a buffer instead of a solid array (Aird *et al*., 2011).

1.4.2 DNA sequences and Data Analysis

Depending on the nature of the sample as well as availability of a reference sequence, sequencing projects are generally considered either *de novo* or *re-sequencing* studies (Rizzi *et al*., 2012). *De novo* differs from *re-sequencing* as it requires the generation of an informative and robust sequence of higher sequencing depth. In order to construct previously unknown sequences, specific *ad hoc* algorithms are used to generate different libraries of different fragment sizes as well assemble the reads. aDNA sequencing studies are *re-sequencing* projects because a reference sequence is required to assemble short reads of damaged DNA. If a new ancient specimen is discovered, sequences need to be adequately selected for reference. In anthropological studies, the *Homo sapiens* genome will be used when the specimen of interest is a close relative such as Neanderthal (Green *et al*., 2010) or Denisovan (Krause *et al*., 2010).

Analysis of re-sequenced DNA can be done to differentiate phylogenetically related species or to identify nucleotide variations specific to a given species. aDNA reads are mapped on a

reference sequencing using BLAST to detect variations in sequences. The actual read can originate not only from the aDNA sample but also from contaminating modern human DNA and hence different criteria are used in order to differentiate between the two. aDNA fragments are generally short, being less than 120 bp, but this is not enough to distinguish from modern DNA. Miscoding lesions present at the 3' and 5' ends of aDNA (Briggs *et al.*, 2007) with higher frequencies of base substitutions from C to T and G to A (approximately 35 % in Pleistocene ancient human samples) compare to modern human samples. The combination of reduced fragment lengths and terminal miscoding lesions is the main feature used to successfully identify and distinguish aDNA from modern DNA.

1.5 Ancient DNA samples and Preservation

The most limiting factor in ancient genomic research is poor DNA preservation. A study focusing on hindering this limitation demonstrated that when targeting the outer layer of roots of teeth (cementum) they obtained up to fourteen times more endogenous DNA than when using the inner dentine (Damgaard *et al.*, 2015). Duplicated results were observed for a variety of aDNA samples from different archaeological contexts. Their results strongly support targeting the cementum-rich root surface when teeth are available for aDNA studies.

The cementum layer in teeth roots and the inner part of the petrous bone, owing to their high levels of endogenous DNA, are currently recognized as the two optimal substrates for genomic analysis (Damgaard *et al.*, 2015, Pinhasi *et al.*, 2015, Gamba *et al.*, 2014, Adler *et al.*, 2011).

A study by Hansen et al (2017) did a comparative analysis of the DNA preservation of these two substrates whereby samples were obtained from the same human skulls from a range of different ages and environments. From their results, they reiterated that tooth cementum and petrous bones are excellent substrates for ancient genomic research, they observed high sample-to-sample variation but concluded that petrous bone performed better overall than cementum. Having said this, they also state that teeth with good molecular preservation performed just as well and sometimes better than petrous bones of the same individual, showing a link between visual and molecular preservation. They also found a higher C-T damage rate for petrous bones as well as a smaller ratio of mtDNA to nuclear DNA in comparison to cementum. This shows that there are pros and cons for using both substrates.

Many challenges arise when sampling ancient human remains with regards to destructive sampling and it is vital to maximize continued preservation of the sample for other non-destructive research. Only two petrous bones are present in each skull and a visible hole is left in the inferior part of the skull when one is sampled which may be problematic for precious ancient skulls. Strontium isotope ratios in petrous bone are able to provide geographic location information of a child during pregnancy (Harvig *et al*., 2014) as well as childhood stable isotopic dietary signals (Jørkov *et al*., 2009). This information is lost if the entire otic capsule is used during DNA extractions. If a large part of the petrous bone is removed for sampling, sex (Norén *et* al., 2005) and childhood disease (Homøe *et al*., 1992) information may also be lost.

Tooth sampling can also be damaging, particularly when only a small number of teeth or only one remain. The clear downfall is losing value for exhibition if teeth are removed from the skull, but morphological studies of teeth can also provide important information regarding population affinities (Scott and Turner, 2000) and analyses of tooth wear can give insight into the diet and age of an individual (Molnar, 1971). The enamel holds information about geographic location during childhood, which can be revealed by strontium isotopic analyses, while tooth calculus is an excellent substrate for ancient proteomic studies (Warinner *et al.*, 2014)). Yet, morphological and biomolecular analyses of tooth crowns does not have to be compromised by aDNA sampling when only sampling the root (Damgaard *et al*., 2015).

It is therefore important to consider each skeleton individually and carefully decipher how to minimize the level of destruction to precious ancient material. A number of loose teeth were available from Poulton skeletons and hence cementum was chosen as the substrate to maximize extraction of endogenous DNA and minimize destructive analysis.

1.6 DNA Library Preparation Techniques

1.6.1 Double strand DNA Library Techniques

After DNA is isolated from an ancient specimen, DNA libraries are to be built for sequencing platforms (Illumina). Traditional methods for dsDNA libraries require ligation of adaptors to both the 5’ and 3’ ends of the target DNA molecules (Wales *et al*., 2015). Adaptor ligation lacks efficiency and is a slow process that requires nanogram (ng) amounts of input DNA (Turchinovich *et al*., 2014). Complementary DNA (cDNA) libraries created using adaptor

ligation methods contain adaptor by-products from cross- and self-ligation. For the contaminating by-products to be removed, before and after pre-amplification, purification steps are required, adding time and cost to the library preparation method (Turchinovich *et al*., 2014).

Traditional library preparation methods use dsDNA as starting material. A library preparation technique for highly multiplexed target capture and sequencing has been well described and tested. This method is robust and works with nanograms amounts of DNA (Meyer & Kircher, 2010), ideal for ancient DNA specimens such as Poulton samples. dsDNA libraries will thus be built using this Meyer library method for the aim of reassessing DNA preservation of Poulton material.

1.6.2 Single strand DNA Library Techniques

Accessing single-strands of DNA for library preparation may significantly increase the yield of endogenous DNA. In 2013, Gansauge and Meyer (2013) described a new library preparation method using single-stranded DNA (ssDNA) instead of dsDNA. When preparing libraries using ssDNA they found an increase in endogenous content, particularly when the dsDNA libraries contained less than 3 % endogenous DNA. Other studies using the same method reported endogenous DNA content to increase more than 20 fold in some samples (Bennett *et al.,* 2014), although these studies did not have consistent results, wherein some cases the endogenous content was considerably lower when using ssDNA library preparation (Wales *et al*., 2015). Reports from two studies showed that using the same DNA extract at the same volume input, more complex libraries were obtained with the ssDNA method over the dsDNA method. More complex libraries allow for deeper and more accurate sequencing (Wales *et al*., 2015). One of these afore mentioned studies by Prüfer et al (2014) reported slightly higher endogenous content in Altai Neanderthal ssDNA libraries in comparison to a dsDNA library. This study shows the potential impact ssDNA methods have on aDNA samples.

Wales et al (2015) set out to discover whether using the ssDNA method would be more beneficial than current dsDNA methods using a range of samples. They noted that a significant influence on the success of the library preparation is likely due to degradation patterns of specific samples. Bennett et al (2014) studied archaeological skeletal remains from warmer locations than most of the samples studied by Wales et al (2015) whereby bones would have been exposed to an increased microbial action soon after death, meaning

increased DNA fragmentation and environmental contamination. It was also noted that samples with more fragmented endogenous DNA than exogenous DNA, the ssDNA method is biased against the longer contaminating molecules that may not denature, while successfully recovering the shorter endogenous DNA. This proving to be a very useful method for highly fragmented DNA samples, such as aDNA from the Poulton excavation site.

If dsDNA libraries contain less than 0.5 % endogenous DNA with mean read lengths between 70 and 100 bp, large fold-increases may be obtained using the ssDNA method, as seen in Bennett et al (2014). ssDNA methods are more biased towards short DNA fragments, making this method again viable for use on aDNA from Poulton. ssDNA methods should out compete dsDNA methods especially in terms of complexity of the library (Wales *et al*., 2015). It has also been noted that ssDNA methods are less biased toward high GC content than dsDNA libraries, an important consideration for highly degraded samples (mean fragment length >50 bp), where dsDNA methods will underrepresent AT rich regions, yielding less homogenous genome coverage than ssDNA libraries (Wales *et al*., 2015).

Although this ssDNA has many described benefits, it is both expensive and time consuming (Wales *et al.,* 2015).

Capture and Amplification by Tailing and Switching

Another recently described ssDNA library method is 'Capture and Amplification by Tailing and Switching (CATS) (Turchinovich *et al*., 2014). CATS is a cheaper, more efficient and quicker method for library preparation that does not require adaptor ligation of 5' and 3' adaptors to the fragmented RNA and DNA molecules of interest. Current techniques require 10-100 fold higher inputs of DNA and RNA compared to the CATS method and are much more time consuming and expensive (Turchinovich *et al*., 2014). The CATS method can be used to generate ready-to-sequence DNA libraries from pictogram amounts of DNA (or RNA) molecules in a few hours. Small (< 150 bp) DNA can be used as an input directly, while longer DNA molecules are first fragmented by sonication.

The process of the CATS method is depicted in Figure 2. Briefly, DNA fragments are polydeoxyadenylated (poly(dA)) with terminal deoxytransferase and subsequently a cDNA strand is synthesized in the presence of the anchored poly(dT) oligonucleotide containing a custom 3' adaptor sequence. Reverse transcriptase (RT) is added and when it reaches the 5' end of the DNA template, the enzymes terminal transferase activity adds additional nucleotides (dC). The template switching oligonucleotide (TSO), having three 3' terminal rG

nucleotides and a custom 5' adaptor sequence is added and acts as a second template for the RT. It is thought that template switching is initiated by the interaction of the three consecutive rG nucleotides and the dC-rich extended sequence of the cDNA. Standard PCR with a forward primer complementary to the 3' terminus of the first cDNA strand amplifies a second cDNA strand. The reverse primer is complementary to the 3' terminus of the second cDNA strand and hence amplification of the cDNA occurs (Turchinovich *et al*., 2014).

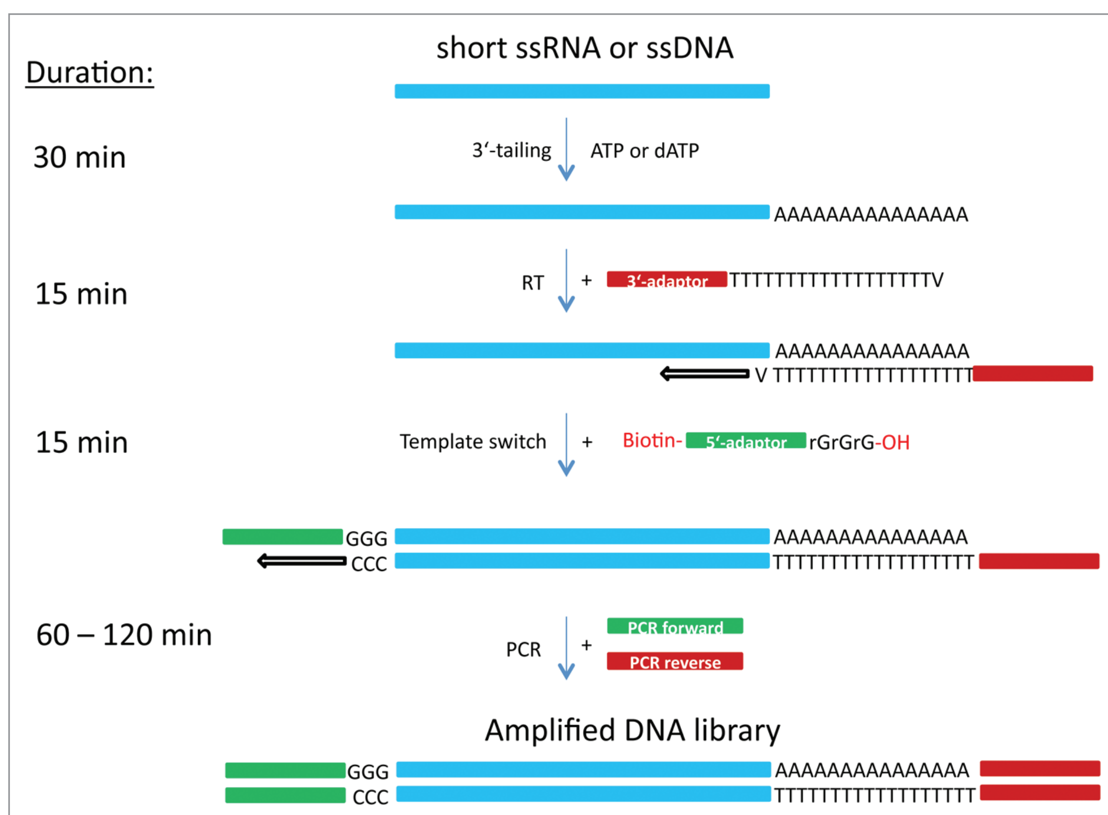


Figure 2: Schematic representation of the CATS library preparation method using the combination of poly(A) or poly(dA) tailing and template switching capacity of MMLV-RT (Turchinovich *et al*., 2014)

This technique can benefit archaeological and forensic sciences allowing for deep sequencing of highly degraded and small amounts of DNA and RNA (Turchinovich *et al*., 2014). This method will potentially provide us with a very sensitive and novel method for accessing ssaDNA and will be implemented in this study using the Poulton material. This will be the second aim of this study.

1.7 Aims and Objectives

The overarching aim of this study is to reassess DNA preservation in human samples excavated at Poulton. This will be achieved by realizing the following objectives:

1. Identify macroscopically well-preserved teeth from which cementum can be isolated
2. Extract DNA from that cementum using protocols optimized for ancient DNA analysis
3. Extract DNA from long bones and test the hypothesis that cementum should produce better results
4. Construct Illumina sequencing libraries and obtain >5M sequence reads/library/sample
5. If DNA is preserved, provide a first insight to the genetic affinity of the Poluton population using mtDNA and nuclear DNA

The working hypothesis of this study is that DNA is poorly preserved in Poulton individuals (Town, 2015). We predict that targeting cementum (Damgaard *et al*., 2015) and using NGS techniques (Millar *et al.*, 2008) will allow sequencing of at least some authentic ancient DNA. The expectation of poorly preserved DNA also leads to the secondary aim of this study; to test a more recently described ssDNA library method (CATS) for potentially yielding better quality DNA results, than the Meyer library method.

Chapter 2: Methods

2.1 Sampling and Powdering

Poulton human samples

Access to the collection of Poulton remains at Liverpool John Moores University was gained for this study. 25 samples were chosen from the most recently excavated skeletons, as these were the least handled by students (2015 and 2016 excavations). All samples have been kept at room temperature. 20 loose teeth (i.e. no longer attached in the mandible or maxilla) and 5 fragmented long bones (arms and legs) were chosen. The 2015 samples had previously been washed by students whilst the majority of the 2016 samples were unwashed.

Targeted sampling/powdering

The samples underwent the following cleaning/wash procedure: samples were wiped with paper towel soaked with 1 % sodium hypochlorite (bleach) until there was no more discoloration on the towel and then wiped with molecular grade water and ethanol. Once the bone and teeth sample were prepared in this manner, the drill (at a low setting of ~100 rpm with small drill bits) was used to lightly remove the required amount of bone powder (cementum: 20mg - 40 mg, long bones: 50mg - 100 mg). In some cases, a small amount of dentine was removed for small teeth unable to provide enough cementum powder. Long bones were covered with aluminum foil, exposing only the area to be accessed for bone powder so as to prevent contamination. The powdered sample was placed into 2 ml Eppendorf tubes, labeled and stored at 7 ° C.

2.2 Ancient DNA Authentication

Laboratory work was performed in the dedicated clean laboratory facilities in the Biological Sciences Building at Liverpool John Moores University according to strict aDNA standards, following Gilbert et al (2005) where possible. aDNA extractions and libraries were carried out in a clean laboratory exclusively dedicated to aDNA manipulation whereby a laminar flow cabinet was used. Bones were also powdered in this laboratory in a ventilated fume hood. PCR and post PCR analysis were done in a separate main laboratory. All surfaces in the aDNA laboratory were cleaned with 5 % bleach and 70 % ethanol and were periodically sterilized by UV irradiation. Equipment and utensils were wiped with bleach and ethanol and UV irradiated in a cross linker at 254 nm for 5 minutes before use. Aqueous solutions were also UV irradiated as above for ~ 10-30 minutes. Protective gear was worn including hooded

forensic suits, face masks, eye covers, hair nets, over sleeve arm covers, shoe covers and two pairs of sterile gloves at all times. To detect possible contamination extraction, PCR, library preparation controls were always undertaken. Positive controls were used during PCR where possible. In addition to this, indexed adapters were also used for amplifying DNA libraries such that any possible contamination entering samples after leaving the aDNA clean laboratory could easily be detected.

2.3 DNA Extractions

DNA extractions followed a modified single silica-based DNA extraction method by Rohland and Hofreiter (2012) (Dabney *et al*., 2013). An extraction buffer was prepared (Urea [8 M] and EDTA [0.5 M]) and 1 ml was added to each sample, followed by the addition of 20 µl Proteinase K [10 mg/ml] and stored rotating at 37 °C for an overnight digestion. The overnight samples were vortexed to ensure optimal digestion of the powder. Samples that appeared inadequately digested, i.e. a lot of powder was still present, were vortexed and stored shaking at ~45°C for ~2 hours and vortexed again. Samples were then spun down at 13000 rpm (14549 x *g*) using Eppendorf centrifuge 5418. PB, EB and PE buffers supplied in QIAquick® PCR Purification Kit (Qiagen) were used. 1 ml of supernatant and 5 ml PB (5 X) were added to the spin column (Amicon filters) and spun for 2 minutes at 1500 rpm (Eppendorf centrifuge 5810 R) and flow through discarded. 1 ml buffer PE was added to the spin column, spun for 1 minute at 1500 rpm and flow through discarded. Another 'empty spin' at 1500 rpm for 1 minute was done to remove residual alcohol from the column. 105 µl buffer EB was added to the spin column, left for 5 minutes to reconstitute the DNA, and spun for 1 minute at 1500 rpm and the flow through collected. The DNA concentration (µg/ml) as well as purity (A260/280) was read using a Thermo Scientific NanoDrop 2000c UV-Vis spectrophotometer. DNA was stored at -23 °C before library preparation.

2.4 DNA Library Preparations

Libraries of the extracted DNA samples were prepared following both the standard protocol (Meyer & Kircher, 2010) as well as the newer CATS protocol (Turchinovich *et al*., 2014). Adapters for the CATS protocol were modified to be compatible with the Meyer adapters and hence compatible with the Illumina sequencing primers, such that samples prepared from both library preparation methods could be sequenced at the same time, saving both time and

money. SciLife (Sweden) sequenced the samples, where the TruSeq Universal Adapter is used. All primer sequences are in Appendix 2 (Tables A1 – 4).

2.4.1 Meyer Library Preparation

Adapter mix preparation

Hybridization mixes for adapter P5 and P7 was made up at 200 µM each. For P5, IS1_adapter_P5.F [200 µM], IS3_adapter_P5+P7.R [200 µM], oligo hybridization buffer [1X] and water to make up final volume, were mixed together. For P7, IS2_adapter_P7.F [200 µM], IS3_adapter_P5+P7.R [200 µM], oligo hybridization buffer [1X] and water to make up final volume, were mixed together. Both mixes were incubated in a thermal cycler for 10 seconds at 95 °C, then decreased to 12 °C at a rate of 0.1°C/sec. Mix P5 and P7 were then combined to provide a ready-to-use adapter mix [100 µM each adapter].

Blunt end repair

Tango Buffer [1X], dNTPs [100 µM], ATP [1 mM], T4 PNK [0.5 U/µl], T4 DNA Polymerase [0.1 U/µl], DNA and water to make up to final volume, were mixed together. This mix was then incubated in a thermal cycler for 15 min at 25°C followed by 5 min at 12°C and then purified using MinElute following the kits protocol (Qiagen).

Adapter Ligation

A master mix was prepared for the required number of ligation reactions; T4 DNA ligase Buffer (1X), PEG-4000 (5 %), adapter mix (2.5 µM each, 1:10 dilution), T4 DNA ligase (5 U/µl), DNA and water to make up the final volume were combined. This mixture was incubated for 30 minutes at 22°C then purified using the MinElute PCR Purification Kit (Qiagen).

Adapter Fill in

The following reagents were mixed together; Thermopol buffer [1X], dNTPs [250 µM], *Bst* polymerase (large fragment) [0.3 U/µl], DNA and water to make up the final volume and incubated for 20 minutes at 37 °C and heat-killed for 20 min at 80 °C.

qPCR

Prior to library amplification, quantitative PCR (qPCR) was performed to determine the number of cycles required for optimal amplification of the libraries. The following PCR mix was set up with a duplicate made for each sample as well as a blank (whereby water replaced

the library DNA): Maxima SYBR Green qPCR master mix [1X], Forward primer (IS4) [200 nM], Reverse primer (any indexing primer) [200 nM], DNA library and water to make up final volume (e.g. 25 µl). PCR steps followed; Initial denaturation (95 °C for 10 minutes, 1 cycle) and 35 cycles of; Denaturation (95 °C for 30 seconds, 1 cycle), Annealing (60 °C for 30 seconds, 1 cycle) and Elongation (72 °C for 30 seconds, 1 cycle). Fluorescence measurement was carried out at the end of each extension step. Linear and log plots of the qPCR process were analyzed to determine baseline and threshold and thus define the adequate number of cycles for each sample. Results from qPCR are included in Appendix 4.

Amplification

6 reactions per DNA library were set up in 25 µl reactions containing: TaqGold Buffer [1X], $MgCl_2$ [2.5 mM], dNTPs [0.25 mM], 10 µM IS4 (Forward Primer) [0.2 µM] and ddH_20 to make up final volume. This was vortexed lightly and left briefly to mix. The following was then added; AmpliTaq Gold [0.1 U/µl], Index primer [0.2 µM], Library DNA (3 µl). The reactions were mixed and spun down and the following thermal cycler program was used: Initial denaturation (94 °C for 12 minutes, 1 cycle), X cycles (dependent on qPCR analysis) of; Denaturation (94 °C for 30 seconds, 1 cycle), Annealing (60 °C for 30 seconds, 1 cycle) and Elongation (72 °C for 45 seconds, 1 cycle), Final extension (72 °C for 10 minutes, 1 cycle) and Hold (4°C for ∞). X varied for each sample (from 8 to 15).

Pooling and purification used AMPure XP (Agencourt-Beckman Coulter A63881) following the manufacturer's guidelines. After amplification, the concentration and size profiles of the purified libraries were determined on a Bioanalyzer 2100 using the High Sensitivity DNA chip (Agilent) for DNA visualization. None of the extraction or PCR blanks showed signals of DNA on agarose gels and Bioanalyzers and were therefore not further sequenced.

2.4.2 Capture and Amplification by Tailing and Switching

Notes:

1) The molarity of IPd primers must be approximately proportional to the molarity of the poly (dA) tails and must not be used at more than 1 µM final concentration, i.e. the following were used dependent upon DNA concentration:

- 10 µM IPd [1 µM final] for 10-100 ng/µl range DNA
- 1µM IPd [100 nM final] for 1 ng range DNA
- 100 nM IPd [10 nM final] for 100 pg range DNA

- 10 nM IPd [1nM final] for 5 pg range DNA

whereby we used [100 nM] final concentration IPd.

2) Up to 7 µL of poly(dA)-tailed DNA were used per 20 µL RT reaction
3) TSO final concentration was always 1 µM
4) The reaction was incubated for 2 hours with TSO, as this increases the yield of the library 5-10 times as compared to 15 min (appears 3 PCR cycles earlier on gel)

First-strand cDNA Synthesis/Tailing and Template Switching

16 µl of DNA (1 pg/µL – 5 ng/µL) was added to 2 µl 10X Terminal Transferase buffer and heated for 2 minutes at 95 °C and then fast cooled on ice. 0.5 µl T4 PNK was added and heated for 5 minutes at 37 °C. 2 µl dATP (the final concentration of dATP was added approximately linearly depending on DNA amount. e.g., 0.1 mM for 1-10 pg/µl DNA or 1mM for 1-10 ng/µl DNA) and 0.5 µl Terminal Deoxynucleotide Transferase [1X] was added and heated for 15 minutes at 37 °C and then for 20 minutes at 70 °C.

2 µl poly(dA) DNA reaction was added to 5 µl RT Buffer (1X), and to 2 µl IPd primer [1 nM – 1 µM] and heated for 2 minutes at 72 °C and 2 minutes at 42 °C. A RT master mix was prepared by adding the following (amount needed for 1 sample); 4 µl RT Buffer (5X) + 2 µl dNTP (10 mM) + 2 µl SMARTScribe RT + 0.5 µl RNAse inhibitor + 0.5 µl DTT (100 mM). 9 µl mastermix was added to 9 µl of sample/IPd, pipette mixed and heated for 15 minutes at 42 °C. 2 µl TSO [1 µM] was added, pipette mixed, and heated for 2 hours at 42 °C and 10 minutes at 70 °C.

Amplification

20 µl RT reaction, 50 µl *Taq* PCR Master Mix (2X), 10 µl PCR forward primer [10 µM], 10 µl PCR reverse primer [10µM] and 10 µl water were added together. The PCR procedure followed; Initial denaturation (94 °C for 30 seconds, 1 cycle), 13-23 cycles* of; Denaturation (94 °C for 15 seconds, 1 cycle), Annealing (62 °C for 30 seconds, 1 cycle) and Elongation (70 °C for 30 seconds, 1 cycle), Final extension (72 °C for 10 minutes, 1 cycle) and Hold (4°C for ∞). *Rough guide: Use 13 PCR cycles for 1 ng DNA and 23 PCR cycles for 5 pg DNA OR determine optimal cycle number prior to amplification by qPCR analysis.

The following two images (Figure 3 and 4) are what the DNA sequences should look like after amplification and show the differences in primers used in the original method versus our edited method, whereby the blue N = sample DNA/target sequence, Magenta = A Tail, Red = IPdTDPo primer, Green = 3' adapter (Illumina poly (dT) primer), Red = 5' adapter (5'-biotin

blocked template switch oligonucleotide (TSO)), Orange = Illumina Custom Sequencing Primer, Purple = Forward and Reverse Primers and Black = newly translated sequences.

Figure 3: Image created using Snapgene portraying the CATS Library preparation method (i.e. original paper using their primers): Blue N = sample DNA/target sequence, Magenta = A Tail, Red = IPdTDPo, Green = 3' adapter (Illumina poly (dT) primer), Red = 5' adapter (5'-biotin blocked template switch oligonucleotide (TSO)), Orange = Illumina Custom Sequencing Primer, Purple = Forward and Reverse Primers and Black = newly translated sequences

Figure 4: Image created using Snapgene portraying the CATS Library preparation method for compatibility with the Meyer libraries: Blue N = sample DNA/target sequence, Magenta = A Tail, Red = IPdTDPo, Green = 3' adapter (Illumina poly (dT) primer), Red = 5' adapter (5'-biotin blocked template switch oligonucleotide (TSO)), Purple = Forward (IS4) and Reverse (index 1) Primers (multiple binding sites) and Black = newly translated sequences

Purification

Libraries not sequenced were stored at 4°C for the opportunity to be used in further research projects.

20 µl of PCR reaction was run on agarose gel to examine library quality, fragment size distribution and the presence of the "empty" libraries (peak of about 150 bp size). If size selection was not required, the DNA library was purified with AMPure XP Beads (Beckman Coulter). The DNA library was quantified and average size estimated by Qubit and Bioanalyser before being loaded on MiSeq/HiSeq. (Illumina Custom Sequencing Primer should be used for Read 1).

Additional studies done with CATS

An initial comparison of CATS and Meyer library preparation techniques was implemented by drawing comparisons from the threshold cycle (Ct) values following amplification by qPCR. Ct is the number of cycles required to reach a threshold signal representing the lowest reliable signal to exceed the background in all samples. Small numbers of template (i.e. low copy number) at the start of a reaction will result in a qPCR curve with high Ct values, while large numbers of template (i.e. high copy number) will result in small Ct values (Lacey, 2012).

The CATS method was first tested on two samples, one of lower and one of higher concentration: sample 3 (3.0 ng/µl) and 10 (10.9 ng/µl). qPCR of these two samples along with Meyer library 10 (ASM010) showed that CATS library 10 appeared to have failed, whilst library 3 was amplified (Appendix 4 Figure A11). CATS was then used to build libraries of 5 samples (11,14,15,17 and 23). The qPCR was run along the equivalent Meyer samples to compare library efficiency. Only 1 CATS sample (17) amplified with similar Ct values as the Meyer libraries, whilst the other 4 samples were all less efficient (Appendix 4 Figure A12).

This prompted changing parameters of the CATS protocol to find which produces the most efficient libraries. It was decided to adjust incubation times of PNK and dATP (Table 1). Sample 23 was diluted 10 fold (20µl in 180 µl H_2O) to have enough sample availability across each of the incubation time tests. The following incubation times were set up within the CATS protocol and these libraries were amplified by qPCR (Appendix 4 Figure A13).

Table 1: The reagents and differing incubation times set for deducing optimal CATS protocol using three different reagent incubation times

Reagent	**Incubation times 1**	**Incubation times 2**	**Incubation times 3**
PNK	5 minutes	10 minutes	15 minutes
dATP	15 minutes	30 minutes	30 minutes

Although this test was done to find the optimal incubation periods, which was found to be the set number 2 (10 min PNK and 30 min dATP), this result also suggested that no difference between library efficiency is visualized when preparing the CATS library from DNA

extraction concentration versus 10 fold dilution. Since some of the DNA concentrations used were higher than those used in the original paper (protocol tested for 1 pg/µL - 10 ng/µL) it was decided to create libraries by CATS and Meyer methods for both diluted and undiluted concentrations of the same samples to compare the amplification curve and Ct values.

Five lower concentration and five higher concentration samples were selected to compare both non-diluted and diluted CATS libraries vs Meyer libraries; low concentration samples: 13, 14, 18, 19 and 20, high concentration samples: 8, 10, 11, 12 and 17. Both CATS and Meyer libraries were created for these 10 samples at non-diluted and diluted (10 fold) concentrations (i.e. totaling 40 libraries) including negative controls. The CATS libraries were created using the new optimal incubation times.

A comparison of diluted versus non-diluted samples was first set up for the Meyer libraries using a positive control (Meyer library 22 ASM022) (Appendix 4 Figure A14). There was not enough positive control to use for the CATS libraries and hence qPCR was performed again for Meyer libraries using a new positive control (diluted the positive control ASM022 from the PCR reactions at two dilutions: 1) 1 in 100 and 2) 1 in 200 (Appendix 4 Figure A15, A16 and A17). The positive control was also used for the CATS qPCR to have a standard positive control across all the reactions (Appendix 4 Figure A18 and A19).

The threshold line was set in the linear phase of the qPCR amplification curves, using the positive control, and the Ct value for each sample was thus determined. The difference in Ct values for diluted versus non-diluted was calculated for the samples from both library preparation techniques.

2.5 Sequencing

Statement: all DNA sequencing and post-sequencing processing of the DNA sequence reads were performed by others.

The amplified Meyer samples were sequenced at the NGI Stockholm sequencing facilities whereby a number of quality certified checks are employed. Sequences were then aligned to the reference human genome and summary statistics of 1) proportion human DNA, 2) average read length, 3) autosomal and mitochondrial genome coverage, and 4) contamination estimations based on mitochondrial DNA were obtained.

2.5.1 Sequencing Process and Alignment

Sequencing of purified libraries pooled at equimolar ratios was performed on the HiSeq X platform (100 base pairs paired-end reads) at the SciLife Sequencing Centre in Stockholm, Sweden. The raw data was processed on the semi-automated ancient DNA pipeline on the Uppmax Milou cluster (Uppsala Multidisciplinary Center for Advanced Computational Science) as part of the 1000 Ancient Genomes project at Uppsala University. In short, read pairs were merged into consensus sequences as adapters were removed using MergeReadsFastQ_cc.py (Kircher, 2012), whereby an overlap of at least 11 bp was required. These were mapped as single-end reads to the human reference genome using *BWA aln* version 0.7.8 (Li & Durbin, 2009) with the non-default parameters –n 0.01 –o 2 and disabled seeding as in (Lazaridis *et al.*, 2014). Consensus sequences were created using FilterUniqueSAMCpns.py (Kircher, 2012) by collapsing PCR duplicate reads with identical start and end coordinates. Reads not meeting the requirement of less than 10% mismatches to the human reference genome and a length of less than 35 bp were discarded.

2.5.2 Mitochondrial DNA Haplogroups

Consensus sequences of the mitochondrial genomes of all samples were called using *mpileup* and *vcfutils.pl (vcf2fq)* from the *samtools* package (Li *et al.*, 2009). Reads were to have a minimum base quality of 30 and a minimum coverage of 3 X to call confident bases. This too was done by SciLife, providing us with the mtDNA consensus sequences in FASTA format.

The consensus sequences were then delivered to LJMU and were assigned to haplotypes using Haplofind (Vianello *et al.*, 2013). Haplofind is based on PhyloTree (Van Oven & Kayser, 2009), the Human Phylogenetic tree. Mitochondrial gene locus and disease associations were obtained from MitoMap Database (MITOMAP, 2011). Alignment and SNP discovery capabilities were undertaken with Mummer (Kurtz *et al.*, 2004) and the sequences used to weight their tree are regularly downloaded from GenBank.

2.5.3 Mitochondrial DNA Authentication

mtDNA contamination was estimated following the protocol described by Green et al (2008) which utilizes private or near-private consensus alleles in modern day individuals (>5 % in 311 modern mtDNAs), and bases with mapping quality ≥30 and coverage of at least 10 X for the aDNA data. To obtain a contamination estimate, the counts of consensus and alternative alleles are added together across all sites.

2.5.4 Biological Sexing

The biological sex of each individual was assessed using the ratio of reads mapping to the X and Y-chromosomes which is then compared to a reference panel (Skoglund *et al.*, 2013). Analysis was restricted to sequence alignments with mapping quality of at least 30. These DNA determined sex was compared to the sex determined by morphology where possible.

2.6 Population Genetics

2.6.1 Principal Component Analysis

Statement: the population genetic analysis was kindly performed by Dr. Gulsah Merve Dal Kilinc at Stockholm University.

The Poulton individuals were merged with the SNP genotype calls of the Human Origins array (Lazaridis *et al.*, 2014 and Patterson *et al.*, 2012). A Principal Component Analysis (PCA) Plot was created at SciLife using Eigensoft to determine the genetic structure of the ten Poulton individuals. This was created using the Poulton individuals, using the merged SNP data, and the European populations from the Human Origins data set. PCA is a technique used for uncovering population structure. It is computationally efficient, handling genome-wide data for thousands of individuals and can be used to extract the fundamental structure of a dataset without the need for any modeling of the data (Paschou *et al.*, 2007).

Chapter 3: Results

3.1 Poulton samples and DNA Extractions

Table 2 summarizes the twenty-five selected Poulton skeletons. A minimum of 50 mg bone powder was successfully collected for each of the 20 tooth samples whilst ~100 mg was collected from the 5 long bone samples. DNA was extracted from all twenty-five samples and the concentration (ng/µl) was recorded (Table 2).

Table 2: Skeleton information of the twenty-five selected Poulton teeth and long bone samples

Sample ID	Skeleton #	Type	Context	Year excavated	Orientation	Coordinates	Washed (yes/no)	Concentration (ng/µl)	Osteological Sex
ASM001	794	Canine	3034	2015	E-W	105 E 10 N	Yes	5.3	Not defined
ASM002	836	Incisor	3198	2016	n/a	n/a	Yes	4.9	n/a
ASM003	797	Pre-molar	3044	2015	E-W	100 E 10 N	Yes	3	Male
ASM004	837	Canine	3202	2016	E-W	105 E 10 N	Yes	4.6	Male
ASM005	800	Canine	3061	2015	n/a	n/a	Yes	4.1	Not defined
ASM006	839	Canine	3209	2016	E-W	105 E 15 N	No	4.9	Female
ASM007	801	Canine	3063	2015	n/a	n/a	Yes	6.8	Not defined
ASM008	847	Pre-molar	n/a	2016	E-W	120 E 20 N	No	7	Female
ASM009	821	Incisor	3159	2015	n/a	n/a	Yes	6.3	Not defined
ASM010	854	Molar	3288	2016	E-W	125 E 25 N	Yes	10.9	Female
ASM011	823	Incisor	3165	2015	E-W	105 E 10 N	Yes	9.4	Not defined
ASM012	856	Pre-molar	3291	2016	E-W	105 E 10 N	No	7.6	Not defined
ASM013	829	Molar	3041	2015	n/a	n/a	Yes	5.5	Not defined
ASM014	865	Canine	3322	2016	E-W	115 E 10 N	No	4.8	Male
ASM015	834	Incisor	3192	2015	n/a	n/a	Yes	6.9	Not defined
ASM016	870	Incisor	3337	2016	E-W	105 E 10 N	Yes	8.4	Male
ASM017	873	Canine	3346	2016	E-W	105 E 10 N	No	10.8	Not defined
ASM018	797	Pre-molar	3165	2015	E-W	105 E 10 N	Yes	4.8	Not defined
ASM019	794	Pre-molar	3044	2015	E-W	105 E 10 N	Yes	4.1	Female
ASM020	865	Canine	3323	2016	E-W	115 E 10 N	No	2.6	Male
ASM021	843	Long Bone (left leg)	3221	2016	E-W	105 E 10 N	Yes	14.5	Not defined
ASM022	839	Long Bone (left leg)	3209	2016	E-W	105 E 15 N	Yes	7	Female
ASM023	825	Long Bone (left leg)	n/a	2015	n/a	n/a	Yes	17.2	Not defined
ASM024	823	Long Bone (left leg)	3165	2015	E-W	105 E 10 N	Yes	36	Not defined
ASM025	822	Long Bone (left leg)	3162	2015	n/a	n/a	Yes	10.8	Not defined

"n/a" & "not defined" due to excavation sheets non existent/not available, Context refers to the remains of an individual stratigraphic event (Darvill, 2003)

3.2 Sequencing and Authentication

Genome-wide DNA sequencing data was successfully obtained from ten Poulton individuals, with coverage greater than 0.1 X for four individuals (Table 3). Six samples have a proportion of human DNA greater than 16 %, at 75 %, 58 %, 45 %, 20 %, 17 % and 16 %. The sex was successfully determined for seven of the samples, of which four are men and three are women. One individual was not assigned, while two were found to be 'consistent with XY but not XX'. Images of the ten sequenced Poulton teeth samples are included in Appendix 1 (Figures A1-10).

Table 3: Raw sequencing data as generated on the semi-automated ancient DNA pipeline at Uppsala University.

Sample	ASM014	ASM008	ASM013	ASM012	ASM018	ASM010	ASM017	ASM007	ASM011	ASM015
Merged sequences	18132239	15837955	17365791	19420771	19946988	17162159	22803414	20552855	19473558	14309451
Human sequences	13863376	9195065	7998128	3925167	3489335	2666204	1175850	124519	111056	69386
Proportion human DNA (%)	75.7089	58.3015	45.5746	20.4736	17.1128	16.1576	5.1326	0.6141	0.5957	0.5016
Average read length (bp)	99.5515	122.123	70.8633	90.3751	84.7437	119.338	92.6263	88.0969	98.221	91.2316
Clonality	13.8255	13.5639	14.1971	15.4098	14.6998	14.8529	13.9572	12.4502	10.7808	11.5387
Too short	0.2633	0.1868	0.2034	0.3518	0.7906	0.1950	1.1230	4.1921	6.8628	5.7220
Genome coverage (%)	36.8402	26.6245	19.3207	10.0681	7.7050	9.2326	2.9580	0.3022	0.3057	0.1762
MtDNA coverage (%)	20.9009	23.7104	16.7282	12.5008	7.74513	20.4933	3.66667	4.43593	2.46877	2.60148
MtDNA sequence	4820	5347	4848	3024	2584	3803	1074	949	512	575
X sequence	391785	255697	413107	198075	95982	138121	31759	3842	3447	2181
Y sequence	104424	70289	5859	2670	27085	1441	8655	758	757	422
Biological sex	XY	XY	XX	XX	XY	XX	XY	Not assigned	Consistent with XY but not XX	Consistent with XY but not XX

mtDNA – mitochondrial DNA

All ten individuals showed characteristic features of ancient DNA (Sawyer *et al.*, 2012). The DNA was fragmented and the cytosine deamination can be seen for one sample in Figure 5 below and the other nine samples in the appendix (Figure A21), whereby the presence of T nucleotides are consistently higher at the 5' terminal ends.

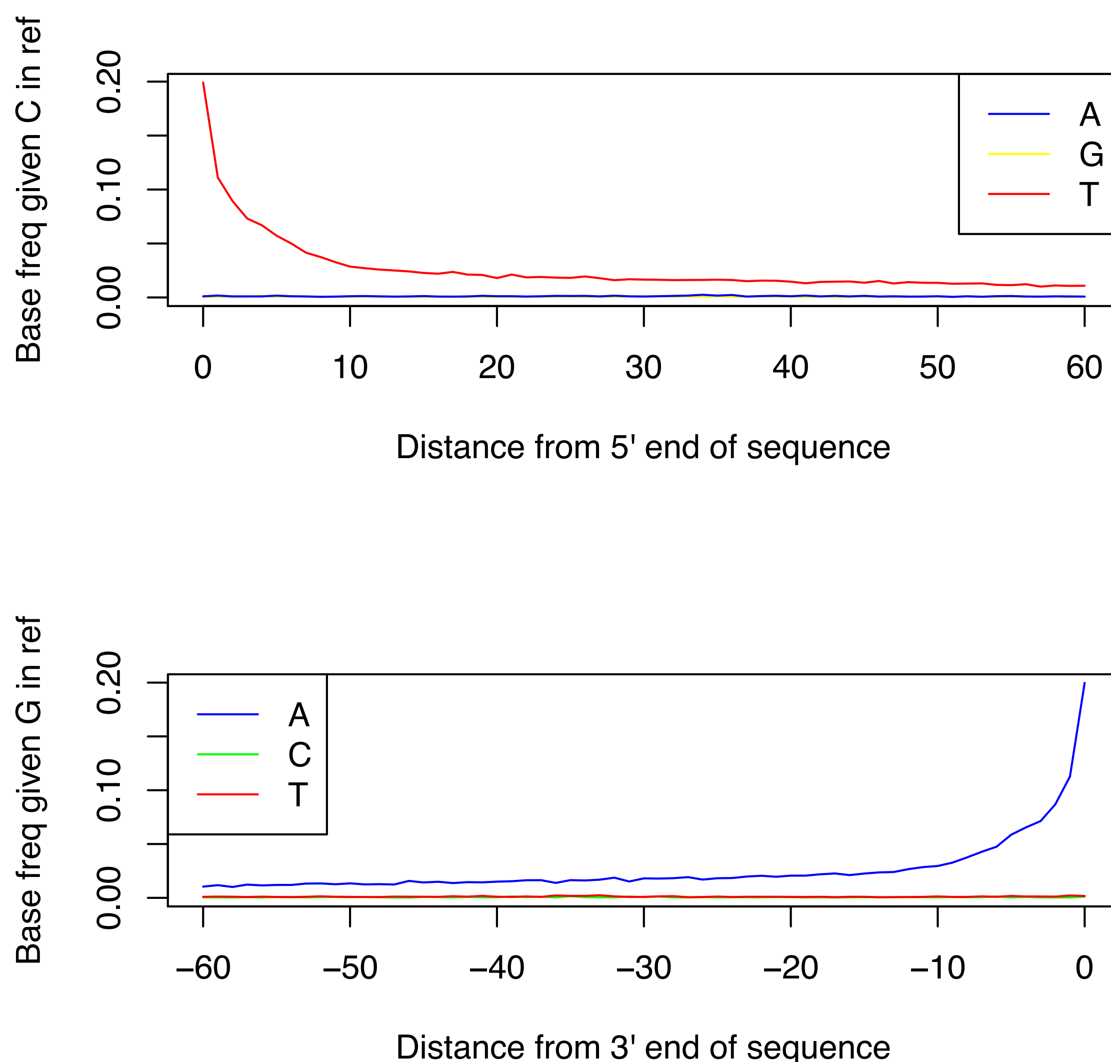


Figure 5: Mitochondrial DNA damage plots of Poulton individual ASM007

The mitochondrial DNA-based contamination estimates were 0 % for all the ancient individuals, with a 95 % confidence interval of 0 % to maximum 5.209 % (Table 4). These were calculated based on consensus alleles (range of 56 to 150) at informative sites (range of 2 to 6). The consensus alleles reflect the number of sequences that are conserved (Schneider, 2002), i.e. unique to the ancient material.

Table 4: Mitochondrial DNA-based contamination estimates for five Poulton samples

Sample	ASM010	ASM012	ASM013	ASM008	ASM014
Point estimates (%)	0	0	0	0	0
Informative Sites	3	6	3	5	2
Consensus alleles	66	85	56	150	56
All alleles	66	85	56	150	56
95 % Confidence Interval	0.0 - 4.438	0.0 - 3.463	0.0 - 5.209	0.0 - 1.977	0.0 - 5.209

3.3 Washed upon excavation versus not washed

Out of the ten sequenced samples, students had not washed four after excavation, whilst the other six had been washed. The unwashed samples were first, second, fourth and seventh in terms of ranking by highest endogenous DNA and genome coverage percentages. The 'seventh' rank had 5% endogenous human DNA, whilst the last three samples (previously washed) had endogenous human DNA percentages below 0.6 % (Table 3).

3.4 Phylogenetic Analysis

Haplogroups

The mitochondrial DNA genomes were uploaded to Haplofind (Vianello *et al*., 2013) and assigned to known haplogroups. The haplogroup and scores are summarized in Table 5 below. The SNPs and mutations generated are in Appendix 3 (Table A5: A to J). The mtDNA haplogroup tree was downloaded from Phylotree (Van Oven & Kayser, 2009), and the number of Poulton individuals, estimated to a specific haplogroup, is indicated on the tree accordingly (Figure 6).

Table 5: Results from Haplofind of mitochondrial DNA sequences; haplogroup and scores

Sample	Haplogroup	Score	Assignment Score
ASM007	I2	1	0.9
ASM008	H105a	1	0.99
ASM010	U5b2c2b	1	0.98
ASM011	HV0f	0.7	0.68
ASM012	rCRS (H2a2a)	1	0.98
ASM013	W5a1	1	0.94
ASM014	H1a1	1	0.93
ASM015	T2a1b1a1b	1	0.81
ASM017	I1a1	0.7	0.74
ASM018	T2f1a	1	0.92

rCRS – Cambridge Reference Sequence (haplogroup H2a2a)

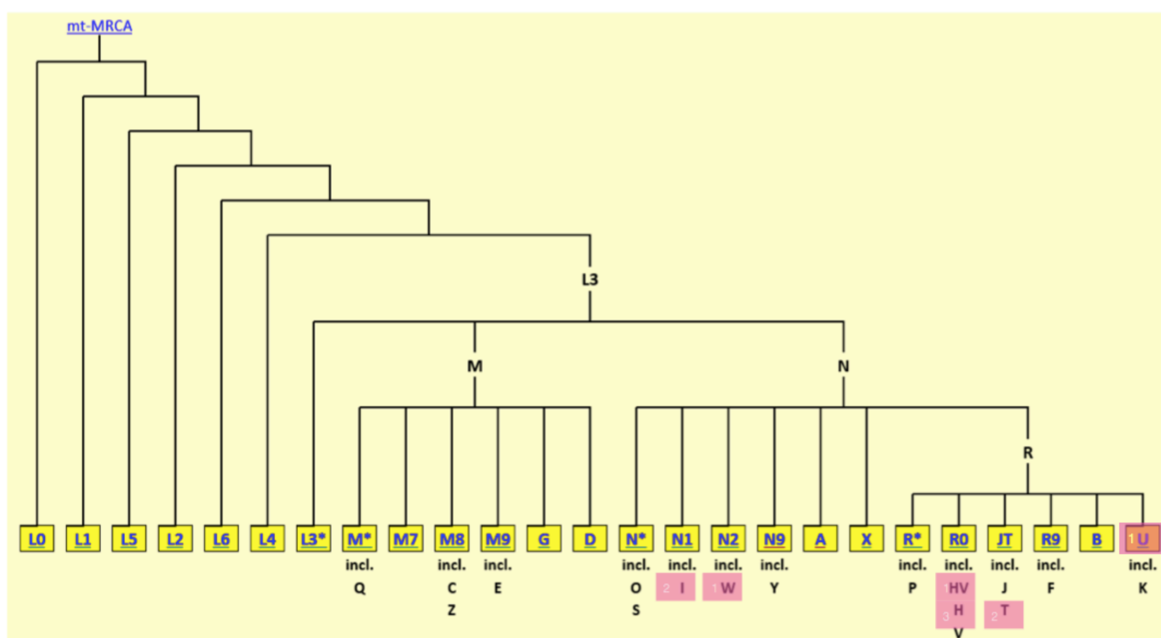


Figure 6: mtDNA haplogroup tree (Van Oven & Kayser, 2009) accompanied with the number of Poulton samples estimated at specific haplogroups. The number of Poulton samples, which fall within a haplogroup, is highlighted in pink

Nuclear DNA

Samples with a minimum number of 10000 overlapping SNPs with human origins array (Table 6) were analysed using PCA (Figure 7). The three samples not included were also the sequences having the lowest genome coverage (Table 3). These were ASM007, ASM011 and ASM015.

Table 6: Number of SNPs overlapping with human origins array for the Poulton samples

Individual	Human Origins Array SNPs
ASM007-b1e1l1p1_CTCGATG.only.map	1240
ASM008-b1e1l1p1_GCTCGAA.only.map	92374
ASM010-b1e1l1p1_CCGGTAC.only.map	34346
ASM011-b1e1l1p1_AACTCCG.only.map	1320
ASM012-b1e1l1p1_TTGAAGT.only.map	39881
ASM013-b1e1l1p1_ACTATCA.only.map	70406
ASM014-b1e1l1p1_TTGGATC.only.map	121016
ASM015-b1e1l1p1_CGACCTG.only.map	752
ASM017-b1e1l1p1_AGGTACC.only.map	11879
ASM018-b1e1l1p1_TGCGTCC.only.map	31141

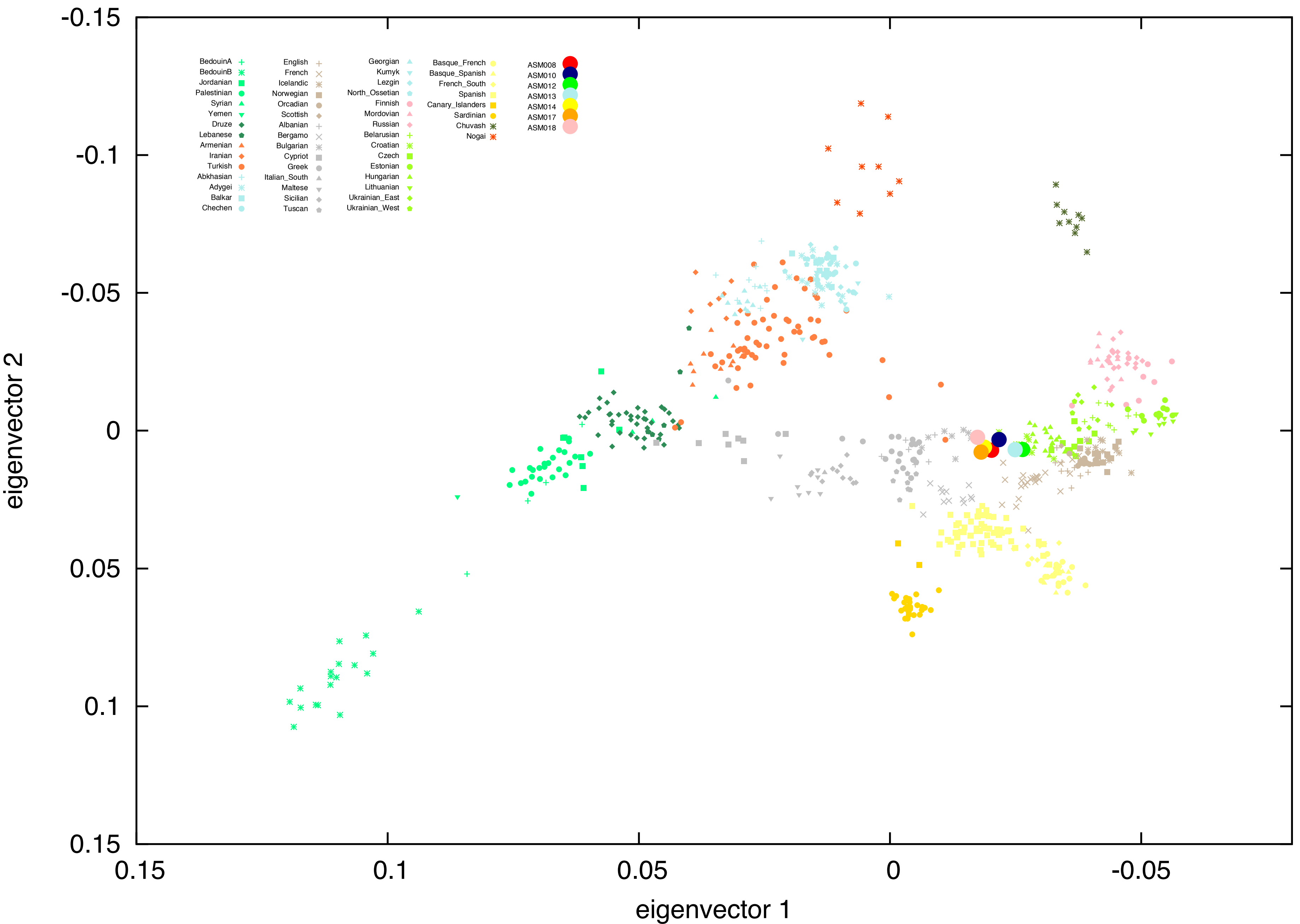


Figure 7: PCA plot of seven ancient Poulton individuals and modern-day individuals across Europe. Only ancient samples with more than 10000 overlapping SNPs with modern-day SNPs are plotted. Larger colour circles show the ancient individuals locality with regards to the modern-day European populations

3.5 Library Preparation Techniques

Table 7 below summarizes the Poulton samples that either underwent both library preparation methods and/or were sequenced. Libraries for a number of other samples were prepared by Meyer method and their amplification curves are included in Appendix 4 (Figure A20)

Table 7: Indication of which Poulton samples were prepared by which library technique and/or sequenced

Sample ID	Meyer	CATS	Note
ASM007	YES	NO	Sequenced
ASM008	YES	YES	Sequenced
ASM010	YES	YES	Sequenced
ASM011	YES	YES	Sequenced
ASM012	YES	YES	Sequenced
ASM013	YES	YES	Sequenced
ASM014	YES	YES	Sequenced
ASM015	YES	YES	Sequenced. Not included in qPCR dilutions
ASM017	YES	YES	Sequenced
ASM018	YES	YES	Sequenced
ASM019	YES	YES	Not sequenced
ASM020	YES	YES	Not sequenced
ASM023	YES	YES	Not sequenced. Not included in qPCR dilutions

3.5.1 Meyer

Analysis by qPCR of the Meyer libraries showed that eleven out of the twelve samples showed successful amplification of the DNA (Figure 8), confirming good DNA preservation as well as good library efficiency. Ten samples displaying good library efficiency were chosen for sequencing. Libraries were prepared by Meyer technique for a number of samples that were not sequenced and were successfully amplified by qPCR (Appendix 4 Figure A20). Included in these, were four long bone samples (Samples 21, 22, 23 and 24), which all amplified successfully.

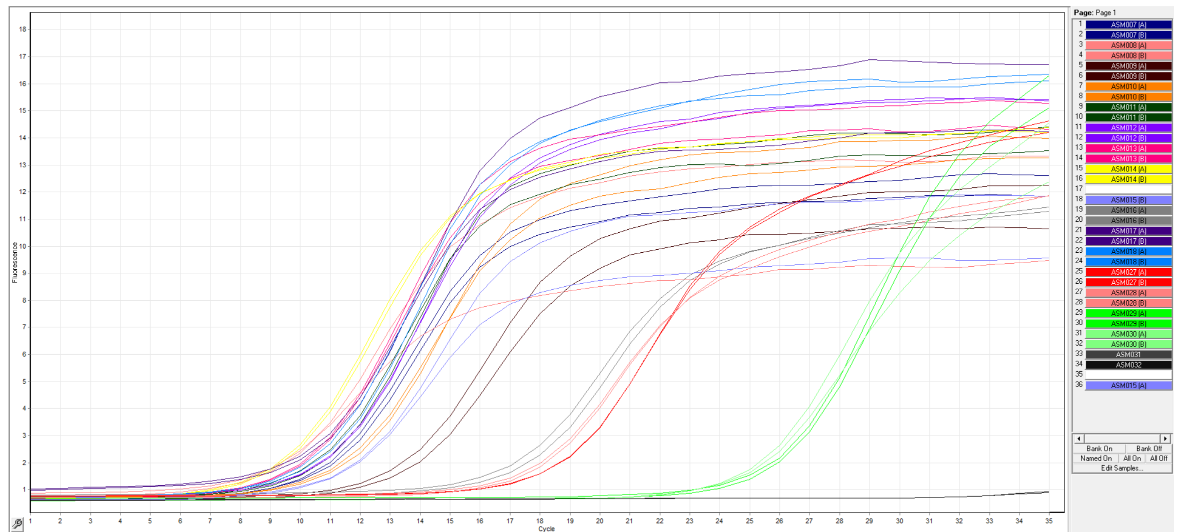


Figure 8: qPCR of Poulton cementum libraries prepared by Meyer methods. Samples ASM007-ASM018, duplicates are indicated as the same colour. ASM027 and ASM028 = DNA extraction controls, ASM029 and ASM030 = Library controls, ASM031 and ASM032 = qPCR controls

3.5.2 CATS

A total of twelve individuals were used to build CATS libraries (Table 7) with successful amplification. Eleven libraries were from cementum samples whilst one library was built from a long bone sample. The qPCR results of the amplified CATS libraries are in Appendix 4.

3.5.3 Comparison of CATS and Meyer

The Meyer libraries built using diluted samples displayed a drop in library efficiency in comparison to the libraries of non-diluted samples of the same individuals (Appendix 4 Figure A14 - 16), whereas the initial test of diluted CATS libraries displayed similar amplification curves to the un-diluted libraries (Appendix 4 Figure A13).

The relative Ct values for diluted and undiluted samples from both library methods obtained from qPCR analysis were calculated (Table 8). The Meyer libraries all have higher Ct values when diluted. The CATS libraries vary, with three samples having lower Ct values when diluted (8, 11 and 12), three samples had very similar Ct values (10, 13 and 20) and the rest with higher Ct values (14, 17, 18 and 19).

Table 8: Relative Ct values for diluted and non-diluted library samples prepared by Meyer and CATS

Sample	CATS undiluted	Meyer undiluted	CATS diluted	Meyer diluted
ASM008	18.05	5.51	14.45	9.99
ASM010	17	6.34	18.19	11.02
ASM011	14.09	8.08	11.14	9.62
ASM012	15.57	7.02	14.45	10.06
ASM013	11.21	7.96	12.3	10.73
ASM014	11.69	6.42	19.49	7.56
ASM017	5.13	6.04	10.31	8.6
ASM018	12.28	4.99	23.28	8.79
ASM019	19.18	10.41	23.34	13.48
ASM020	18.83	9.41	19.87	10.25

The raw difference in Ct values of non-diluted and diluted for Meyer and CATS libraries was calculated as well as the ratio of non-diluted to diluted Ct values (Table 9 and 10). All of the Meyer samples have ratios >1.1, with five samples >1.4. The CATS ratios were more varied with three samples <1, four between 1 and 1.3, and three were between 1.5 and 2.

Table 9: Difference and ratio of non-diluted to diluted CT values for CATS and Meyer DNA libraries

	CATS				Meyer			
Sample	**Non diluted**	**Diluted**	**Difference**	**Ratio**	**Non diluted**	**Diluted**	**Difference**	**Ratio**
ASM008	18.05	14.45	3.6	0.80	5.51	9.99	-4.48	1.81
ASM010	17	18.19	-1.19	1.07	6.34	11.02	-4.68	1.74
ASM011	14.09	11.14	2.95	0.79	8.08	9.62	-1.54	1.19
ASM012	15.57	14.45	1.12	0.93	7.02	10.06	-3.04	1.43
ASM013	11.21	12.3	-1.09	1.10	7.96	10.73	-2.77	1.35
ASM014	11.69	19.49	-7.8	1.67	6.42	7.56	-1.14	1.18
ASM017	5.13	10.31	-5.18	2.01	6.04	8.6	-2.56	1.42
ASM018	12.28	23.28	-11	1.90	4.99	8.79	-3.8	1.76
ASM019	19.18	23.34	-4.16	1.22	10.41	13.48	-3.07	1.29
ASM020	18.83	19.87	-1.04	1.06	9.41	10.25	-0.84	1.09

Chapter 4: Discussion

4.1 Assessment of Poulton DNA preservation

The principle aim of reassessing DNA preservation in Poulton material was successfully implemented using the discussed updated methods for DNA library preparation, and subsequent results gave clear answers to this question. Meyer libraries were amplified by qPCR, allowing for the successful sequencing of ten Poulton samples. The sequencing data revealed that the Poulton samples do in fact have good DNA preservation and are of high quality. This was not found in previous studies that adapted PCR techniques for DNA extraction and analysis of Poulton samples (Town, 2015). This aforementioned study concluded that no amplifiable DNA had been extracted and that it was impossible to determine from where the extracted DNA derived. Town (2015) used four different methods for DNA extraction. The first was Silica Columns whereby the bone was purified using a revised protocol and the QIAamp DNA Micro kit (Qiagen, U.K.), the second was Phenol Chloroform which is commonly used with degraded bone samples and involves separating the DNA into an aqueous phase, the third was Chelex which is a useful technique for extracting DNA from tissue but has not been widely tested on old archaeological bones (Walsh *et al*., 1991) and the fourth was Trizol which is commonly used to prepare RNA (Town, 2015). The use of these older methods for extraction, the low quality samples and use of one type of bone was the probable cause of their results in comparison to our study, which employed up-to-date methods for aDNA extraction, library preparation and sequencing.

Our study is therefore the first to successfully extract and sequence DNA from the Poulton collection. In addition to the successful library preparation that were sequenced and analysed, seven other Poulton libraries were prepared and amplified. These were stored for possible further research. The qPCR results (Appendix 4, Figure A20) for these samples are comparable to the qPCR results of the sequenced samples, and hence it can be said that good preservation was seen across majority of the selected Poulton samples, both teeth and long bones.

DNA was recovered from all of the ten sequenced samples with high proportions of human DNA (>15 %) for six samples. The average read length visualized across all the samples (96 bp) is similar to other ancient DNA studies (e.g. Günther *et al.*, 2015) all displaying good DNA preservation. This supports the authenticity of recovery of aDNA (Pääbo *et al*., 2004) and suggests that the Poulton DNA is better preserved than initially expected. Martiniano et al

(2016) produced endogenous DNA ranging from 20 % to 59 % with their British Roman period samples, similar to our Poulton medieval samples. This aforementioned study used the same dsDNA library preparation technique (Meyer & Kircher, 2010) as was used in our study. The low contamination estimates recorded for the mtDNA; less than 5.2 % for five greater than 0.09 X coverage individuals (Table 3), indicate that the methods used for cleaning the samples, extracting the DNA, preparing libraries and sequencing the samples were all performed correctly and are of high standard.

4.2 Analysis of Poulton sequences

Biological sex

The sex was successfully determined for seven of the samples, of which four are men and three are women. One individual was not assigned, while two were found to be 'consistent with XY but not XX'. The skeletons have not yet been (professionally) catalogued at LJMU, as they are the more recent samples to be excavated (2015 and 2016). Notes that were taken at the time of excavation give an indication of the sex for some samples. The notes available which overlap with our chosen samples are for: Skeleton 797 (ASM018) noted to be male, which matches our DNA sex determination, Skeleton 847 (ASM008) was noted to be a female based on sciatic notch and mandibular morphology, which does not correlate with our result, Skeleton 854 (ASM010) was noted as being female which correlates with our sex determination and Skeleton 865 (ASM014) was noted as being male which matches our sex outcome. Excavation notes were available for another three of our chosen skeletons, of which no estimated sex was noted. This possibly shows the limitation in determining sex by osteological methods, whilst highlighting the importance of DNA sequencing for sex determination.

Washing observation

Another observation was that samples excavated more recently as well as were not washed after excavation and rather washed in the laboratory before DNA extraction, produced higher endogenous DNA levels. This suggests that samples unwashed and rather cleaned, as described in our methods, are more likely to produce better sequencing data. This is in correspondence to studies that have looked at exactly this.

Tap water contains DNA and is thus a serious source for contamination and hence potential damage (Bollongino *et al.*, 2008). Washing enhances further degradation, most notably on porous bone and teeth structures, which are completely penetrated by water, contaminating

deeply into the tissue. Contaminants inside the sample will always be co extracted and not merely removed from the sample surface.

A study by Pruvost et al (2006) analysed the influence of standard post excavation treatments on the preservation of DNA within archaeological bones by comparing bones from museum collection and freshly excavated bones kept under conditions resembling those in sediment after excavation. The growth of microorganisms, which destroy preserved biomolecules by oxidation and hydrolytic process, was prevented by specific handling techniques of samples. Dissolution and degradation of endogenous DNA as well as contamination by exogenous DNA was also prevented, by avoiding standard archaeological techniques such as washing, brushing and treatment with consolidants and other chemicals. They found that 46 % of the 'fresh' fossils yielded authenticated amplification products, whereas only 18 % of old fossils yielded these products. They were able to amplify larger DNA fragments (201 bp) from 15 % of the fresh bones whilst only 4 % was amplifiable from the old bones. They concluded that the evidence they obtained gave a clear indication that DNA preservation is better when bones are freshly excavated and untreated and that the post excavation treatments and/or storage conditions negatively affect DNA preservation (Pruvost *et al.*, 2006). Further to this, knowing that taphonomic conditions strongly affect preservation of DNA, they compared bones collected under various conditions from the same preservation site (Telleilat-Mezraa, a Neolithic site in Turkey). They compared old bones that had been excavated several years earlier and had been brushed with water, dried and stored under light exclusion conditions in collections at room temperature, to fresh bones recently excavated following protocols to optimise recovery of bio molecular evidence. The difference was profound; they were able to amplify DNA from five of eight fresh fossil bones (with ~39,900 – 1600 molecules/gram of bone) of opposed to 0 of 11 old fossil bones (Pruvost *et al.*, 2006). This showed that suitable and optimal post excavation conditions were vital for DNA preservation of bones from the same preservation context. Their final and conclusive evidence was analysis of fossil material, which shared diagenetic history but experienced different post excavation histories. Both types of fossils were selected from the same skeletal elements (ribs) to minimise preservation differences. DNA was not amplified from ten samples from two ribs excavated in 1947, whereas nine samples from the 2004 excavation had a 100 % success rate. This lead to their final conclusion that ancient DNA preserved for thousands of years in fossil bones can degrade relatively quickly when removed from the preserving conditions of their original location. This degradation is due to the change of macro environment and/or standard handling and storage procedures in collections (Pruvost *et al.*, 2006).

Our results support these previous findings, emphasizing again the importance of handling specimens from the moment of excavation to storage. This is especially an important factor for Poulton material, which is handled by a number of students and often used as teaching material.

4.3 Population Genetics

PCA Plot

Current British populations are on average, 36.94 % British or Anglo-Saxon, 21.59 % Irish (Celtic) and 19.91 % Western European – the region today covered by Germany and France (Ancestry, 2015). Early migrations to Britain with the largest contribution to the present British populations are western Germany, Belgium and northwestern France (Leslie *et al.*, 2015). Ancestry profiles of the North Wales population have substantial contribution from northern Germany and northwest France (Leslie *et al.*, 2015).

Our PCA Plot illustrates the relation of seven Poulton samples to current European populations. The close positioning of all the samples to one another and the lack of outliers supports that the samples are from the same population, the sequenced DNA is from the Poulton material and not contaminating DNA, and that the DNA is well preserved. The Poulton samples are in closest proximity to samples from Bulgaria and Croatia as well as Hungary and then France. England, Norway and Czechoslovakia follow as the other nearby surrounding countries on the plot. The positioning of the samples does not appear as 'British' as might have been expected (Leslie *et al.*, 2015).

A study assessing population structure of Ireland and Britain presented a PCA Plot of their samples compared to Portugal, Bulgaria, Sweden and Utah HapMap European Americans (Utah residents of northern and western European ancestry) (O'Dushlaine *et al.*, 2010). Unlike our results, their samples were not in close proximity to the Bulgarian samples but rather samples from the Utah HapMap European Americans and Sweden.

Analysis of the nuclear genome suggests that the Poulton material shares higher genetic affinity with more Southeastern European countries than today's British population. The overlap of ancestry from France could suggest minor shared ancestry with the Poulton population and today's British population.

Mitochondrial Haplogroups

The seven samples assigned to haplogroups that are all common in present-day European populations (Kivisild, 2015), as was expected (Table 5). The lack of overlapping mtDNA haplotypes implies that contamination in the laboratory is very unlikely.

Four samples (8, 11, 12 and 14) all fall under the main haplogroup HV0. Haplogroup HV0 is evenly distributed across all Europe and North Africa with a 2-8 % frequency in nearly all countries. The Sami of northern Scandinavia and Finland are the only population with higher incidence of HV0 (42 %) and the Cantabrians with 19 % frequency. HV0 is overall slightly more common in around the Baltic, in Iberia and in the Maghreb (Barral-Arca *et al*., 2016). Sample 11 is haplogroup HV0, subclade HV0f, found in Sweden and Italy. Samples 8, 12 and 14 belong more specifically to Haplogroup H, by far the most common all over Europe, about 40 % of the European population (Achili *et al.*, 2004). It is also found in lower frequencies in North Africa, the Middle East, Central Asia, and Northern Asia and along the East coast of Africa as far down as Madagascar. Sample 8 and 14 are haplogroup H1, which peaks in Norway (30 % of the population) and Iberia (18-25 %), is high among the Sardinians, Finns and Estonians (16 %) is present in Western and Central Europe (10-12 %) and North West Africans (10-20 %) (Roostalu *et al.*, 2007). Sample 12 is the Cambridge reference sequence, which is H2a2a, found throughout Europe.

Three samples (7, 13 and 17) fall under the main haplogroup N. Samples 7 and 17 are more specifically haplogroup I. This is a fairly rare lineage, found at moderate to low frequencies in East Africa, Europe, West Asia and South Asia (Fernandes *et al.*, 2012). Sample 13 is haplogroup W having frequency peaks in India, the Near East and the Caucasus, possibly suggesting an origin in the Near East and subsequent rapid spread into Europe (Olivieri *et al.*, 2013).

Three samples (10, 15 and 18) are in the main haplogroup R. Sample 10 is Haplogroup U5b, which is present at low frequencies across Europe with 57 % in Norwegian Sami, 41 % in Finnish and 26 % in Swedish (Ingman & Gyllensten, 2006). Samples 15 and 18 are within the main haplogroup JT, and more specifically T2. Haplogroup T has low frequencies throughout Europe (10 %) and Western and Central Asia. It is common in modern day Iranians, roughly 8.3 % of the population (Kivisild *et al.*, 2004).

When assessing the haplogroups of the Poulton samples alone, a more northern European population may be deciphered, although this cannot be confirmed as it is not consistent across

all samples as well as Europe has a major overlap of haplogroups across all countries. For example, HV0f has an even distribution all across Europe and North Africa (Barral-Arca *et al*., 2016) and hence a sample with this haplogroup, like one of our Poulton samples, would be impossible to define to be more northern or southern Europe.

The majority of haplogroups found in today's populations from Croatia and Bulgaria are H, U and J (Svjetlana *et al.*, 2004 and Karachanak *et al.*, 2012). 7 out of 10 of our samples have haplogroups within these three groups. This does not necessarily define our samples as sharing most of their mitochondrial genome with these countries, but when looking at it together with the PCA result it does provide support between these two sets of data.

Despite a distinct and clear pattern of geographic spread of haplogroups in extant populations, it is not straightforward to associate specific haplogroups with prehistoric events or time periods. Phylogeographic inferences have suggested that the majority of the common European haplogroups of today derive from the Late Glacial re-colonization event (Soares *et al.*, 2010), whilst evidence from aDNA studies shows that only a subset of haplogroup U variation is likely to hold ancestry in pre-Neolithic Europe and other haplogroups likely to be related with more recent episodes of gene flow and demographic events (Soares *et al.*, 2010).

Other studies of ancient British remains have been analysed for comparison to the haplogroups of the Poulton samples. There is an overlap of haplogroups with some studies of ancient material from a similar location and era. Hassan et al (2014) studied the ancient remains of putative infanticide victims from the Yewden Roman villa site at Hambleden, England, which is dated as 1st-4th century AD. Although these remains are older than ours as well as a southern England location, a number of their samples have a common haplogroup overlap, e.g., H, T2 and U5. Töpf et al (2006) investigated the history of women migration by amplifying mtDNA from ancient Britons who lived between approximately 300-1000 AD from five sites around England. They then compared these to 3549 modern mtDNA databases from England. A large proportion of their samples have haplogroups which overlap with six of the Poulton samples; U5b (Sample 10), I (Sample 7 and Sample 17), W (Sample 13), T2 (Sample 15 and Sample 18). Although the sub-haplogroups differ, this high proportion of overlap shows that the Poulton population, although according to the PCA plot appear more southern European than British, still has a high proportion of overlapping mtDNA with other Britons living at a similar time period.

A study of Iron Age and Anglo-Saxon genomes from East England estimated that on average the contemporary East English population derives 38 % of its ancestry from Anglo-Saxon migrations (Schiffels *et al.*, 2016). They also found that Anglo-Saxon samples are closely related to modern Dutch and Danish populations, while the Iron Age samples share ancestry with multiple Northern European populations including Britain. According to our PCA Plot, the Poulton samples do not show close relation to these European populations, suggesting that they do not share ancestry with Anglo-Saxon or Iron Age samples, although this would need to be confirmed using e.g. the Rarecoal method as implemented in Schiffels et al (2016).

The Poulton population was of Cistercian order, comprising monks and nuns, their families and other associates. By the end of the twelfth century, the Cistercian order had spread throughout France and into England, Wales, Scotland, Ireland and other European countries. The movement had already begun in the eleventh century (Peters, 2016). The high level of mtDNA diversity and autosomal affinity to other parts of Europe could possibly reflect this.

The genetic contribution of the Yamnaya migration is seen by mtDNA haplogroups I, T1, U2, U4, U5a, W and subtypes of H (Haak *et al.*, 2015). Seven out of our ten sequenced Poulton samples fall within these haplogroups, indicating a link to the steppe populations. It is also thought that the steppe migrants would have mixed with eastern European agriculturists on their way to Central Europe (Haak *et al.*, 2015), possibly explaining a more Southern Eastern affinity of the Poulton samples.

4.4 Capture and Amplification by Tailing and Switching

We originally hypothesised that DNA would not be well preserved in the Poulton material due to the result of a previous study (Town, 2015) and hence decided to simultaneously explore the possibility of a new ssDNA method with the potential of retrieving more endogenous DNA. This was based on previous studies (Bennett *et al.,* 2014*)* recommending the use of ssDNA library methods for poorly preserved ancient material. We aimed to both test and optimize the CATS ssDNA library preparation technique.

CATS showed successful and promising results upon qPCR analysis with the construction of a number of libraries. After editing some of the incubation times, amplification by qPCR was more prominent with lower Ct values recorded. Most of the CATS libraries displayed lower Ct values across the samples compared to the Meyer libraries but with correct protocol editing we believe would produce good DNA libraries for the Poulton samples ready for sequencing.

The TSO primer from the original paper was edited by us so as to be compatible with the Meyer libraries for ease of potential sequencing. The sequence change of this primer was very likely to have decreased library efficiency. TSO is used in the vital part of the CATS method of template switching. Without adequate template switching, the library will not be built optimally. The sequence was edited to be compatible with the Meyer P5 sequence (or the TruSeq Universal Adapter). This sequence has a run of three cytosines, which is thought to have hybridized to the G-overhang, blocking optimal binding of TSO and hence decreasing the template switching activity.

Effect of dilution on CATS libraries

When amplifying diluted CATS DNA libraries, the results obtained were very similar to when amplifying CATS libraries built with original concentrations. Previous studies have shown the potential of using the ssDNA method when dsDNA methods have low endogenous DNA (<3 % (Gansuage and Meyer, 2013) and <0.5 % (Bennett *et al.*, 2014)). The average proportion of human DNA across our ten sequenced samples was 24 % with only three of our Meyer libraries (ASM007, ASM011, ASM015) having less than 3 % DNA. Out of these three samples only one (ASM011) was used to build a library with the CATS method. This was the second best performing CATS library from qPCR analysis with a similar amplification curve as the Meyer library of the same individual. Library performance was better when diluted (Ct 11.14) versus non-diluted (Ct 14.09), supporting findings of the previous mentioned studies, however this still did not show better library efficiency than the Meyer method (Ct 9.62).

4.5 Cementum versus Long Bone

Four long bone library samples were created and amplified successfully, with lower Ct values than two cementum samples in the same qPCR (Appendix 4 Figure A20). Although this suggests long bone may provide higher copy number DNA extracts (Lacey, 2012), the two cementum samples (Sample 19 and 20) used for comparison were not the top performing Meyer libraries (Appendix 4 Figure A14), but rather the bottom. Comparing these two different qPCRs (Appendix 4 Figures A14 and A20) suggests that the long bone libraries amplified in a similar fashion to the top performing cementum libraries. Evidence for preference of cementum or long bone as preferential ancient DNA material therefore cannot be confirmed, however it can be stated that the cementum samples used did provide sequences of high quality, confirming good preservation of Poulton material and allowing for a first insight to the genetic affinity of this population.

4.6 Conclusions

The overarching aim of reassessing DNA preservation in human samples excavated at Poulton was achieved by addressing each of the objectives set out; macroscopically well-preserved teeth were identified by assessing all of the skeletons excavated in 2015 and 2016, selecting twenty-five samples for analysis, DNA was extracted from cementum of twenty teeth samples and five long bones, using optimized protocols for aDNA analysis, Illumina sequencing libraries were constructed using up-to-date and standard dsDNA library preparation methods, ten libraries were sequenced with the targeted >5M sequence reads/library/samples with the lowest amount of sequence reads at 14M. DNA preservation allowed for an initial insight into the genetic affinity of the Poulton population using mtDNA and nuclear DNA, revealing a more Southeastern European affinity than modern Britains. Targeting of cementum and using NGS techniques allowed for this study to be the first to extract and sequence DNA from the Poulton material. The secondary aim of the study was also achieved by initial assessment and optimization of the ssDNA library preparation method (CATS) for Poulton samples, successfully amplifying thirteen individuals.

Our study has several implications. Firstly, we were able to demonstrate that the Poulton material is not poorly preserved as originally supposed and that it is possible to sequence good coverage DNA from this material allowing for population analysis. We highlighted the importance of following proper post excavation handling of ancient material. The noticeable implication of CATS was that library efficiency improved or remained roughly the same when samples were diluted 10 fold, in comparison to Meyer libraries which all exhibited lower library efficiency when the sample was diluted. This finding sheds light on CATS being an optimal method for samples of lower concentration, i.e. samples of lower quality or badly preserved, backing up previous studies relaying similar results. With optimized CATS method, based on these results and previous studies, it is likely that samples with lower endogenous DNA could outperform the Meyer libraries. Having said this, it is possible that CATS is simply better at recovering more, small, DNA but not necessarily DNA that maps to the human genome. This would only be determined with sequencing of the CATS libraries.

4.7 Future Work

Optimization of the ssDNA CATS library preparation technique should be done to adequately compare the two library methods and infer which is a better fit for the Poulton material. This should be done by using the TSO primer sequence in the original paper and not our edited

sequence, when amplifying CATS libraries ensure optimal methods presented in original paper are used and not edited as we did for Meyer compatibility, using the optimal incubation times we determined as well as to test for better incubation times as well as concentrations of reagents and primers as these are very sensitive and slight shifts may produce noticeable increase in library efficiency.

The successful library building and amplification of nearly twenty samples, and sequencing of ten chosen samples confirming good DNA preservation, indicates the ability as well as importance of future studies of this material such as; direct radiocarbon dating for the Poulton individuals to confirm medieval age, obtaining more sequencing data and coupling this with extraction of phenotypic data (e.g. HIrsiPlex for eye and hair colour (Walsh *et al.*, 2013)) to reveal more genetic knowledge of the Poulton population, definitive sample selection and sequencing to answer questions surrounding burial, family, population and sex structure etc., and include more high medieval/late medieval samples from around the UK to resolve the population genetic mark obtained on the PCA data. Our results also indicate the importance of aDNA advancement, hence future studies of similar aims should employ up to date techniques for optimal results.

Reference List:

Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., Cruciani, Fulvio Zeviani, M., Briem, E., Carelli, V., Moral, P.,Dugoujon, J-M., Roostalu, U., Loogväli, E-L., Kivisild, T., Bandelt, H-J., Richards, M., Villems, R., Santachiara-Benerecetti, A.S., Semino, O., Torroni, A., 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

Adkins, L., & Adkins, R. A., 1982, The Handbook of British Archaeology, *Constable*, London

Adler, C. J., Haak, W., Donlon, D., & Cooper, A., 2011, Survival and recovery of DNA from ancient teeth and bones, *Journal of Archaeological Science*, *38*(5): 956–964

Aird, D., Ross, M. G., Chen, W., Danielsson, M., Fennell, T., Russ, C., Jaffe, D. B., Nusbaum, C., Gnirke, A., 2011, Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, 12: 1–14

Allentoft, M. E., Collins, M., Harker, D., Haile, J., Oskam, C. L., Hale, M. L., Campos, P. F., Samaniego, J. A., Gilbert, M. T. P., Willerslev, E., Zhang, G., Scofield, R. P., Holdaway, R. N., Bunce, M., 2012, The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings of the Royal Society B: Biological Sciences*

Ancestry,.com, 2015, DNA of the nation revealed...and we're not as 'British' as we think [online], Accessed: 10/10/2017, Available at: https://www.ancestry.com/corporate/international/press-releases/DNA-of-the-nation-revealedand-were-not-as-British-as-we-think

Barral-Arca, R., Pischedda, S., Gómez-Carballa, A., Pastoriza, A., Mosquera-Miguel, A., López-Soto, M., Martinón-Torres, F., Álvarez-Iglesias, V., Salas, A., 2016, Meta-Analysis of Mitochondrial DNA Variation in the Iberian Peninsula, *PLOS ONE*, *11*(7): e0159735

Bastiaens, M., Huurne, J. T., Gruis, N., Bergman, W., Westendorp, R., Vermeer, B. J., Bavinck, J. N. B., 2001, The melanocortin-1-receptor gene is the major freckle gene, *Human Molecular Genetics*, 10: 1701-1708

Bennett, E, A., Massilani, D., Lizzo, G., Daligaut, J., Geigl, E, M., Grange, T., 2014, Library construction for ancient genomics: single stranded or double strand?, *Biotechniques*, 56: 289-298

Bollongino, R., Tresset, A., & Vigne, J. D., 2008, Environment and excavation: Pre-lab impacts on ancient DNA analyses, *Comptes Rendus Palevol*, *7*(2): 91–98

Brandt, G., Haak, W., Adler, C. J., Roth, C., Szécsényi-Nagy, A., Karimnia, S., Möller-Rieker, S., Meller, H., Ganslmeier, R., Friederich, S., Dresely, V., Nicklisch, N., Pickrell, J. K., Sirocko, F., Reich, D., Cooper, A., Karimnia, S., 2013, Ancient DNA Reveals Key Stages in the Formation of Central European Mitochondrial Genetic Diversity, *Science*, *342*(6155): 257-261

Briggs, A. W., Good, J. M., Green, R. E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, Carles Rudan, P., Brajković, D., Kućan, Ž., Gušić, I., Schmitz, R., Doronichev, V. B., Golovanova, L.V., de la Rasilla, M., Fortea, J.,Rosas, A., Pääbo, S., 2009, Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes. *Science*, 325: 318-321

Briggs, A. W., Stenzel, U., Johnson, P. L. F., Green, R. E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M., Pääbo, S., 2007, Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences*, 104: 14616–14621

Brotherton, P., Endicott, P., Sanchez, J. J., Beaumont, M., Barnett, R., Austin, J., & Cooper, A., 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: 5717–5728

Burrell, C. L., Carpenter, R. J., 2014, Analysis of Human Skeletal Material from the Poulton Research Project: 1995–2013, *Poulton Archaeology Press*, Williamsburg, Virginia, Available at: https://www.researchgate.net/publication/313060935_Analysis_of_Human_Skeletal_Material_from_the_Poulton_Research_Project_1995-2013, Accessed 10/05/2016

Capelli, C., Redhead, N., Abernethy, J. K., Gratrix, F., Wilson, J. F., Moen, T., Hervig, T., Richards, M., Stumpf, M. P. H., Underhill, P. A.,Bradshaw, P., Shaha, A., Thomas, M. G.,

Bradman, N., Goldstein, D. B., 2003, A Y Chromosome Census of the British Isles, *Current Biology*, *13*(11): 979–984

Choi, M., Scholl, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P., Nayir, A., Bakkaloğlu, A., Özen, S., Sanjad, S., Nelson-Williams, C., Farhi, A., Mane, S., Lifton, R. P., 2009, Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences*, 106: 19096–19101

Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, Ward R., 2001, Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature*, 409: 704–707

Cooper, A., Mourer-Chauviré, C., Chambers, G. K., von Haeseler, A., Wilson, A. C., & Pääbo, S., 1992, Independent origins of New Zealand moas and kiwis, *Proceedings of the National Academy of Sciences* , *89*(18): 8741–8744

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J. L., Meyer, M., 2013, Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, 110: 15758–15763

Damgaard, P, B., Margaryan, A., Schroeder, H., Orlando, L., Willerslev, E., & Allentoft, M, E., 2015, Improving access to endogenous DNA in ancient bones and teeth, *Scientific Reports,* 5: 11184

Darvill, T., 2003, The concise Oxford dictionary of archaeology, *Oxford University Press*, Oxford, 98

Driscoll, C. A., Yamaguchi, N., Bar-Gal, G. K., Roca, A. L., Luo, S., Macdonald, D. W., & O'Brien, S. J., 2009, Mitochondrial Phylogeography Illuminates the Origin of the Extinct Caspian Tiger and Its Relationship to the Amur Tiger, *PLOS ONE*, *4*(1): e4125

Eglinton G, Logan GA., 1991, Molecular preservation, *Philosophical Transactions*, later *Philosophical Transactions of the Royal Society*, London Series B, 333: 315–27; discussion 27–28

Emery, M., 2000, The Poulton Chronicles: Tales from a Medieval Chapel, *Poulton Archaeology Press*, Willamsburg, Virginia, 1-59

Endicott, P., Gilbert, M. T. P., Stringer, C., Lalueza-Fox, C., Willerslev, E., Hansen, A. J., & Cooper, A., 2003, The Genetic Origins of the Andaman Islanders. *The American Journal of Human Genetics*, 72: 178–184

Friedberg, E. C., 2003, DNA damage and repair, *Nature*, 421: 436–440

Fernandes, V., Alshamali, F., Alves, M., Costa, M. D., Pereira, J. B., Silva, N. M., Cherni, L., Harich, N., Cerny, V., Soares, P., Richards, M. B., Pereira, L., 2012, The Arabian Cradle: Mitochondrial Relicts of the First Steps along the Southern Route out of Africa, *American Journal of Human Genetics*, *90*(2): 347–355

Gamba, C., Jones, E. R., Teasdale, M. D., McLaughlin, R. L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., Anders, A., Whittle, A., Dani, J., Raczky, P., Higham, T. F. G., Hofreiter, M., Bradley, D. G., Pinhasi, R., 2014, Genome flux and stasis in a five millennium transect of European prehistory, *Nature Communications, 5*: 5257

Gansauge, M. T., Meyer, M., 2013, Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA, *Nature Protocols*, 8: 737-748

Gigli, E., Rasmussen, M., Civit, S., Rosas, A., de la Rasilla, M., Fortea, J., Gilbert, M. T. P., Willerslev, E., Lalueza-Fox, C., 2009, An improved PCR method for endogenous DNA retrieval in contaminated Neandertal samples based on the use of blocking primers. *Journal of Archaeological Science*, 36: 2676–2679

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

Gilbert, M. T. P., Rudbeck, L., Willerslev, E., Hansen, A. J., Smith, C., Penkman, K. E. H., Prangenberg, K., Nielsen-Marsh, C. M., Jans, M. E., Arthur, P., Lynneruo, N., Turner-Walker, G., Biddle, M., Collins, M. J., 2005, Biochemical and physical correlates of DNA contamination in archaeological human bones and teeth excavated at Matera, Italy. *Journal of Archaeological Science*, 32: 785–793

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

Golenberg, E. M., Giannasi, D. E., Clegg, M. T., Smiley, C. J., Durbin, M., Henderson, D., Zurawski, G., 1990, Chloroplast DNA sequence from a Miocene Magnolia species. *Nature*, 344: 656–658

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H., Hansen, N. F., Durand, E. Y., Malaspinas, A. S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., Good, J. M., Schultz, R., Aximu-Petri, A., Butthof, A., Höber, B., Höffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E. S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Doronichev, V. B., Golovanova, L. V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R. W., Johnson, P. L., Eichler, E. E., Falush, D., Birney, E., Mullikin, J. C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., Pääbo, S., 2010, A Draft Sequence of the Neandertal Genome. *Science*, 328: 710–722

Green, R. E., Malaspinas, A.-S., Krause, J., Briggs, A. W., Johnson, P. L. F., Uhler, C., Meyer, M., Good, J. M., Maricic, T., Stenzel, U., Prüfer, K., Siebauer, M., Burbano, H. A., Ronan, M., Rothberg, J. M., Egholm, M., Rudan, P., Brajković, D., Kućan, Ž., Gušić, I., Wikström, M., Laakkonen, L., Kelso, J., Slatkin, M., Pääbo, S., 2008, A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing, *Cell*, *134*(3): 416–426

Greenwood, A. D., Castresana, J., Feldmaier-Fuchs, G., & Pääbo, S., 2001, A Molecular Phylogeny of Two Extinct Sloths, *Molecular Phylogenetics and Evolution*, *18*(1): 94–103

Günther, T., Valdiosera, C., Malmström, H., Ureña, I., Rodriguez-Varela, R., Sverrisdóttir, Ó. O., Daskalaki, E. A., Skoglund, P., Naidoo, T., Svensson, E. M., Bermúdez de Castro, J. M., Carbonell, E., Dunn, M., Storå, J., Iriarte, E., Arsuaga, J. L., Carretero, J-M., Götherström, A., Jakobsson, M., 2015, Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques, *Proceedings of the National Academy of Sciences*, *112*(38): 11917–11922

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

Hagelberg, E., Hofreiter, M., Keyser, C., 2014, Ancient DNA: the first three decades, *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1660)

Hänni, C., Laudet, V., Stehelin, D., & Taberlet, P., 1994, Tracking the origins of the cave bear (Ursus spelaeus) by mitochondrial DNA sequencing, *Proceedings of the National Academy of Sciences* , *91*(25): 12336–12340

Harvig, L., Frei, K. M., Price, T. D., & Lynnerup, N., 2014, Strontium isotope signals in cremated petrous portions as indicator for childhood origin. *PloS One*, *9*(7), e101603

Hauswirth, W. W., Dickel, C. D., Rowold, D. J., & Hauswirth, M. A., 1994, Inter- and intrapopulation studies of ancient humans, *Experientia*, 50: 585–591

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A., & Wilson, A. C., 1984, DNA sequences from the quagga, an extinct member of the horse family. *Nature*, 312: 282–284
Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M., & Paabo, S., 2001, Ancient DNA. *Nature Reviews Genetics*, 2: 353–359

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

Homøe, P., Lynnerup, N., & Videbæk, H., 1992, CT-scanning of Ancient Greenlandic Inuit Temporal Bones. *Acta Oto-Laryngologica*, 112: 674–679

Ingman, M., & Gyllensten, U., 2006, A recent genetic link between Sami and the Volga-Ural region of Russia, *European Journal of Human Genetics*, *15*(1): 115–120

Jørkov, M. L. S., Heinemeier, J., & Lynnerup, N., 2009, The petrous bone--a new sampling site for identifying early dietary patterns in stable isotopic studies. *American Journal of Physical Anthropology*, 138: 199–209

Kang, D., & Hamasaki, N., 2005, Mitochondrial DNA in somatic cells: A promising target of routine clinical tests, *Clinical Biochemistry*, *38*(8): 685–695

Karachanak, S., Carossa, V., Nesheva, D. et al., 2012, Bulgarians vs the other European populations: a mitochondrial DNA perspective, *International Journal of Legal Medicine*, 126 (4): 497-503

Kircher, M., 2012, Analysis of High-Throughput Ancient DNA Sequencing Data BT - Ancient DNA: Methods and Protocols, In B. Shapiro & M. Hofreiter, pp. 197–228, Totowa, NJ: Humana Press

Kivisild. T., 2015, Maternal ancestry and population history from whole mitochondrial genomes, *Investigative Genetics*, 6: 3

Kivisild, T., Reidla, M., Metspalu, E., Rosa, A., Brehm, A., Pennarun, E., Parik, J., Geberhiwot, T., Usanga, E., Villems, R., 2004, Ethiopian Mitochondrial DNA Heritage: Tracking Gene Flow Across and Around the Gate of Tears, *American Journal of Human Genetics*, *75*(5): 752–770

Kolman, C. J., & Tuross, N., 2000, Ancient DNA analysis of human populations. *American Journal of Physical Anthropology*, 111: 5–23

Krause, J., Dear, P. H., Pollack, J. L., Slatkin, M., Spriggs, H., Barnes, I., Lister, A. M., Ebersberger, I., Pääbo, S., Hofreiter, M., 2006, Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae, *Nature*, 439: 724–727

Krause, J., Briggs, A. W., Kircher, M., Maricic, T., Zwyns, N., Derevianko, A., & Pääbo, S., 2010, A Complete mtDNA Genome of an Early Modern Human from Kostenki, Russia, *Current Biology*, 20: 231–236

Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R. E., Burbano, H. A., Hublin, J., Hänni, C., Fortea, J., de la Rasilla, M., Bertranpetit, J., Rosas, A., Pääbo, S., 2016, The Derived *FOXP2* Variant of Modern Humans Was Shared with Neandertals. *Current Biology*, 17: 1908–1912

Krings, M., Stone, A., Schmitz, R. W., Krainitzki, H., Stoneking, M., Pääbo, S., 1997, Neandertal DNA Sequences and the Origin of Modern Humans, *Cell*, 90: 19–30

Krzewińska, M., Bjørnstad, G., Skoglund, P., Olason, P. I., Bill, J., Götherström, A., & Hagelberg, E., 2014, Mitochondrial DNA variation in the Viking age population of Norway, *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1660)

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L., 2004, Versatile and open software for comparing large genomes, *Genome Biology*, 5: R12

Lacey, L. A., 2012, Manual of Techniques in Invertebrate Pathology, *Elsevier Ltd*, London, Great Britain, 224

Lalueza-Fox, C., Römpler, H., Caramelli, D., Stäubert, C., Catalano, G., Hughes, D., Rohland, N., Pilli, E., Longo, L., Condemi, S., de la Rasilla, M., Fortea, J., Rosas, A., Stoneking, M., Schöneberg, T., Bertranpetit, J., Hofreiter, M., 2007, A Melanocortin 1 Receptor Allele Suggests Varying Pigmentation Among Neanderthals. *Science*, 318: 1453 –1455

Lambert DM, Ritchie PA, Millar CD, Holland B, Drummond AJ, Baroni C., 2002, Rates of evolution in ancient DNA from Adelie penguins, *Science*, 295: 2270–2273

Lari, M., Rizzi, E., Milani, L., Corti, G., Balsamo, C., Vai, S., Catalano, G., Pilli, E., Longo, L., Condemi, S., Giunti, P., Hänni, C., De Bellis, G., Orlando, L., Barbujani, G., Caramelli, D, 2010, The *Microcephalin* Ancestral Allele in a Neanderthal Individual. *PLoS ONE*, 5: e10648

Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Krause, J., 2014, Ancient human genomes suggest three ancestral populations for present-day Europeans, *Nature*, *513*(7518), 409–413

Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T., Hutnik, K., Royrvik, E. C., Cunliffe, B., Lawson, D. J., Falush, D., Freeman, C., Pirinen, M., Myers, S., Robinson, M., Donnelly, P., Bodmer, W., 2015, The fine-scale genetic structure of the British population, *Nature*, *519*(7543): 309–314

Li, H., & Durbin, R., 2009, Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics*, *25*

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing., 2009, The Sequence Alignment/Map format and SAMtools, *Bioinformatics*, *25*(16), 2078–2079

Lindahl T., 1993, Instability and decay of the primary structure of DNA, *Nature*, 362:709–15

LJMU Blog, 2013, Uncovering history, [online], Available at: http://blog.ljmu.ac.uk/2013/uncovering-history/ Accessed: 01/05/2016

Llorente, M. G., Jones, E. R., Eriksson, A., Siska, V., Arthur, K. W., Arthur, J. W., Curtis, M. C., Stock, J. T., Coltorti, M., Pieruccini, P., Stretton, S., Brock, F., Higham, T., Park, Y., Hofreiter, M., Bradley, D. G., Bhak, J., Pinhasi, R., Manica, A., 2015, Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa, *Science*, 350: 820-822

Malmström, H., Linderholm, A., Skoglund, P., Storå, J., Sjödin, P., Gilbert, M. T. P., Götherström, A., 2014, Ancient mitochondrial DNA from the northern fringe of the Neolithic farming expansion in Europe sheds light on the dispersion process, *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1660)

Martiniano, R., Caffell, A., Holst, M., Hunter-Mann, K., Montgomery, J., Müldner, G., McLaughlin, R. L., Teasdale, M. D., van Rheenen, W., Veldink, J. H., van den Berg, L. H., Hardiman, O., Carroll, M., Roskams, S., Oxley, J., Morgan, C., Thomas, M. G., Barnes, I., McDonnell, C., Collins, M. J., Bradley, D. G., 2016, Genomic signals of migration and continuity in Britain before the Anglo-Saxons, *Nature Communications*, 7: 10326

Meyer, M., Briggs, A. W., Maricic, T., Höber, B., Höffner, B., Krause, J., Weihmann, A., Pääbo, S., Hofreiter, M., 2008, From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing, *Nucleic Acids Research*, 36: e5–e5

Meyer, M., Kircher, M., 2010, Illumina sequencing library preparation for highly multiplexed target capture and sequencing, *Cold Spring Harbor Protocols* 2010: 1-7

Millar, C. D., Huynen, L., Subramanian, S., Mohandesan, E., & Lambert, D. M., 2008, New developments in ancient genomics. *Trends in Ecology & Evolution*, 23: 386–393

Miller, W., Drautz, D. I., Ratan, A., Pusey, B., Qi, J., Lesk, A. M., Tomsho, L. P., Packard, M. D., Zhao, F., Sher, A., Tikhonov, A., Raney, B., Patterson, N., Lindblad-Toh, K., Lander, E. S., Knight, J. R., Irzyk, G. P., Fredrikson, K. M., Harkins, T. T., Sheridan, S., Pringle, T., Schuster, S. C., 2008, Sequencing the nuclear genome of the extinct woolly mammoth, *Nature*, 456: 387–390

MITOMAP: A Human Mitochondrial Genome Database, 2011, Available at: http://www.mitomap.org, Accessed: 05/07/17

Molnar, S., 1971, Human tooth wear, tooth function and cultural variability, *American Journal of Physical Anthropology*, 34: 175–189

Norén, A., Lynnerup, N., Czarnetzki, A., & Graw, M., 2005, Lateral angle: a method for sexing using the petrous bone, *American Journal of Physical Anthropology*, 128: 318–23

O'Dushlaine, C. T., Morris, D., Moskvina, V., Kirov, G., Consortium, I. S., Gill, M., Corvin, A., Wilson, J. F., Cavalleri, G. L., 2010, Population structure and genome-wide patterns of variation in Ireland and Britain, *European Journal of Human Genetics*, *18*(11): 1248–1254

Olivieri, A., Pala, M., Gandini, F., Kashani, B. H., Perego, U. A., Woodward, S. R., Grugni, V., Battaglia, V., Semino, O., Achilli, A., Richards, M. B., Torroni, A., 2013, Mitogenomes from Two Uncommon Haplogroups Mark Late Glacial/Postglacial Expansions from the Near East and Neolithic Dispersals within Europe, *PLoS ONE*, *8*(7): e70492

Orlando, L., Calvignac, S., Schnebelen, C., Douady, C. J., Godfrey, L. R., & Hänni, C., 2008, DNA from extinct giant lemurs links archaeolemurids to extant indriids, *BMC Evolutionary Biology*, *8*(1): 121

Orlando, L., Darlu, P., Toussaint, M., Bonjean, D., Otte, M., & Hänni, C., 2006, Revisiting Neandertal diversity with a 100,000 year old mtDNA sequence, *Current Biology*, 16: 400–402

Pääbo, S., 1985, Molecular cloning of Ancient Egyptian mummy DNA, *Nature*, 314: 644–645

Pääbo, S., Higuchi, R. G., Wilson, A. C.,1989, Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. *Journal of Biological Chemistry*, 264: 9709–9712

Pääbo S., 1989, Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Science of the United States of America*, 86: 1939–1943

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Després, V., Hebler, J., Rohland, N. Kuch, M., Krause, J., Vigilant, L., Hofreiter, M., 2004, Genetic Analyses from Ancient DNA, *Annual Review of Genetics*, 38: 645–679

Parks, M., Subramanian, S., Baroni, C., Salvatore, M. C., Zhang, G., Millar, C. D., Lambert, D. M., 2014, Ancient population genomics and the study of evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370

Paschou, P., Ziv, E., Burchard, E. G., Choudhry, S., Rodriguez-Cintron, W., Mahoney, M. W., & Drineas, P., 2007, PCA-Correlated SNPs for Structure Identification in Worldwide Human Populations, *PLOS Genetics*, *3*(9):e160

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., Reich, D., 2012, Ancient Admixture in Human History, *Genetics*, *192*(3): 1065–1093

Paxinos, E. E., James, H. F., Olson, S. L., Sorenson, M. D., Jackson, J., & Fleischer, R. C., 2002, mtDNA from fossils reveals a radiation of Hawaiian geese recently derived from the Canada goose (Branta canadensis), *Proceedings of the National Academy of Sciences*, *99*(3): 1399–1404

Peters, C., 2016, Cistercian expansion and industrial water mill infrastructure in twelfth-century France: An unrecognized dependency, Master of Arts, University of Houston, USA

Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpaslan-Roodenberg, S., Gerritsen, F., Moiseyev, V., Gromov, A., Raczky, P., Anders, A., Pietrusewsky, M., Rollefson, G., Jovanovic, M., Trinhhoang, H., Bar-Oz, G., Oxenham, M., Matsumura, H., Hofreiter, M., 2015, Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone, *PLOS ONE*, *10*(6): e0129102

Poinar, H. N., Schwarz, C., Qi, J., Shapiro, B., MacPhee, R. D. E., Buigues, B., Tikhonov, A., Huson, D. H., Tomsho, L. P., Auch, A., Rampp, M., Miller, W., Schuster, S. C., 2006, Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA. *Science*, 311: 392 -394

Poulton Research Project, 2014, History of Poulton, [online], Available at: http://www.poultonresearchproject.co.uk/history-of-poulton/ Accessed: 01/05/2016

Prufer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P. H., de Filippo, C., Li, H., Mallick, S., Dannemann, M., Fu, Q., Kircher, M., Kuhlwilm, M.,Lachmann, M.,Meyer, M., Ongyerth, M., Siebauer, M., Theunert, C., Tandon, A., Moorjani, P., Pickrell, J., Mullikin, J. C., Vohr, S. H., Green, R. E.,Hellmann, InesJohnson, P. L. F., Blanche, H., Cann, H., Kitzman, J. O., Shendure, J., Eichler, E. E., Lein, E. S., Bakken, T. E., Golovanova, L. V., Doronichev, V. B., Shunkov, M. V., Derevianko, A. P., Viola, B., Slatkin, M., Reich, D., Kelso, J., Paabo, S., 2014, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, 505: 43–49

Pruvost, M., Schwarz, R., Correia, V. B., Champlot, S., Braguier, S., Morel, N., Fernandez-Jalvo, Y., Grange, T., Geigl, E. M., 2007, Freshly excavated fossil bones are best for amplification of ancient DNA, *Proceedings of the National Academy of Sciences*, 104: 739–744

Ramírez, O., Gigli, E., Bover, P., Alcover, J. A., Bertranpetit, J., Castresana, J., & Lalueza-Fox, C., 2009, Paleogenomics in a Temperate Environment: Shotgun Sequencing from an Extinct Mediterranean Caprine, *PLOS ONE*, *4*(5): e5670

Rizzi, E., Lari, M., Gigli, E., De Bellis, G., Caramelli, D., 2012, Ancient DNA studies: new perspectives on old samples. *Genetics Selection Evolution*, 44: 1–19

Roostalu, U., Kutuev, I., Loogväli, E.-L., Metspalu, E., Tambets, K., Reidla, M., Khusnutdinova, E. K., Usanga, E., Kivisild, T., Villems, R., 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

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., & Pääbo, S., 2012, Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. *PLOS ONE*, *3*: e34131

Schiffels, S., Haak, W., Paajanen, P., Llamas, B., Popescu, E., Loe, L., Clarke, R., Lyons, A., Mortimer, R., Sayer, D., Tyler-Smith, C., Cooper, A., Durbin, R., 2016, Iron Age and Anglo-Saxon genomes from East England reveal British migration history, 7: 10408.

Schneider, T. D., 2002, Consensus Sequence Zen, *Applied Bioinformatics*, *1*(3), 111–119

Schwarz, C., Debruyne, R., Kuch, M., McNally, E., Schwarcz, H., Aubrey, A. D., Bada, J., Poinar, H., 2009, New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains, *Nucleic Acids Research*, *37*(10), 3215–3229

Scott, G. R., Turner, C. G., 2000, The anthropology of modern human teeth: dental morphology and its variation in recent human populations: Cambridge University Press

Shendure, J., Mitra, R. D., Varma, C., Church, G. M., 2004, Advanced sequencing technologies: methods and goals. *Nature Reviews Genetics*, 5: 335–344

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P., 2014, Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*, 15: 121–132

Siska, V., Jones, E. R., Jeon, S., Bhak, Y., Kim, H.-M., Cho, Y. S.,Kim, H., Lee, K., Veselovskaya, E., Balueva, T., Gallego-Llorente, M., Hofreiter, M., Bradley, D. G., Eriksson, A., Pinhasi, R., Bhak, J., Manica, A., 2017, Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. *Science Advances*, *3*(2)

Skoglund, P., Storå, J., Götherström, A., & Jakobsson, M., 2013, Accurate sex identification of ancient human remains using DNA shotgun sequencing, *Journal of Archaeological Science*, *40*(12): 4477–4482

Soares, P., Achilli, A., Semino, O., Davies, W., Macaulay, V., Bandelt, H.J., Torroni, A., Richards, M. B., 2010, The Archaeogenetics of Europe, *Current Biology*, *20*(4): R174–R183

Svjetlana, C., Tolk, H. V., Barac-Lauc, L., Colak, I., Dordević, D., Efremovska, L., Janićijević, B., Kvesić, A., Martinović Klarić, I., Metspalu, E., Pericic Salihovic, M., Parik, J., Popović, D., Sijacki, A., Terzić, R., Villems, R., Rudan, P., 2004, Frequencies of mtDNA Haplogroups in Southeastern Europe-Croatians, Bosnians and Herzegovinians, Serbians, Macedonians and Macedonian Romani, *Collegium antropologicum,* 28

Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V. J., Sawyer, S. K., Greenfield, D. L., Germonpré, M. B., Sablin, M. V., López-Giráldez, F., Domingo-Roura, X., Napierala, H., Uerpmann, H-P., Loponte, D. M., Acosta, A. A., Giemsch, L., Schmitz, R. W., Worthington, B., Buikstra, J. E., Druzhkova, A., Graphodatsky, A. S., Ovodov, N. D., Wahlberg, N., Freedman, A. H., Schweizer, R. M., Koepfli, K. P.,Leonard, J. A., Meyer, M., Krause, J., Pääbo, S., Green, R. E.,Wayne, R. K., 2013, Complete Mitochondrial Genomes of Ancient Canids Suggest a European Origin of Domestic Dogs. *Science*, *342*(6160), 871-874

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

Thomas, R. H., Schaffner, W., Wilson, A. C., & Paabo, S., 1989, DNA phylogeny of the extinct marsupial wolf, *Nature*, *340*(6233): 465–467

Tobin, S., 1996, *The Cistercians: Monks and Monasteries of Europe*, Overlook Press, Retrieved from https://books.google.co.uk/books?id=Ij83AQAAIAAJ, Accessed 12/10/16

Töpf, A. L., Gilbert, M. T. P., Dumbacher, J. P., & Hoelzel, A. R., 2006, Tracing the Phylogeography of Human Populations in Britain Based on 4th–11th Century mtDNA Genotypes. *Molecular Biology and Evolution*, *23*(1): 152–161

Town, N. J., 2015, The Analysis Of Preserved And Degraded Human Skeletal Material: Understanding Relationships Between Bone And The Soil Environment, PhD Thesis, Liverpool John Moores University, United Kingdom

Turchinovich, A., Surowy, H., Serva, A., Zapatka, M., Lichter, P., & Burwinkel, B., 2014, Capture and Amplification by Tailing and Switching (CATS), *RNA Biology*, 11: 817–828

Van Oven, M., & Kayser, M., 2009, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, 30: E386–E394

Vestheim, H., Jarman, S. N., 2008, Blocking primers to enhance PCR amplification of rare sequences in mixed samples – a case study on prey DNA in Antarctic krill stomachs. *Frontiers in Zoology*, 5: 12

Vianello, D., Sevini, F., Castellani, G., Lomartire, L., Capri, M., & Franceschi, C., 2013, HAPLOFIND: A New Method for High-Throughput mtDNA Haplogroup Assignment, *Human Mutation*, *34*(9): 1189–1194

Wales, N., Caroe, C., Sandoval-Velasco, M., Gamba, C., Barnett, R., Samaniego, A, J., Madrigal, J, R., Orlando, L., Gilbert, M, T, P., 2015, New insights on single-stranded versus double-stranded DNA library preparation for ancient DNA, *BioTechniques*, 59:368-371

Walsh, P. S., Metzger, D. A., & Higushi, R., 1991, Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material, *BioTechniques*, *54*(3), 134–139

Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branicki, W., Kayser, M., 2013, The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Science International: Genetics*, 7: 98–115

Warinner, C., Rodrigues, J. F. M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R. Y., Fiddyment, S., Speller, C., Hendy, J., Charlton, S., Luder, H. U., Salazar-Garcia, D. C., Eppler, E., Seiler, R., Hansen, L. H., Castruita, J. A. S., Barkow-Oesterreicher, S., Teoh, K. Y., Kelstrup, C. D., Olsen, J. V., Nanni, P., Kawai, T., Willerslev, E., von Mering, C., Lewis Jr, C. M., Collins, M. J., Gilbert, M. T. P., Ruhli, F., Cappellini, E.,

2014, Pathogens and host immunity in the ancient human oral cavity. *Nature Genetics*, 46: 336–344.

Weale, M. E., Weiss, D. A., Jager, R. F., Bradman, N., Thomas, M. G., 2002, Y Chromosome Evidence for Anglo-Saxon Mass Migration. *Molecular Biology and Evolution*, 19: 1008–1021

Woodward, S. R., Weyand, N. J., Bunnell, M., 1994, DNA sequence from Cretaceous period bone fragments. *Science*, 266: 1229–1232

Zhang, J., Chiodini, R., Badr, A., & Zhang, G., 2011, The impact of next-generation sequencing on genomics, *Journal of Genetics and Genomics = Yi Chuan Xue Bao*, *38*(3): 95–109

Appendix

Appendix 1: Images of ten Poulton tooth samples that were sequenced

Figure A1: Skeleton 801 (Sample ASM007) tooth (Canine)

Figure A2: Skeleton 847 (Sample ASM008) tooth (Pre-molar)

Figure A3: Skeleton 854 (Sample ASM010) tooth (Molar)

Figure A4: Skeleton 823 (Sample ASM011) tooth (Canine)

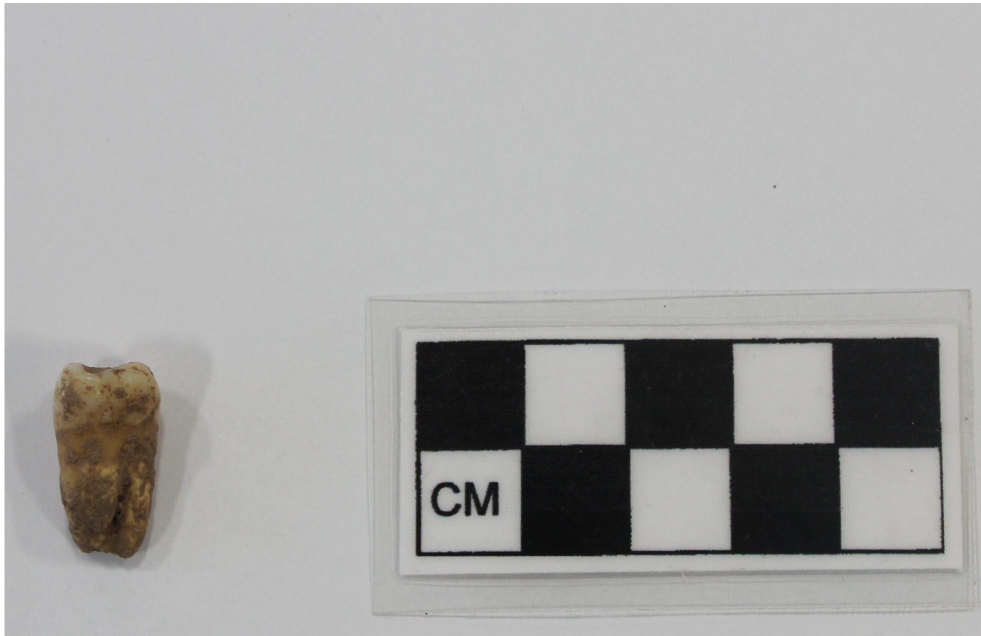


Figure A5: Skeleton 856 (Sample ASM012) tooth (Pre-molar)

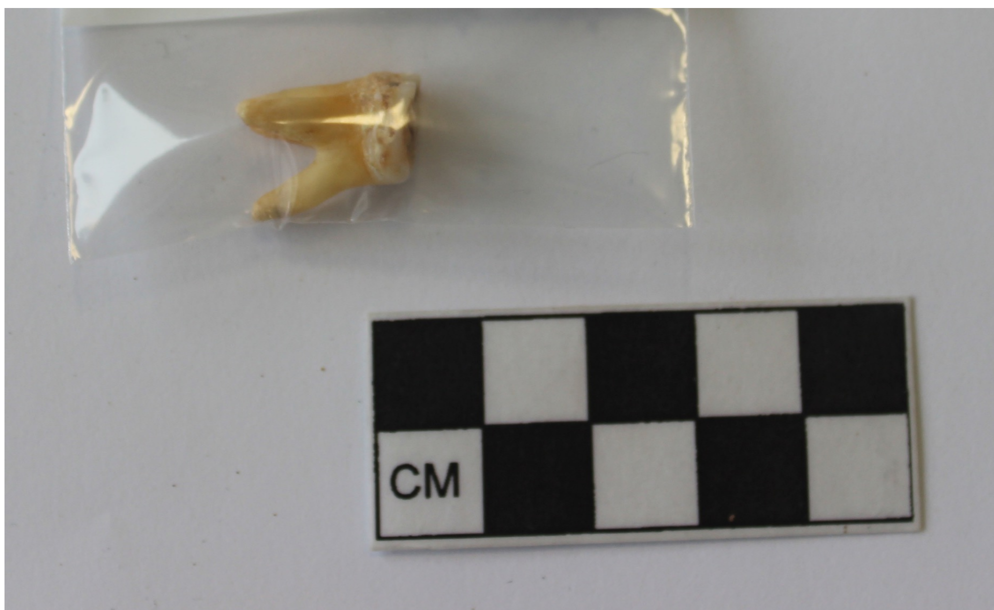


Figure A6: Skeleton 829 (Sample ASM013) tooth (Molar)

Figure A7: Skeleton 865 (Sample ASM014) tooth (canine)

Figure A8: Skeleton 834 (Sample ASM015) tooth (incisor)

Figure A9: Skeleton 873 (Sample ASM017) tooth (canine)

Figure A10: Skeleton 797 (Sample ASM018) tooth (Pre-molar)

Appendix 2: Primers and Indexing Oligos

2.1 Meyer Primers

Table A1: Primers for Meyer library preparation method *indicates a PTO bond

Oligo ID	Oligo sequence (5'-3')
IS1_adapter.P5	A*C*A*C*TCT TTC CCT ACA CGA CGC TCT TCC G*A*T*C*T
IS2_adapter.P7	G*T*G*A*CTG GAG TTC AGA CGT GTG CTC TTC CG*A*T*C*T
IS3_adapter.P5+P7	A*G*A*T*CGG AA*G*A*G*C
PCRf/IS4	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT

2.2 CATS Primers

Primers for dsDNA:

Table A2: Primers for dsDNA CATS library preparation method

Oligo ID	Oligo sequence (5'-3')
IPdTDPo	AGA CGT GTG CTC TTC CGA TCT TTTTTTTTTTTTTTTTTTTTTTTTTTTT V, where V= A+G+C
TSO	Biotin-5'-GTT CAG AGT TCT ACA GTC CGA CGA TC rGrGrG, where rG = riboguanidine
PCRf	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG TTC AGA GTT CTA CAG TCC GA
PCRr	CAA GCA GAA GAC GGC ATA CGA GAT CGT GAT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT
Illumina Custom Sequencing Primer	GTT CAG AGT TCT ACA GTC CGA CGA TCG GG

In our case primers were edited to be compatible with Meyer (the changes are in bold, adapted from Meyer primers):

Table A3: Edited primers for dsDNA CATS library preparation method

Oligo ID	Oligo sequence (5'-3')
IPdTDPo (used the P7 sequence, i.e. IS2)	**GTG ACT GGA GTT C**AG ACG TGT GCT CTT CCG ATC TTTTTTTTTTTTTTTTTTTT TTTTTTTTTT V 3', where V= A+G+C
TSO (used the P5 sequences, i.e. IS1)	Biotin-5' **ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT** rGrGrG 3', where rG = riboguanidine
PCRf (IS4 primer; complementary to P5)	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT
PCRr can be any indexing oligo (complementary to P7; index in bold) e.g.	CAA GCA GAA GAC GGC ATA CGA GAT **cctgcga** GTG ACT GGA GTT CAG ACG TGT
TruSeq Universal Adapter	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T

2.3 Indexing oligos

Table A4: Indexing oligos for library preparation

Oligo ID	Oligo sequence (5'-3') (index marked with small letters)	Index sequence (5'-3')
indexing1	CAAGCAGAAGACGGCATACGAGATcctgcgaGTGACTGGAGTTCAGACGTGT	TCGCAGG
indexing2	CAAGCAGAAGACGGCATACGAGATtgcagagGTGACTGGAGTTCAGACGTGT	CTCTGCA
indexing3	CAAGCAGAAGACGGCATACGAGATacctaggGTGACTGGAGTTCAGACGTGT	CCTAGGT
indexing4	CAAGCAGAAGACGGCATACGAGATttgatccGTGACTGGAGTTCAGACGTGT	GGATCAA
indexing5	CAAGCAGAAGACGGCATACGAGATatcttgcGTGACTGGAGTTCAGACGTGT	GCAAGAT
indexing6	CAAGCAGAAGACGGCATACGAGATtctccatGTGACTGGAGTTCAGACGTGT	ATGGAGA
indexing7	CAAGCAGAAGACGGCATACGAGATcatcgagGTGACTGGAGTTCAGACGTGT	CTCGATG
indexing8	CAAGCAGAAGACGGCATACGAGATttcgagcGTGACTGGAGTTCAGACGTGT	GCTCGAA
indexing9	CAAGCAGAAGACGGCATACGAGATagttggtGTGACTGGAGTTCAGACGTGT	ACCAACT
indexing10	CAAGCAGAAGACGGCATACGAGATgtaccggGTGACTGGAGTTCAGACGTGT	CCGGTAC
indexing11	CAAGCAGAAGACGGCATACGAGATcggagttGTGACTGGAGTTCAGACGTGT	AACTCCG
indexing12	CAAGCAGAAGACGGCATACGAGATacttcaaGTGACTGGAGTTCAGACGTGT	TTGAAGT
indexing13	CAAGCAGAAGACGGCATACGAGATtgatagtGTGACTGGAGTTCAGACGTGT	ACTATCA
indexing14	CAAGCAGAAGACGGCATACGAGATgatccaaGTGACTGGAGTTCAGACGTGT	TTGGATC
indexing15	CAAGCAGAAGACGGCATACGAGATcaggtcgGTGACTGGAGTTCAGACGTGT	CGACCTG

indexing16	CAAGCAGAAGACGGCATACGAGATcgcattaGTGACTGGAGTTCAGACGTGT	TAATGCG
indexing17	CAAGCAGAAGACGGCATACGAGATggtacctGTGACTGGAGTTCAGACGTGT	AGGTACC
indexing18	CAAGCAGAAGACGGCATACGAGATggacgcaGTGACTGGAGTTCAGACGTGT	TGCGTCC
indexing19	CAAGCAGAAGACGGCATACGAGATgagattcGTGACTGGAGTTCAGACGTGT	GAATCTC
indexing20	CAAGCAGAAGACGGCATACGAGATgagcatgGTGACTGGAGTTCAGACGTGT	CATGCTC
indexing21	CAAGCAGAAGACGGCATACGAGATgttgcgtGTGACTGGAGTTCAGACGTGT	ACGCAAC
indexing22	CAAGCAGAAGACGGCATACGAGATccaatgcGTGACTGGAGTTCAGACGTGT	GCATTGG
indexing23	CAAGCAGAAGACGGCATACGAGATcgagatcGTGACTGGAGTTCAGACGTGT	GATCTCG
indexing24	CAAGCAGAAGACGGCATACGAGATcatattgGTGACTGGAGTTCAGACGTGT	CAATATG
indexing25	CAAGCAGAAGACGGCATACGAGATgacgtcaGTGACTGGAGTTCAGACGTGT	TGACGTC

Appendix 3: Mitochondrial DNA

SNPS and mutations generated from Haplofind results

Table A5: Mitochondrial SNP mutations and resulting diseases for the ten sequenced Poulton samples A:ASM007 B:ASM008 C:ASM010 D:ASM011 E:ASM012 F:ASM013 J:ASM014 H:ASM015 I:ASM017 J:ASM018

A: ASM007

SNP	Loci	Diseases
146T	MT-DLOOP	
195T	MT-DLOOP	
199C	MT-DLOOP	
204C	MT-DLOOP	
207A	MT-DLOOP	
247G	MT-DLOOP	
250C	MT-DLOOP	
409del	MT-DLOOP	
436_437del	MT-DLOOP	
534del	MT-DLOOP	
559T	MT-DLOOP	
626del	MT-TF	
722del	MT-RNR1	
724del	MT-RNR1	
736del	MT-RNR1	
769G	MT-RNR1	
825T	MT-RNR1	
1018G	MT-RNR1	
1156del	MT-RNR1	
1719A	MT-RNR2	
1785del	MT-RNR2	
1880del	MT-RNR2	
1883del	MT-RNR2	
1886del	MT-RNR2	
2140del	MT-RNR2	
2758G	MT-RNR2	
2885T	MT-RNR2	
3090del	MT-RNR2	
3594C	MT-ND1 (Syn)	
3780del	MT-ND1	
3794_3795del	MT-ND1	
3884T	MT-ND1 (T->I pos:193)	
3889T	MT-ND1 (R->W pos:195)	
3952A	MT-ND1 (A->T pos:216)	
4104A	MT-ND1 (Syn)	
4259del	MT-ND1	
4312C	MT-TI	

4878del	MT-ND2	
5532del	MT-TW	
5646T	MT-TA	
5651T	MT-TA	
5657T	MT-TN	
5671T	MT-TN	
6239del	MT-CO1	
6347del	MT-CO1	
6363del	MT-CO1	
6871del	MT-CO1	
6888del	MT-CO1	
7223T	MT-CO1 (Syn)	
7229T	MT-CO1 (Syn)	
7236A	MT-CO1 (D->N pos:445)	
7256C	MT-CO1 (Syn)	
8369_8371del	MT-ATP8	
8468C	MT-ATP8 (Syn)	
8592del	MT-ATP6	
8607T	MT-ATP6 (Syn)	
8655C	MT-ATP6 (Syn)	
8701A	MT-ATP6 (A->T pos:59)	
9176del	MT-ATP6	
9408T	MT-CO3 (Q->* pos:68)	
9469del	MT-CO3	
9500del	MT-CO3	
9540T	MT-CO3 (Syn)	
9655del	MT-CO3	
9671del	MT-CO3	
10034C	MT-TG	
10238C	MT-ND3 (Syn)	
10530del	MT-ND4L	
10556del	MT-ND4L	
10565del	MT-ND4L	
10664C	MT-ND4L (Syn)	
10688G	MT-ND4L (Syn)	
10810T	MT-ND4 (Syn)	
10873T	MT-ND4 (Syn)	
10915T	MT-ND4 (Syn)	
11914G	MT-ND4 (Syn)	
11919del	MT-ND4	
12000del	MT-ND4	
12002del	MT-ND4	
12016del	MT-ND4	
12018del	MT-ND4	
12174del	MT-TH	
12372del	MT-ND5	
12501A	MT-ND5 (Syn)	

13105A	MT-ND5 (V->I pos:257)	
13276A	MT-ND5 (V->M pos:314)	
13506C	MT-ND5 (Syn)	
13650C	MT-ND5 (Syn)	
13780G	MT-ND5 (I->V pos:482)	
13791del	MT-ND5	
13825del	MT-ND5	
14054del	MT-ND5	
14066del	MT-ND5	
14124del	MT-ND5	
14471del	MT-ND6	
14816del	MT-CYB	
14920del	MT-CYB	
14925del	MT-CYB	
15043A	MT-CYB (Syn)	MDD-associated (Reported)
15758G	MT-CYB (I->V pos:338)	
15924G	MT-TT	LIMM (Point Mutation - Non Pathogenic)
16187C	MT-DLOOP	
16189T	MT-DLOOP	
16230A	MT-DLOOP	
16234T	MT-DLOOP	
16278C	MT-DLOOP	
16311T	MT-DLOOP	
16391A	MT-DLOOP	
16569del	MT-DLOOP	

B: ASM008

SNP	**Loci**	**Diseases**
73A	MT-DLOOP	
146T	MT-DLOOP	
152T	MT-DLOOP	
195T	MT-DLOOP	
247G	MT-DLOOP	
769G	MT-RNR1	
825T	MT-RNR1	
1018G	MT-RNR1	
1415A	MT-RNR1	
2706A	MT-RNR2	
2758G	MT-RNR2	
2885T	MT-RNR2	
3010A	MT-RNR2	Cyclic Vomiting Syndrome with Migraine (Reported; also common pm)
3594C	MT-ND1 (Syn)	
4104A	MT-ND1 (Syn)	
4221del	MT-ND1	
4312C	MT-TI	
4769A	MT-ND2 (Syn)	SZ-associated (Reported)
7028C	MT-CO1 (Syn)	

7146A	MT-CO1 (A->T pos:415)	
7256C	MT-CO1 (Syn)	
7521G	MT-TD	
8468C	MT-ATP8 (Syn)	
8655C	MT-ATP6 (Syn)	
8701A	MT-ATP6 (A->T pos:59)	
9540T	MT-CO3 (Syn)	
9921A	MT-CO3 (A->T pos:239)	
10398A	MT-ND3 (A->T pos:114)	Invasive Breast Cancer risk factor; AD; PD; BD lithium response; Type 2 DM (Reported; haplogroup HNTUVWXK2 marker)
10664C	MT-ND4L (Syn)	
10688G	MT-ND4L (Syn)	
10810T	MT-ND4 (Syn)	
10873T	MT-ND4 (Syn)	
10915T	MT-ND4 (Syn)	
11719G	MT-ND4 (Syn)	
11914G	MT-ND4 (Syn)	
12705C	MT-ND5 (Syn)	
13105A	MT-ND5 (V->I pos:257)	
13276A	MT-ND5 (V->M pos:314)	
13506C	MT-ND5 (Syn)	
13650C	MT-ND5 (Syn)	
14766C	MT-CYB (I->T pos:7)	
16111T	MT-DLOOP	
16129G	MT-DLOOP	
16187C	MT-DLOOP	
16189T	MT-DLOOP	
16223C	MT-DLOOP	
16230A	MT-DLOOP	
16278C	MT-DLOOP	
16311T	MT-DLOOP	
16568_16569del	MT-DLOOP	

C: ASM010

SNP	Loci	Diseases
146T	MT-DLOOP	
150T	MT-DLOOP	Longevity / Cervical Carcinoma / HPV infection risk (Conflicting reports)
152T	MT-DLOOP	
195T	MT-DLOOP	
247G	MT-DLOOP	
412del	MT-DLOOP	
723G	MT-RNR1	
769G	MT-RNR1	
825T	MT-RNR1	
1018G	MT-RNR1	
1721T	MT-RNR2	
2758G	MT-RNR2	

2885T	MT-RNR2	
3197C	MT-RNR2	
3594C	MT-ND1 (Syn)	
3861G	MT-ND1 (Syn)	
4104A	MT-ND1 (Syn)	
4312C	MT-TI	
4461T	MT-TM	
4691T	MT-ND2 (Syn)	
4692T	MT-ND2 (L->F pos:75)	
4720A	MT-ND2 (W->* pos:84)	
4769A	MT-ND2 (Syn)	SZ-associated (Reported)
5836G	MT-TY	
5851del	MT-TY	
6038T	MT-CO1 (Syn)	
6132del	MT-CO1	
7146A	MT-CO1 (A->T pos:415)	
7256C	MT-CO1 (Syn)	
7287del	MT-CO1	
7478A	MT-TS1	
7521G	MT-TD	
7768G	MT-CO2 (Syn)	
8141del	MT-CO2	
8294del	MT-NC7	
8468C	MT-ATP8 (Syn)	
8655C	MT-ATP6 (Syn)	
8701A	MT-ATP6 (A->T pos:59)	
9086del	MT-ATP6	
9477A	MT-CO3 (V->I pos:91)	
9540T	MT-CO3 (Syn)	
10262G	MT-ND3 (Syn)	
10398A	MT-ND3 (A->T pos:114)	Invasive Breast Cancer risk factor; AD; PD; BD lithium response; Type 2 DM (Reported; haplogroup HNTUVWXK2 marker)
10664C	MT-ND4L (Syn)	
10688G	MT-ND4L (Syn)	
10810T	MT-ND4 (Syn)	
10873T	MT-ND4 (Syn)	
10915T	MT-ND4 (Syn)	
11467G	MT-ND4 (Syn)	Altered brain pH (Reported)
11914G	MT-ND4 (Syn)	
12218del	MT-TS2	
12308G	MT-TL2	CPEO / Stroke / CM / Breast & Renal & Prostate Cancer Risk/ Altered brain pH (Hg U marker)
12372A	MT-ND5 (Syn)	Altered brain pH (Reported)
12705C	MT-ND5 (Syn)	
13017G	MT-ND5 (Syn)	
13105A	MT-ND5 (V->I pos:257)	
13276A	MT-ND5 (V->M pos:314)	
13506C	MT-ND5 (Syn)	
13617C	MT-ND5 (Syn)	
13637G	MT-ND5 (Q->R pos:434)	Possible LHON factor (Reported)
13650C	MT-ND5 (Syn)	

14182C	MT-ND6 (Syn)	
16129G	MT-DLOOP	
16187C	MT-DLOOP	
16192T	MT-DLOOP	
16223C	MT-DLOOP	
16230A	MT-DLOOP	
16249C	MT-DLOOP	
16270T	MT-DLOOP	
16278C	MT-DLOOP	
16311T	MT-DLOOP	
16519T	MT-DLOOP	Cyclic Vomiting Syndrome with Migraine (Reported)
16568_16569del	MT-DLOOP	

D: ASM011

SNP	**Loci**	**Diseases**
73A	MT-DLOOP	
85del	MT-DLOOP	
146T	MT-DLOOP	
437del	MT-DLOOP	
494T	MT-DLOOP	
497T	MT-DLOOP	
501T	MT-DLOOP	
601del	MT-TF	
611del	MT-TF	
652del	MT-RNR1	
687A	MT-RNR1	
723G	MT-RNR1	
769G	MT-RNR1	
825T	MT-RNR1	
847A	MT-RNR1	
1320del	MT-RNR1	
1333del	MT-RNR1	
1393del	MT-RNR1	
1539del	MT-RNR1	
1549del	MT-RNR1	
1551del	MT-RNR1	
1652del	MT-TV	
2004del	MT-RNR2	
2028del	MT-RNR2	
2364del	MT-RNR2	
2585del	MT-RNR2	
2592del	MT-RNR2	
2683del	MT-RNR2	
2687del	MT-RNR2	
2706A	MT-RNR2	
2757del	MT-RNR2	
2885T	MT-RNR2	
3010A	MT-RNR2	Cyclic Vomiting Syndrome with Migraine (Reported; also common pm)

3240del	MT-TER|MT-TL1	
3292T	MT-TL1	
3324T	MT-ND1 (Syn)	
3594C	MT-ND1 (Syn)	
4249T	MT-ND1 (P->S pos:315)	
4312C	MT-TI	
4826del	MT-ND2	
5354del	MT-ND2	
6119T	MT-CO1 (Syn)	
6174A	MT-CO1 (D->N pos:91)	
6190del	MT-CO1	
6260A	MT-CO1 (Syn)	
6261A	MT-CO1 (A->T pos:120)	Prostate Cancer/LHON (Reported)
6265A	MT-CO1 (G->E pos:121)	
6434del	MT-CO1	
6591del	MT-CO1	
6596_6597del	MT-CO1	
6639del	MT-CO1	
6959T	MT-CO1 (Syn)	
7256C	MT-CO1 (Syn)	
7499T	MT-TS1	
7521G	MT-TD	
8080T	MT-CO2 (Syn)	
8155del	MT-CO2	
8292A	MT-NC7	
8468C	MT-ATP8 (Syn)	
8655C	MT-ATP6 (Syn)	
8701A	MT-ATP6 (A->T pos:59)	
9081del	MT-ATP6	
9118T	MT-ATP6 (Syn)	
9153T	MT-ATP6 (Syn)	
9213T	MT-CO3 (H->Y pos:3)	
9444del	MT-CO3	
9469del	MT-CO3	
9540T	MT-CO3 (Syn)	
9562del	MT-CO3	
9841del	MT-CO3	
9851del	MT-CO3	
10325A	MT-ND3 (Syn)	
10398A	MT-ND3 (A->T pos:114)	Invasive Breast Cancer risk factor; AD; PD; BD lithium response; Type 2 DM (Reported; haplogroup HNTUVWXK2 marker)
10466del	MT-TR	
10473_10474del	MT-ND4L	
10497del	MT-ND4L	
10664C	MT-ND4L (Syn)	
10688G	MT-ND4L (Syn)	
10810T	MT-ND4 (Syn)	
10873T	MT-ND4 (Syn)	
10915T	MT-ND4 (Syn)	
11426del	MT-ND4	

11453A	MT-ND4 (A->T pos:232)	
11634A	MT-ND4 (S->N pos:292)	
11651A	MT-ND4 (V->M pos:298)	
11719G	MT-ND4 (Syn)	
11798del	MT-ND4	
11864C	MT-ND4 (Syn)	
11914G	MT-ND4 (Syn)	
12079del	MT-ND4	
12147del	MT-TH	
12213del	MT-TS2	
12224del	MT-TS2	
12227del	MT-TS2	
12402T	MT-ND5 (Syn)	
12411del	MT-ND5	
12541del	MT-ND5	
12844del	MT-ND5	
12958T	MT-ND5 (P->S pos:208)	
12959T	MT-ND5 (P->L pos:208)	
13105A	MT-ND5 (V->I pos:257)	
13160del	MT-ND5	
13239del	MT-ND5	
13244del	MT-ND5	
13270del	MT-ND5	
13276A	MT-ND5 (V->M pos:314)	
13448del	MT-ND5	
13506C	MT-ND5 (Syn)	
13650C	MT-ND5 (Syn)	
13889del	MT-ND5	
13940A	MT-ND5 (R->H pos:535)	
13977del	MT-ND5	
13981_13982del	MT-ND5	
14225T	MT-ND6 (Syn)	
14603del	MT-ND6	
14753del	MT-CYB	
14766C	MT-CYB (I->T pos:7)	
14804del	MT-CYB	
14957del	MT-CYB	
15285T	MT-CYB (T->I pos:180)	
15297C	MT-CYB (I->T pos:184)	
15575del	MT-CYB	
15665T	MT-CYB (L->F pos:307)	
15676T	MT-CYB (Syn)	
15700del	MT-CYB	
15704del	MT-CYB	
15723del	MT-CYB	
15807T	MT-CYB (A->V pos:354)	
15871del	MT-CYB	
15882del	MT-CYB	
15888del	MT-TT	
15927del	MT-TT	

16000del	MT-TP	
16129G	MT-DLOOP	
16187C	MT-DLOOP	
16189T	MT-DLOOP	
16230A	MT-DLOOP	
16255del	MT-DLOOP	
16278C	MT-DLOOP	
16310del	MT-DLOOP	
16311T	MT-DLOOP	
16390del	MT-DLOOP	
16412del	MT-DLOOP	
16449del	MT-DLOOP	
16478del	MT-DLOOP	
16496del	MT-DLOOP	
16503del	MT-DLOOP	
16560_16569del	MT-DLOOP	

E: ASM012

SNP	Loci	Diseases
73A	MT-DLOOP	
146T	MT-DLOOP	
152T	MT-DLOOP	
195T	MT-DLOOP	
247G	MT-DLOOP	
263A	MT-DLOOP	
750A	MT-RNR1	SZ-associated (Reported)
769G	MT-RNR1	
825T	MT-RNR1	
1018G	MT-RNR1	
1203del	MT-RNR1	
1268del	MT-RNR1	
1438A	MT-RNR1	SZ-associated (Reported)
2212del	MT-RNR2	
2217del	MT-RNR2	
2706A	MT-RNR2	
2713del	MT-RNR2	
2758G	MT-RNR2	
2816del	MT-RNR2	
2885T	MT-RNR2	
3352del	MT-ND1	
3594C	MT-ND1 (Syn)	
3893del	MT-ND1	
4104A	MT-ND1 (Syn)	
4140del	MT-ND1	
4170T	MT-ND1 (Syn)	
4194T	MT-ND1 (Syn)	
4312C	MT-TI	
4428A	MT-TM	
4440A	MT-TM	

4769A	MT-ND2 (Syn)	SZ-associated (Reported)
5223del	MT-ND2	
6111del	MT-CO1	
6117del	MT-CO1	
6657del	MT-CO1	
7028C	MT-CO1 (Syn)	
7146A	MT-CO1 (A->T pos:415)	
7256C	MT-CO1 (Syn)	
7327del	MT-CO1	
7521G	MT-TD	
7539del	MT-TD	
7554del	MT-TD	
7565del	MT-TD	
8468C	MT-ATP8 (Syn)	
8655C	MT-ATP6 (Syn)	
8701del	MT-ATP6	
8702A	MT-ATP6 (A->D pos:59)	
8790A	MT-ATP6 (Syn)	
8860A	MT-ATP6 (A->T pos:112)	
9540T	MT-CO3 (Syn)	
9677del	MT-CO3	
9908del	MT-CO3	
10398A	MT-ND3 (A->T pos:114)	Invasive Breast Cancer risk factor; AD; PD; BD lithium response; Type 2 DM (Reported; haplogroup HNTUVWXK2 marker)
10664C	MT-ND4L (Syn)	
10688G	MT-ND4L (Syn)	
10810T	MT-ND4 (Syn)	
10873T	MT-ND4 (Syn)	
10915T	MT-ND4 (Syn)	
11331del	MT-ND4	
11719G	MT-ND4 (Syn)	
11778del	MT-ND4	
11914G	MT-ND4 (Syn)	
12389del	MT-ND5	
12705C	MT-ND5 (Syn)	
13105A	MT-ND5 (V->I pos:257)	
13138del	MT-ND5	
13276A	MT-ND5 (V->M pos:314)	
13296del	MT-ND5	
13457del	MT-ND5	
13506C	MT-ND5 (Syn)	
13650C	MT-ND5 (Syn)	
13882del	MT-ND5	
14528del	MT-ND6	
14766C	MT-CYB (I->T pos:7)	
14804del	MT-CYB	
15326A	MT-CYB (A->T pos:194)	
16129G	MT-DLOOP	
16187C	MT-DLOOP	
16189T	MT-DLOOP	
16223C	MT-DLOOP	

16230A	MT-DLOOP	
16278C	MT-DLOOP	
16311T	MT-DLOOP	
16519T	MT-DLOOP	Cyclic Vomiting Syndrome with Migraine (Reported)

F: ASM013

SNP	**Loci**	**Diseases**
146T	MT-DLOOP	
152T	MT-DLOOP	
194T	MT-DLOOP	
204C	MT-DLOOP	
207A	MT-DLOOP	
247G	MT-DLOOP	
329del	MT-DLOOP	
709A	MT-RNR1	
769G	MT-RNR1	
825T	MT-RNR1	
1018G	MT-RNR1	
1243C	MT-RNR1	
2356G	MT-RNR2	
2758G	MT-RNR2	
2885T	MT-RNR2	
3505G	MT-ND1 (T->A pos:67)	
3594C	MT-ND1 (Syn)	
4104A	MT-ND1 (Syn)	
4221T	MT-ND1 (Syn)	
4241del	MT-ND1	
4312C	MT-TI	
5046A	MT-ND2 (V->I pos:193)	
5460A	MT-ND2 (A->T pos:331)	AD/PD (Point Mutation - Non Pathogenic)
6356del	MT-CO1	
6528T	MT-CO1 (Syn)	
7146A	MT-CO1 (A->T pos:415)	
7256C	MT-CO1 (Syn)	
7521G	MT-TD	
8251A	MT-CO2 (Syn)	
8292A	MT-NC7	
8468C	MT-ATP8 (Syn)	
8655C	MT-ATP6 (Syn)	
8701A	MT-ATP6 (A->T pos:59)	
8994A	MT-ATP6 (Syn)	
9540T	MT-CO3 (Syn)	
9655del	MT-CO3	
10097G	MT-ND3 (Syn)	
10398A	MT-ND3 (A->T pos:114)	Invasive Breast Cancer risk factor; AD; PD; BD lithium response; Type 2 DM (Reported; haplogroup HNTUVWXK2 marker)
10410C	MT-TR	
10664C	MT-ND4L (Syn)	
10688G	MT-ND4L (Syn)	
10810T	MT-ND4 (Syn)	

10873T	MT-ND4 (Syn)	
10915T	MT-ND4 (Syn)	
11674T	MT-ND4 (Syn)	
11914G	MT-ND4 (Syn)	
11947G	MT-ND4 (Syn)	
12414C	MT-ND5 (Syn)	
13105A	MT-ND5 (V->I pos:257)	
13276A	MT-ND5 (V->M pos:314)	
13506C	MT-ND5 (Syn)	
13626del	MT-ND5	
13650C	MT-ND5 (Syn)	
15775G	MT-CYB (Syn)	
15884C	MT-CYB (A->P pos:380)	
16129G	MT-DLOOP	
16187C	MT-DLOOP	
16189T	MT-DLOOP	
16230A	MT-DLOOP	
16278C	MT-DLOOP	
16292T	MT-DLOOP	
16311T	MT-DLOOP	
16568_16569del	MT-DLOOP	

G: ASM014

SNP	Loci	Diseases
146T	MT-DLOOP	
152T	MT-DLOOP	
195T	MT-DLOOP	
247G	MT-DLOOP	
769G	MT-RNR1	
825T	MT-RNR1	
1018G	MT-RNR1	
2706A	MT-RNR2	
2758G	MT-RNR2	
2885T	MT-RNR2	
3010A	MT-RNR2	Cyclic Vomiting Syndrome with Migraine (Reported; also common pm)
3594C	MT-ND1 (Syn)	
4104A	MT-ND1 (Syn)	
4312C	MT-TI	
5203T	MT-ND2 (P->L pos:245)	
6365C	MT-CO1 (Syn)	
7028C	MT-CO1 (Syn)	
7109T	MT-CO1 (Syn)	
7146A	MT-CO1 (A->T pos:415)	
7256C	MT-CO1 (Syn)	
7521G	MT-TD	
8468C	MT-ATP8 (Syn)	
8655C	MT-ATP6 (Syn)	
8701A	MT-ATP6 (A->T pos:59)	

9540T	MT-CO3 (Syn)	
10398A	MT-ND3 (A->T pos:114)	Invasive Breast Cancer risk factor; AD; PD; BD lithium response; Type 2 DM (Reported; haplogroup HNTUVWXK2 marker)
10664C	MT-ND4L (Syn)	
10688G	MT-ND4L (Syn)	
10810T	MT-ND4 (Syn)	
10873T	MT-ND4 (Syn)	
10915T	MT-ND4 (Syn)	
11719G	MT-ND4 (Syn)	
11914G	MT-ND4 (Syn)	
12705C	MT-ND5 (Syn)	
13105A	MT-ND5 (V->I pos:257)	
13276A	MT-ND5 (V->M pos:314)	
13506C	MT-ND5 (Syn)	
13650C	MT-ND5 (Syn)	
14252del	MT-ND6	
14766C	MT-CYB (I->T pos:7)	
16129G	MT-DLOOP	
16187C	MT-DLOOP	
16189T	MT-DLOOP	
16209C	MT-DLOOP	
16223C	MT-DLOOP	
16230A	MT-DLOOP	
16278C	MT-DLOOP	
16311T	MT-DLOOP	
16568_16569del	MT-DLOOP	

H: ASM015

SNP	Loci	Diseases
146T	MT-DLOOP	
152T	MT-DLOOP	
195T	MT-DLOOP	
247G	MT-DLOOP	
263A	MT-DLOOP	
349T	MT-DLOOP	
709A	MT-RNR1	
769G	MT-RNR1	
825T	MT-RNR1	
1018G	MT-RNR1	
1079A	MT-RNR1	
1395del	MT-RNR1	
1405del	MT-RNR1	
1642A	MT-TV	MELAS (Reported)
1677del	MT-RNR2	
1681_1682del	MT-RNR2	
1725del	MT-RNR2	
1732del	MT-RNR2	
1764del	MT-RNR2	
1816A	MT-RNR2	

2141C	MT-RNR2	
2440del	MT-RNR2	
2758G	MT-RNR2	
2804del	MT-RNR2	
2867del	MT-RNR2	
2870del	MT-RNR2	
2885T	MT-RNR2	
2933del	MT-RNR2	
2943del	MT-RNR2	
3235del	MT-TER|MT-TL1	
3350C	MT-ND1 (I->T pos:15)	
3406del	MT-ND1	
3522del	MT-ND1	
3524del	MT-ND1	
3570del	MT-ND1	
3594C	MT-ND1 (Syn)	
3625A	MT-ND1 (A->T pos:107)	
3697del	MT-ND1	
3896del	MT-ND1	
3918del	MT-ND1	
3952A	MT-ND1 (A->T pos:216)	
4312C	MT-TI	
4799T	MT-ND2 (Syn)	
5066T	MT-ND2 (Syn)	
5112A	MT-ND2 (A->T pos:215)	
5730T	MT-OLR	
6207del	MT-CO1	
6314del	MT-CO1	
6335del	MT-CO1	
6551T	MT-CO1 (Syn)	
7121T	MT-CO1 (Syn)	
7146A	MT-CO1 (A->T pos:415)	
7256C	MT-CO1 (Syn)	
7259del	MT-CO1	
7274del	MT-CO1	
7521G	MT-TD	
7917_7918del	MT-CO2	
7920del	MT-CO2	
8018del	MT-CO2	
8393_8394del	MT-ATP8	
8468C	MT-ATP8 (Syn)	
8655C	MT-ATP6 (Syn)	
8697A	MT-ATP6 (Syn)	
8701A	MT-ATP6 (A->T pos:59)	
9117C	MT-ATP6 (Syn)	
9253A	MT-CO3 (W->* pos:16)	
9262A	MT-CO3 (T->K pos:19)	
9313del	MT-CO3	
9540T	MT-CO3 (Syn)	
9752del	MT-CO3	
9908del	MT-CO3	

9933del	MT-CO3	
9952del	MT-CO3	
10069del	MT-ND3	
10536T	MT-ND4L (R->C pos:23)	
10589A	MT-ND4L (Syn)	
10619del	MT-ND4L	
10664C	MT-ND4L (Syn)	
10688G	MT-ND4L (Syn)	
10810T	MT-ND4 (Syn)	
10873T	MT-ND4 (Syn)	
10915T	MT-ND4 (Syn)	
10938T	MT-ND4 (P->L pos:60)	
11061del	MT-ND4	
11099del	MT-ND4	
11193del	MT-ND4	
11251G	MT-ND4 (Syn)	
11312del	MT-ND4	
11423del	MT-ND4	
11427del	MT-ND4	
11434del	MT-ND4	
11494del	MT-ND4	
11497del	MT-ND4	
11512_11513del	MT-ND4	
11812G	MT-ND4 (Syn)	
11914G	MT-ND4 (Syn)	
12207del	MT-TS2	
12300del	MT-TL2	
12341del	MT-ND5	
12406del	MT-ND5	
12572del	MT-ND5	
12705C	MT-ND5 (Syn)	
12741T	MT-ND5 (Syn)	
12765del	MT-ND5	
12775del	MT-ND5	
12792del	MT-ND5	
13100T	MT-ND5 (A->V pos:255)	
13105A	MT-ND5 (V->I pos:257)	
13114del	MT-ND5	
13199del	MT-ND5	
13276A	MT-ND5 (V->M pos:314)	
13368A	MT-ND5 (Syn)	
13506C	MT-ND5 (Syn)	
13650C	MT-ND5 (Syn)	
13965C	MT-ND5 (Syn)	
13966G	MT-ND5 (T->A pos:544)	
14054del	MT-ND5	
14066del	MT-ND5	
14233G	MT-ND6 (Syn)	
14243A	MT-ND6 (L->* pos:32)	
14258A	MT-ND6 (V->E pos:37)	
14279A	MT-ND6 (L->Q pos:44)	LHON (Reported)

14807T	MT-CYB (L->F pos:21)	
14897T	MT-CYB (Syn)	
14905A	MT-CYB (Syn)	
14958del	MT-CYB	
14993del	MT-CYB	
14995del	MT-CYB	
15205_15206del	MT-CYB	
15223del	MT-CYB	
15438_15439del	MT-CYB	
15446del	MT-CYB	
15452A	MT-CYB (L->I pos:236)	
15607G	MT-CYB (Syn)	
15839T	MT-CYB (Syn)	
15915A	MT-TT	Encephalomyopathy (Reported)
15928A	MT-TT	Multiple Sclerosis/idiopathic repeat miscarriage/AD protection (P.M./possible helper mutation)
16067del	MT-DLOOP	
16069del	MT-DLOOP	
16071del	MT-DLOOP	
16082del	MT-DLOOP	
16126C	MT-DLOOP	
16129G	MT-DLOOP	
16187C	MT-DLOOP	
16189T	MT-DLOOP	
16223del	MT-DLOOP	
16230A	MT-DLOOP	
16273A	MT-DLOOP	
16278C	MT-DLOOP	
16311T	MT-DLOOP	
16320T	MT-DLOOP	
16324C	MT-DLOOP	
16332T	MT-DLOOP	
16384del	MT-DLOOP	
16398del	MT-DLOOP	
16552_16569del	MT-DLOOP	

I: ASM017

SNP	**Loci**	**Diseases**
146T	MT-DLOOP	
152T	MT-DLOOP	
164del	MT-DLOOP	
195T	MT-DLOOP	
247G	MT-DLOOP	
250C	MT-DLOOP	
564T	MT-DLOOP	
698del	MT-RNR1	
709del	MT-RNR1	
769G	MT-RNR1	
825T	MT-RNR1	
893del	MT-RNR1	

1018G	MT-RNR1	
1248T	MT-RNR1	
1290del	MT-RNR1	
1293del	MT-RNR1	
1472del	MT-RNR1	
1473R	MT-RNR1	
1475del	MT-RNR1	
1719A	MT-RNR2	
1836G	MT-RNR2	
1933del	MT-RNR2	
2758G	MT-RNR2	
2796del	MT-RNR2	
2885T	MT-RNR2	
3225del	MT-RNR2|MT-RNR3	
3551del	MT-ND1	
3594C	MT-ND1 (Syn)	
4043T	MT-ND1 (T->M pos:246)	
4104A	MT-ND1 (Syn)	
4250T	MT-ND1 (P->L pos:315)	
4312C	MT-TI	
4392T	MT-TQ	Poss. hypertension factor (Reported)
4393T	MT-TQ	
4394T	MT-TQ	
4982del	MT-ND2	
4987_4988del	MT-ND2	
5014del	MT-ND2	
5046del	MT-ND2	
5052del	MT-ND2	
5060del	MT-ND2	
5365del	MT-ND2	
6168_6170del	MT-CO1	
6267del	MT-CO1	
6361del	MT-CO1	
6734A	MT-CO1 (Syn)	
7146A	MT-CO1 (A->T pos:415)	
7256C	MT-CO1 (Syn)	
7650del	MT-CO2	
8030_8031del	MT-CO2	
8036del	MT-CO2	
8251A	MT-CO2 (Syn)	
8468C	MT-ATP8 (Syn)	
8616T	MT-ATP6 (L->F pos:30)	
8655C	MT-ATP6 (Syn)	
8701A	MT-ATP6 (A->T pos:59)	
9540T	MT-CO3 (Syn)	
9612del	MT-CO3	
9764del	MT-CO3	
9947A	MT-CO3 (Syn)	
9984_9985del	MT-CO3	
10034C	MT-TG	
10163T	MT-ND3 (Syn)	

10165T	MT-ND3 (P->L pos:36)	
10169T	MT-ND3 (Syn)	
10238C	MT-ND3 (Syn)	
10664C	MT-ND4L (Syn)	
10688G	MT-ND4L (Syn)	
10769del	MT-ND4	
10810T	MT-ND4 (Syn)	
10873T	MT-ND4 (Syn)	
11119T	MT-ND4 (Syn)	
11145T	MT-ND4 (T->I pos:129)	
11146T	MT-ND4 (Syn)	
11185_11186del	MT-ND4	
11552del	MT-ND4	
11692del	MT-ND4	
11704del	MT-ND4	
11914G	MT-ND4 (Syn)	
12322del	MT-TL2	
12412del	MT-ND5	
12501A	MT-ND5 (Syn)	
12653del	MT-ND5	
13105A	MT-ND5 (V->I pos:257)	
13276A	MT-ND5 (V->M pos:314)	
13559_13560del	MT-ND5	
13587del	MT-ND5	
13650C	MT-ND5 (Syn)	
13719_13720del	MT-ND5	
13725del	MT-ND5	
13742del	MT-ND5	
13780G	MT-ND5 (I->V pos:482)	
13977_13978del	MT-ND5	
14663del	MT-ND6	
15162del	MT-CYB	
15166del	MT-CYB	
15177del	MT-CYB	
15924G	MT-TT	LIMM (Point Mutation - Non Pathogenic)
16118del	MT-DLOOP	
16172C	MT-DLOOP	
16187C	MT-DLOOP	
16189T	MT-DLOOP	
16211del	MT-DLOOP	
16230A	MT-DLOOP	
16265C	MT-DLOOP	
16278C	MT-DLOOP	
16467del	MT-DLOOP	
16551_16569del	MT-DLOOP	

J: ASM018

SNP	Loci	Diseases
146T	MT-DLOOP	
152T	MT-DLOOP	
247G	MT-DLOOP	
709A	MT-RNR1	
769G	MT-RNR1	
825T	MT-RNR1	
1018G	MT-RNR1	
1226T	MT-RNR1	
1445del	MT-RNR1	
1888A	MT-RNR2	
1894del	MT-RNR2	
1899del	MT-RNR2	
2758G	MT-RNR2	
2885T	MT-RNR2	
3359del	MT-ND1	
3594C	MT-ND1 (Syn)	
4058_4060del	MT-ND1	
4104A	MT-ND1 (Syn)	
4216C	MT-ND1 (Y->H pos:304)	LHON/Insulin Resistance (P.M. - haplogroup J/T marker)
4312C	MT-TI	
4399del	MT-TQ	
4412del	MT-TM	
4709T	MT-ND2 (Syn)	
4769A	MT-ND2 (Syn)	SZ-associated (Reported)
4917G	MT-ND2 (N->D pos:150)	LHON/Insulin Resistance/AMD/NRTI-PN (Reported; haplogroup T marker)
5060T	MT-ND2 (Syn)	
5277C	MT-ND2 (F->L pos:270)	
5424A	MT-ND2 (H->N pos:319)	
5426C	MT-ND2 (Syn)	
5549del	MT-TW	
6489A	MT-CO1 (L->I pos:196)	Therapy-Resistant Epilepsy (Reported)
7028C	MT-CO1 (Syn)	
7146A	MT-CO1 (A->T pos:415)	
7256C	MT-CO1 (Syn)	
7326del	MT-CO1	
7521G	MT-TD	
7793del	MT-CO2	
7967T	MT-CO2 (Syn)	
8255A	MT-CO2 (V->M pos:224)	
8468C	MT-ATP8 (Syn)	
8619_8620del	MT-ATP6	
8655C	MT-ATP6 (Syn)	
8697A	MT-ATP6 (Syn)	
8701A	MT-ATP6 (A->T pos:59)	
9451del	MT-CO3	
9540T	MT-CO3 (Syn)	
10398A	MT-ND3 (A->T pos:114)	Invasive Breast Cancer risk factor; AD; PD; BD lithium response;

		Type 2 DM (Reported; haplogroup HNTUVWXK2 marker)
10463C	MT-TR	
10664C	MT-ND4L (Syn)	
10688G	MT-ND4L (Syn)	
10810T	MT-ND4 (Syn)	
10873T	MT-ND4 (Syn)	
10915T	MT-ND4 (Syn)	
11251G	MT-ND4 (Syn)	
11812G	MT-ND4 (Syn)	
11914G	MT-ND4 (Syn)	
11979_11980del	MT-ND4	
12705C	MT-ND5 (Syn)	
13105A	MT-ND5 (V->I pos:257)	
13276A	MT-ND5 (V->M pos:314)	
13338del	MT-ND5	
13506C	MT-ND5 (Syn)	
13650C	MT-ND5 (Syn)	
14233G	MT-ND6 (Syn)	
14323A	MT-ND6 (Y->N pos:59)	
14905A	MT-CYB (Syn)	
14989del	MT-CYB	
15002del	MT-CYB	
15043A	MT-CYB (Syn)	MDD-associated (Reported)
15257del	MT-CYB	
15452A	MT-CYB (L->I pos:236)	
15607G	MT-CYB (Syn)	
15928A	MT-TT	Multiple Sclerosis/idiopathic repeat miscarriage/AD protection (P.M./possible helper mutation)
16126C	MT-DLOOP	
16129G	MT-DLOOP	
16187C	MT-DLOOP	
16189T	MT-DLOOP	
16223C	MT-DLOOP	
16230A	MT-DLOOP	
16278C	MT-DLOOP	
16294T	MT-DLOOP	
16296T	MT-DLOOP	
16298C	MT-DLOOP	
16311T	MT-DLOOP	
16384del	MT-DLOOP	
16567_16569del	MT-DLOOP	

Appendix 4: Quantitative PCR Results

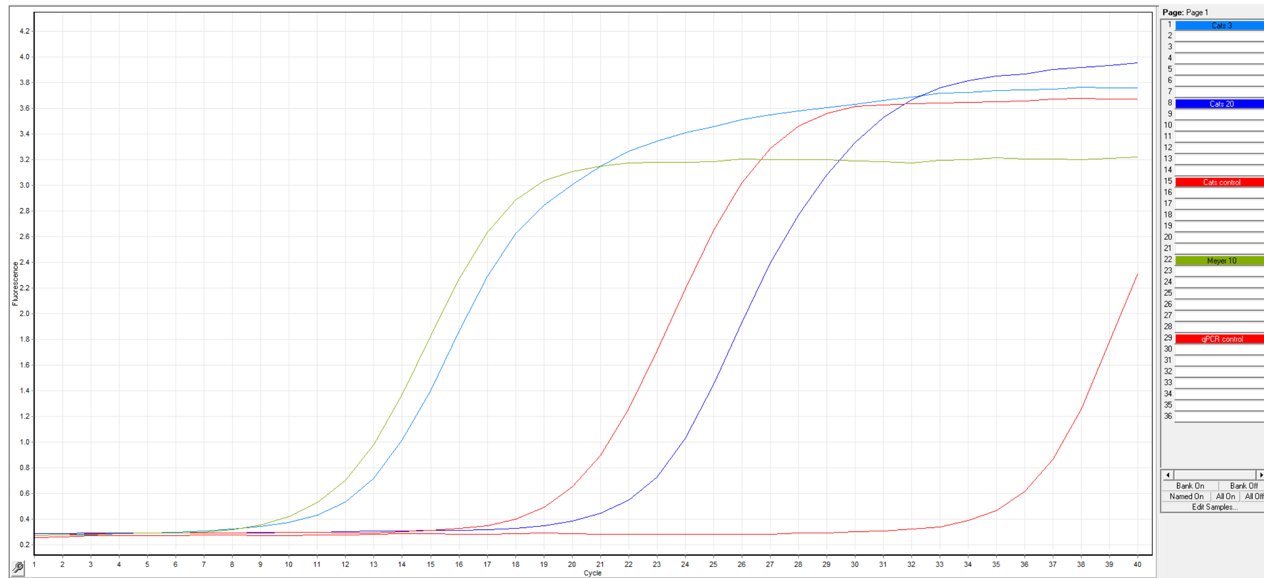


Figure A11: qPCR graph of Poulton DNA library samples by CATS method: sample 3 (light blue) and 10 (dark blue) and Meyer method: sample 10 (green), and CATS and qPCR controls (both red)

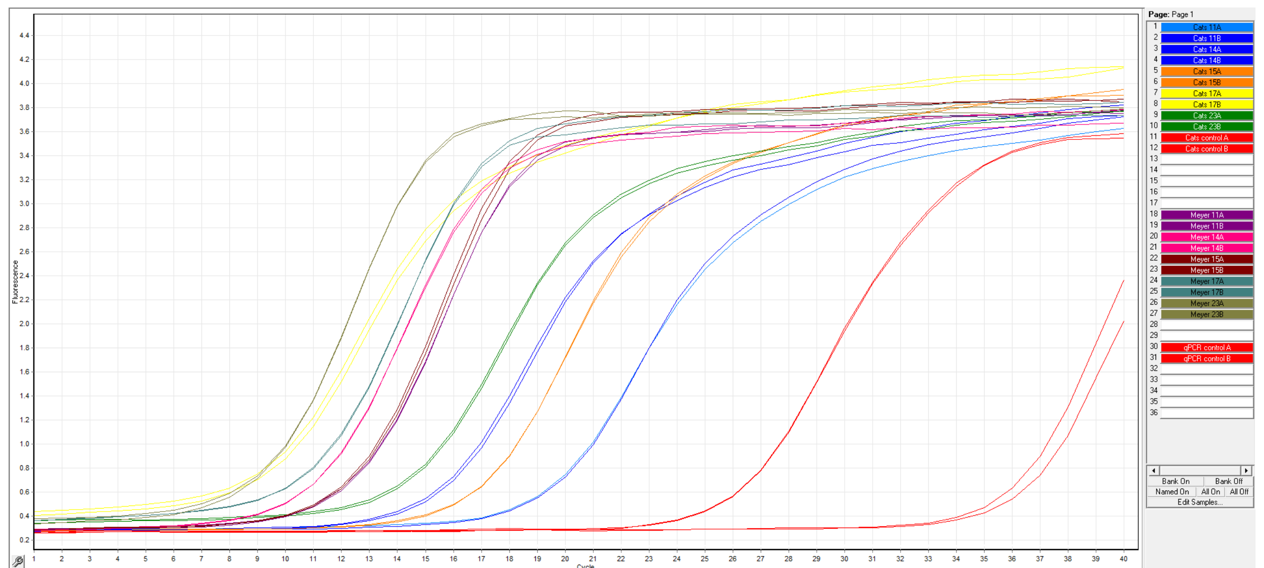


Figure A12: qPCR analysis of the CATS and Meyer library method on Poulton samples: 11 (light blue[a]/purple[b]), 14 (dark blue[a]/pink[b]),15 (orange[a]/brown[b]), 17 (yellow[a]/turquoise[b]) and 23 (green[a]/khaki green[b]) and Meyer library method on Poulton samples. (Where [a]=CATS and [b]=Meyer)

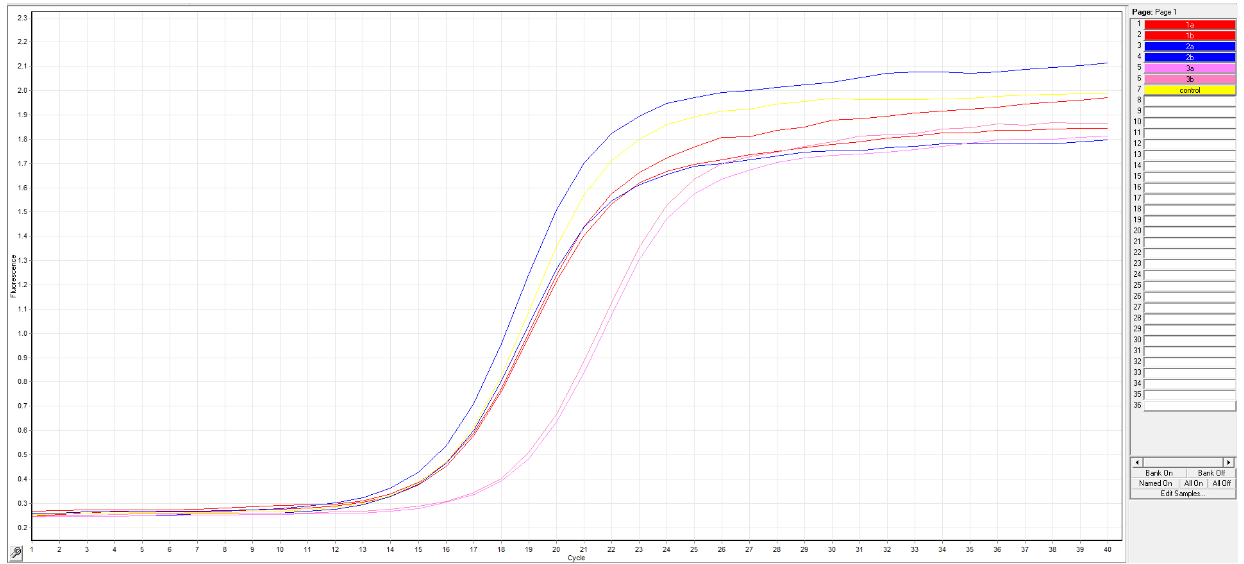


Figure A13: qPCR analysis of three CATS library methods of differing incubation times on diluted Poulton sample 23: 1 (PNK 5 minutes and dATP 15 minutes, red), 2 (PNK 10 minutes and dATP 30 minutes, blue), 3 (PNK 15 minutes and dATP 30 minutes, pink) and control (yellow)

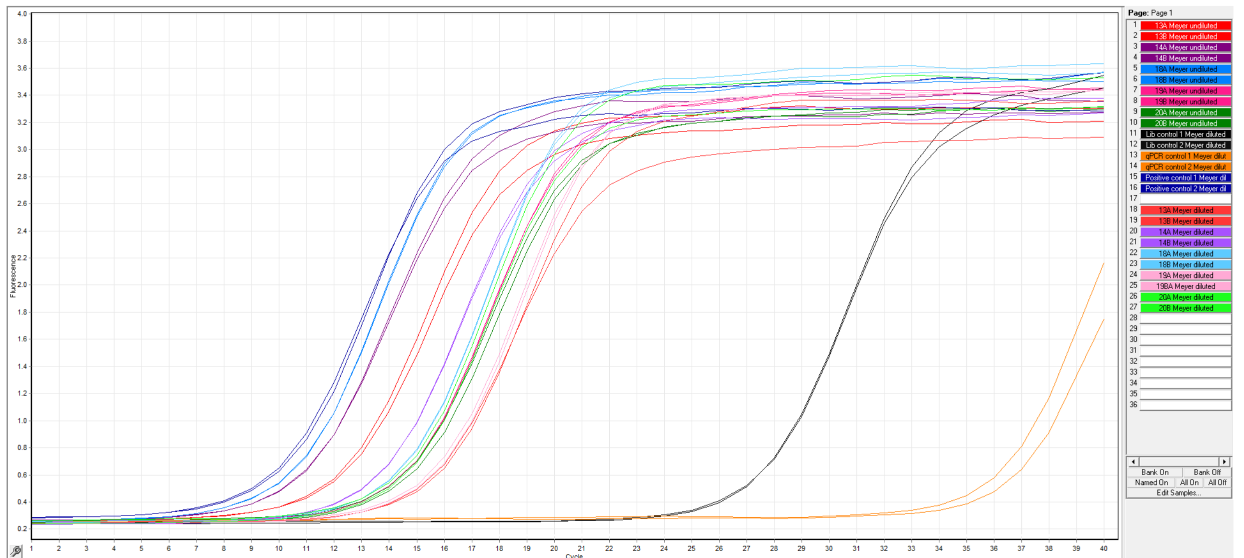


Figure A14: qPCR analysis of lower concentrations undiluted and diluted Poulton Meyer library samples: 13 (light red[a]/dark red[b]), 14 (dark purple[a]/light purple[b]), 18 (dark blue[a]/light blue[b]), 19 (dark pink[a]/light pink[b]), 20 (dark green[a]/light green[b]), qPCR controls (orange) and positive controls (darkest blue). Where ([a]=undiluted and [b]= diluted)

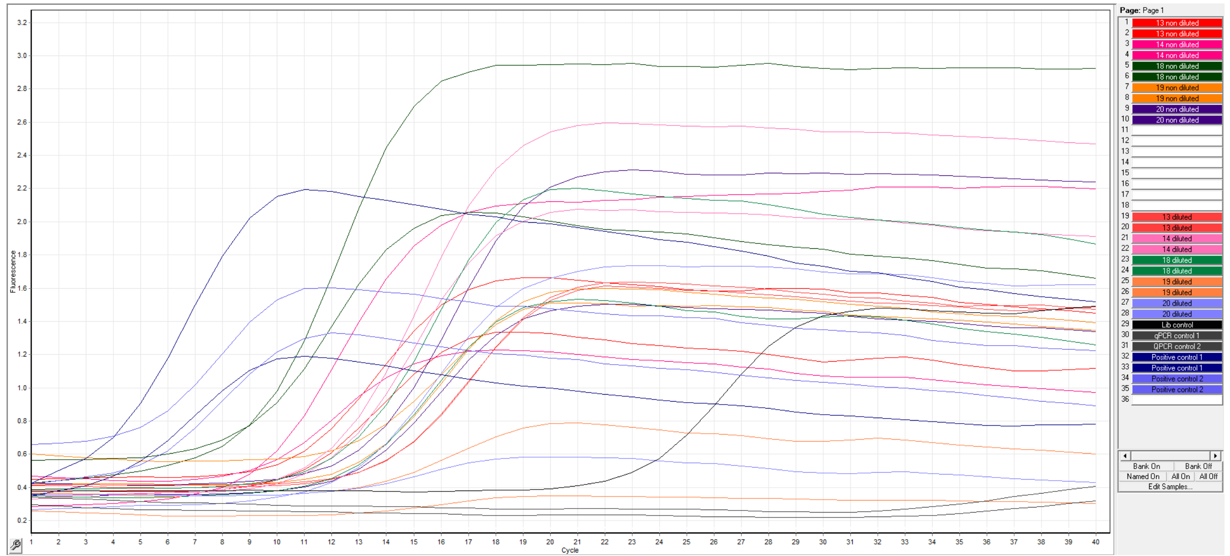
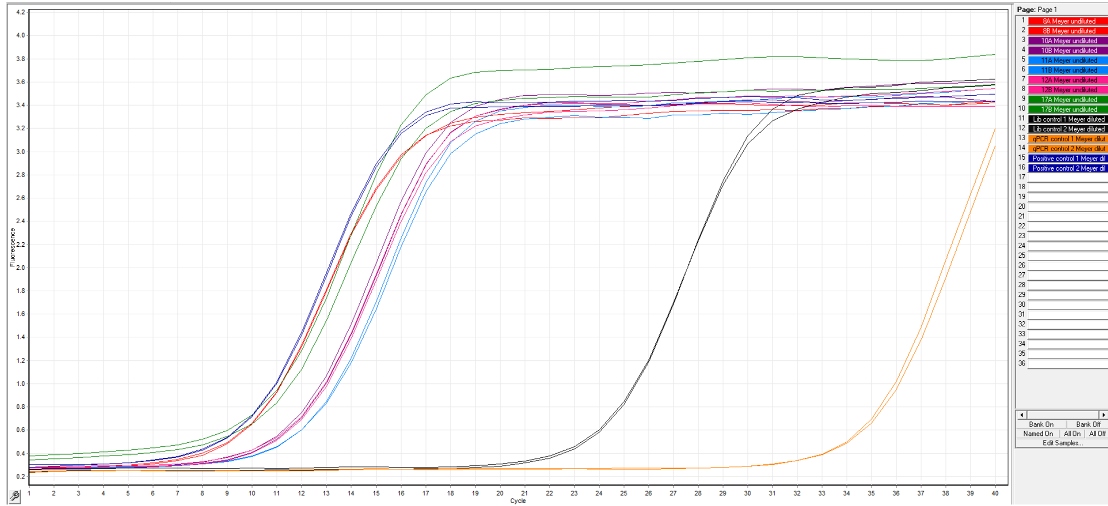


Figure A15: qPCR analysis of lower concentrations undiluted and diluted Poulton Meyer library samples: 13 (light red[a]/dark red[b]), 14 (dark pink[a]/light pink[b]), 18 (dark green[a]/light green[b]), 19 (orange[a and b]), 20 (dark purple[a]/light purple[b]), library control (black), qPCR controls (grey), positive control 1 (darkest blue) and positive control 2 (medium purple). Where ([a]=undiluted and [b]= diluted)

A



B

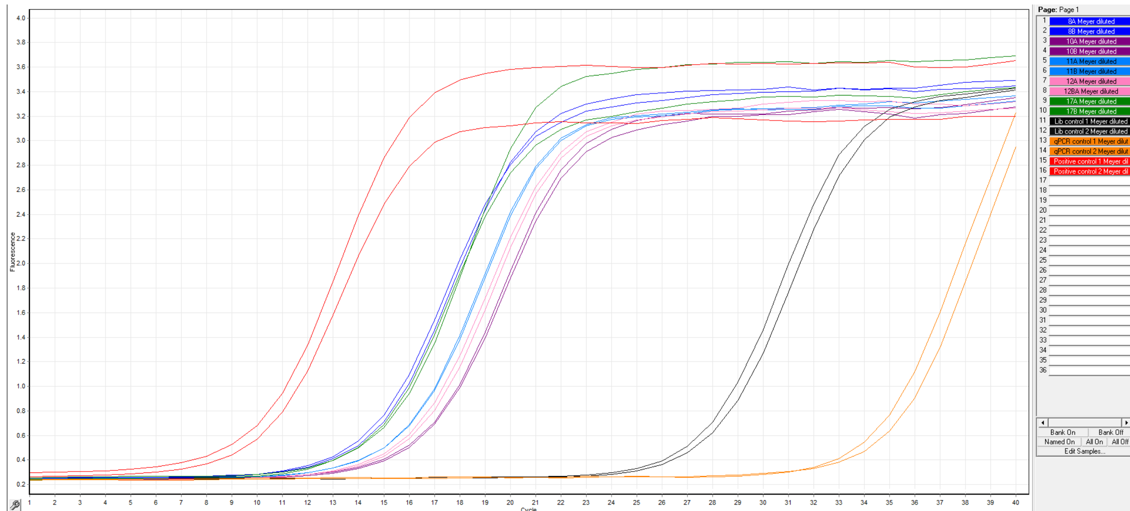


Figure A16: qPCR analysis of higher concentrations A) undiluted and B) diluted Poulton Meyer library samples: 8 (red[a]/dark blue[b]), 10 (dark purple[a and b]), 11 (light blue[a and b]), 12 (dark pink[a]/light pink[b]), 17 (dark green[a and b]), library controls (black), qPCR controls (orange) and positive controls (A) darkest blue and B) red). Where ([a]=undiluted and [b]= diluted)

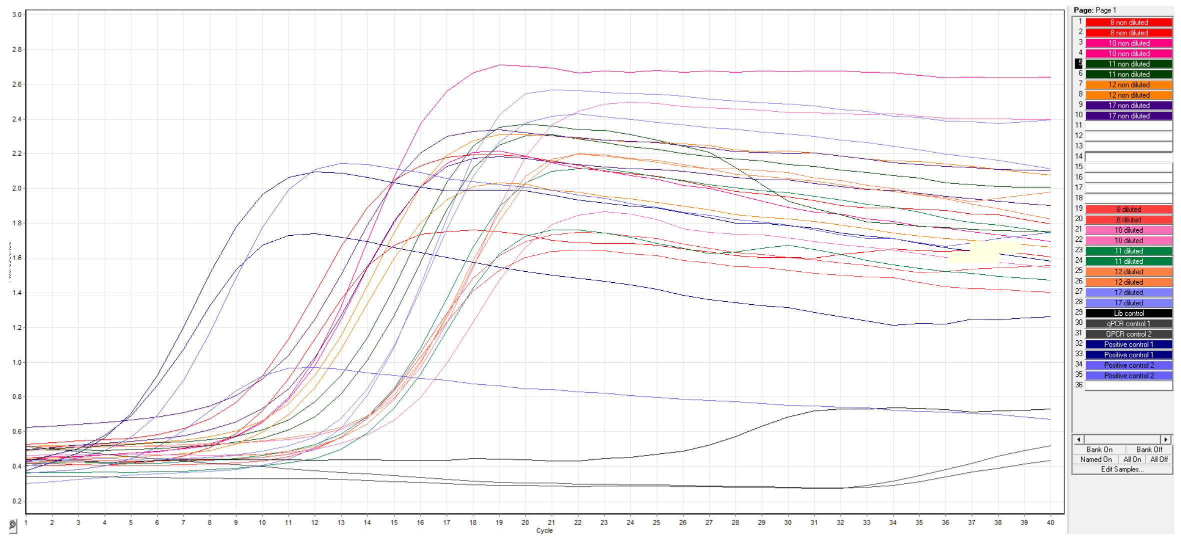


Figure A17: qPCR analysis of higher concentrations undiluted and diluted Poulton Meyer library samples: 8 (red[a and b]), 10 (dark pink[a]/light pink[b]), 11 (dark green[a]/light green[b]), 12 (orange[a and b]), 17 (dark purple[a]/light purple[b]), library control (black), qPCR controls (grey), positive control 1 (darkest blue) and positive control 2 (medium purple). Where ([a]=undiluted and [b]= diluted)

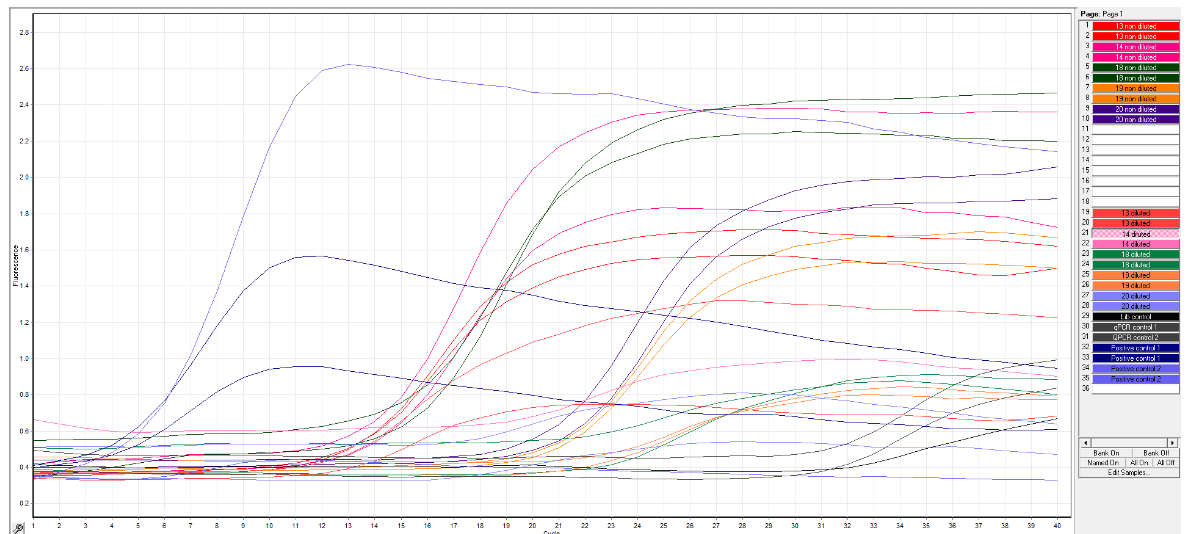


Figure A18: qPCR analysis of lower concentrations undiluted and diluted Poulton CATS library samples: 13 (light red[a]/dark red[b]), 14 (dark pink[a]/light pink[b]), 18 (dark green[a]/light green[b]), 19 (orange[a and b]), 20 (dark purple[a]/light purple[b]), library control (black), qPCR controls (grey), positive control 1 (darkest blue) and positive control 2 (medium purple). Where ([a]=undiluted and [b]= diluted)

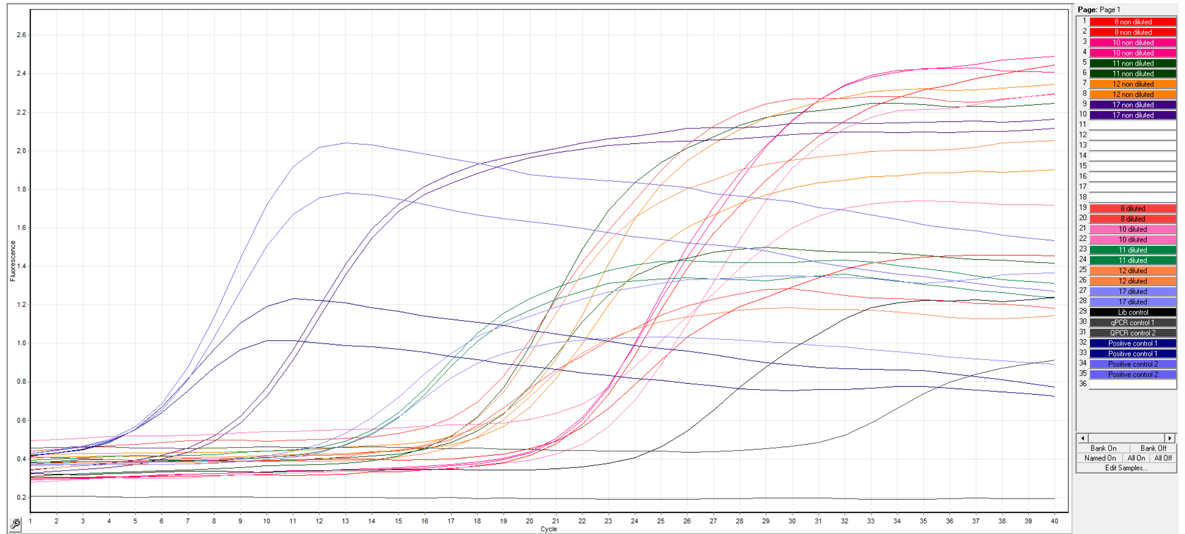


Figure A19: qPCR analysis of higher concentrations undiluted and diluted Poulton CATS library samples: 8 (light red[a]/dark red[b]), 10 (dark pink[a]/light pink[b]), 11 (dark green[a]/light green[b]), 12 (orange[a and b]), 17 (dark purple[a]/light purple[b]), library control (black), qPCR controls (grey), positive control 1 (darkest blue) and positive control 2 (medium purple). Where ([a]=undiluted and [b]= diluted)

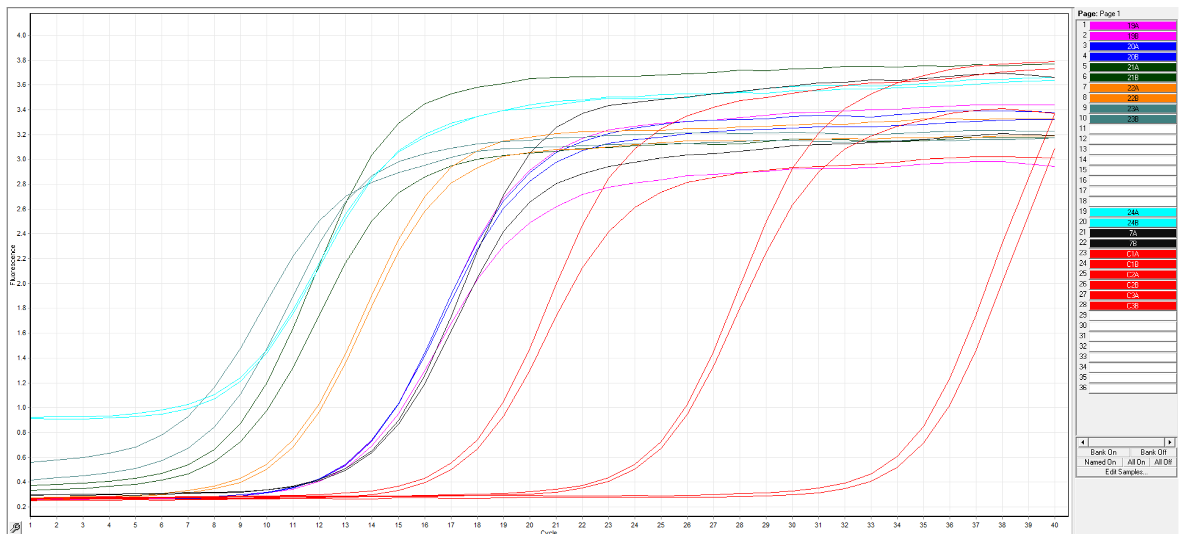
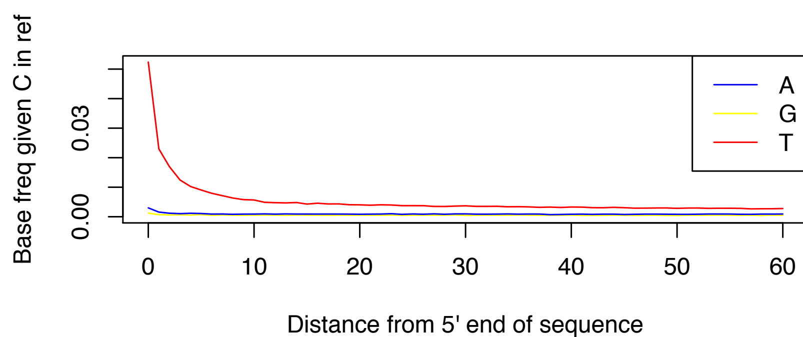
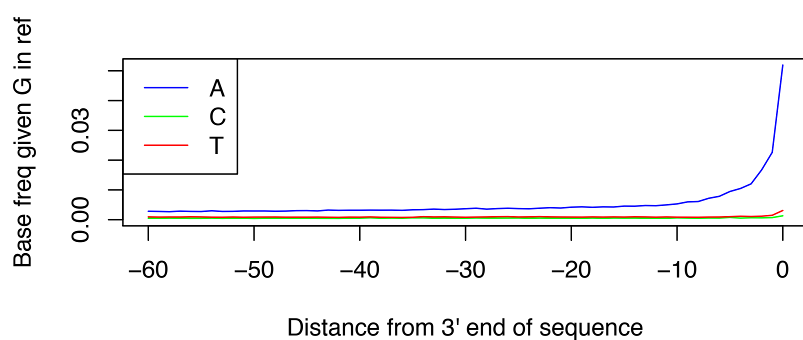


Figure A20: qPCR analysis of Meyer libraries of Poulton individuals: 19 (purple), 20 (blue), 21 (dark green), 22 (orange), 23 (turquoise), 24 (light blue), 7 (black), controls (red)

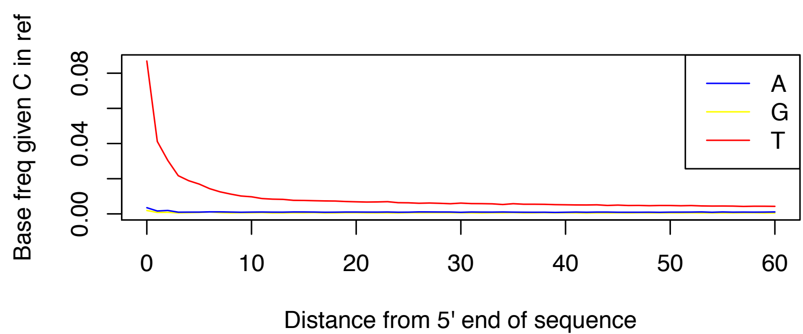
Appendix 5: Mitochondrial DNA damage plots of Poulton

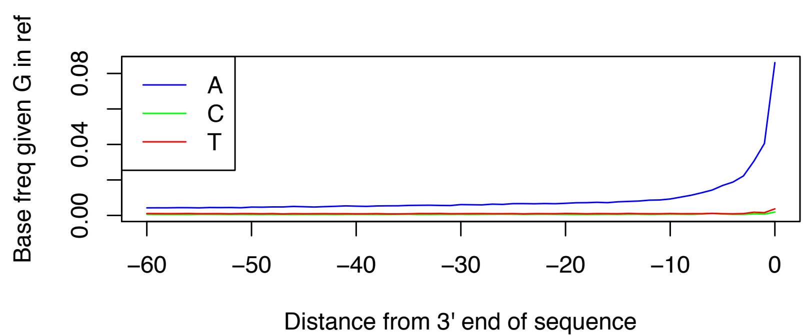
A



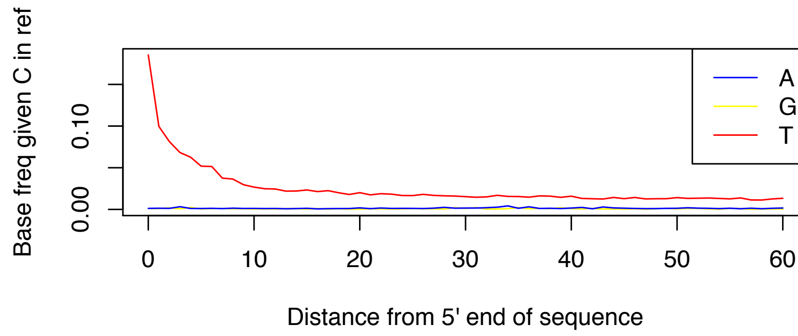


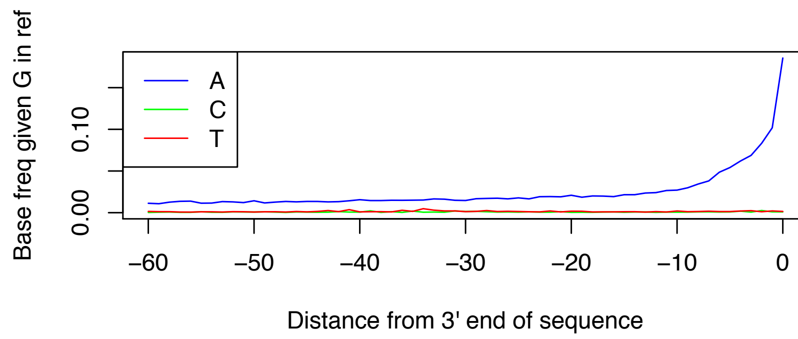
B



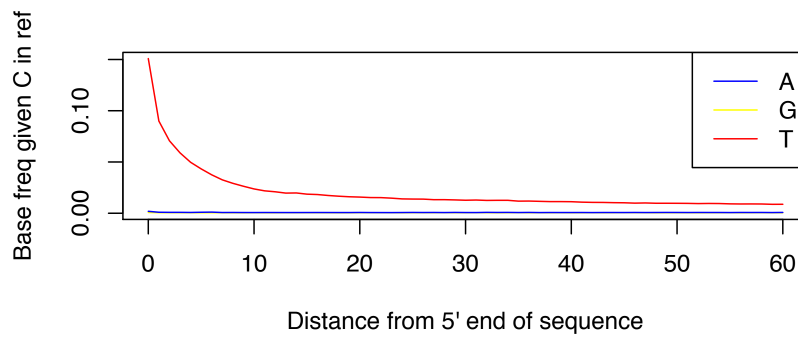


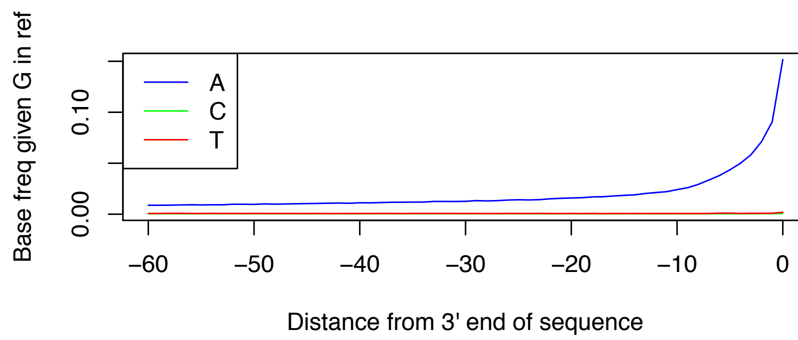
C



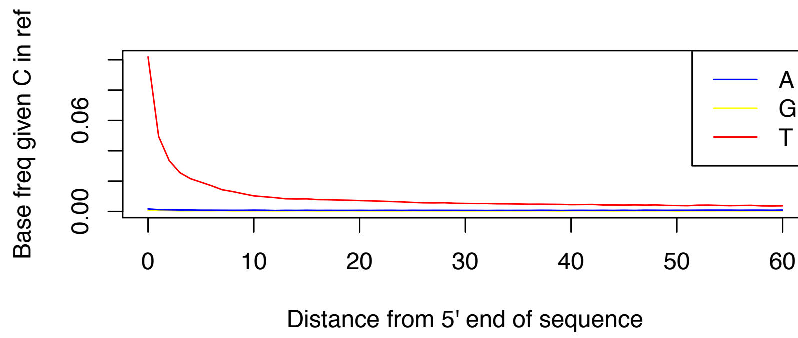


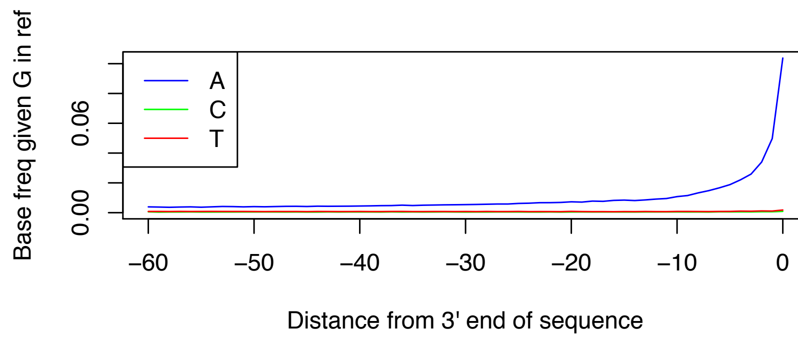
D



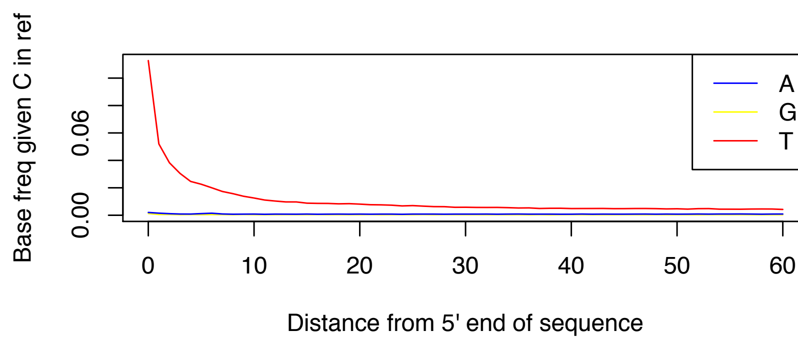


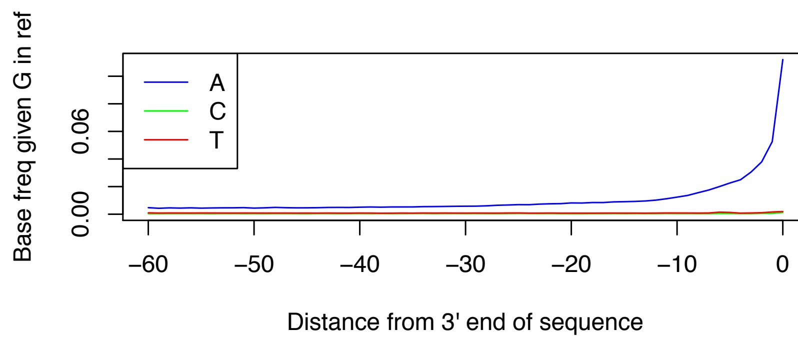
E



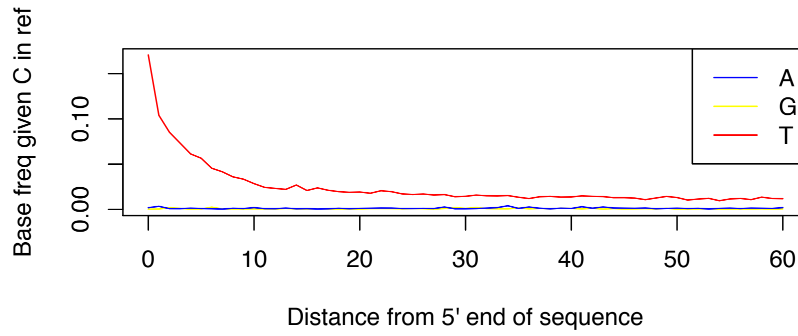


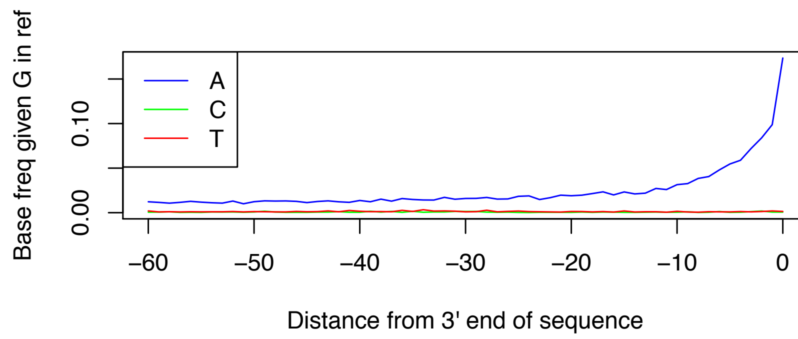
F



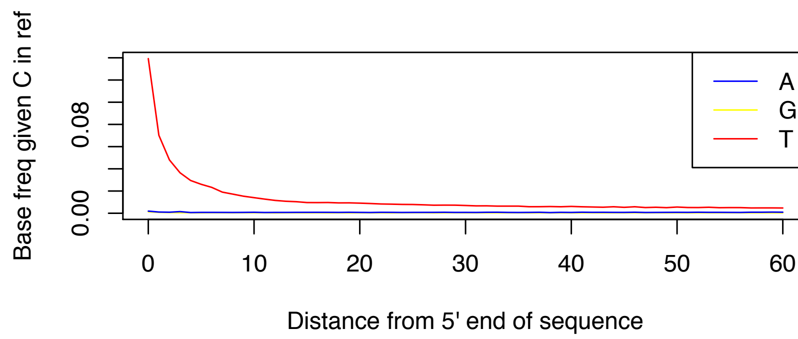


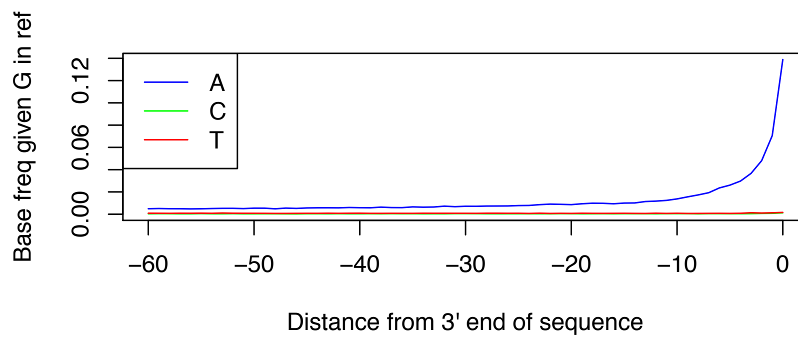
G





H





I

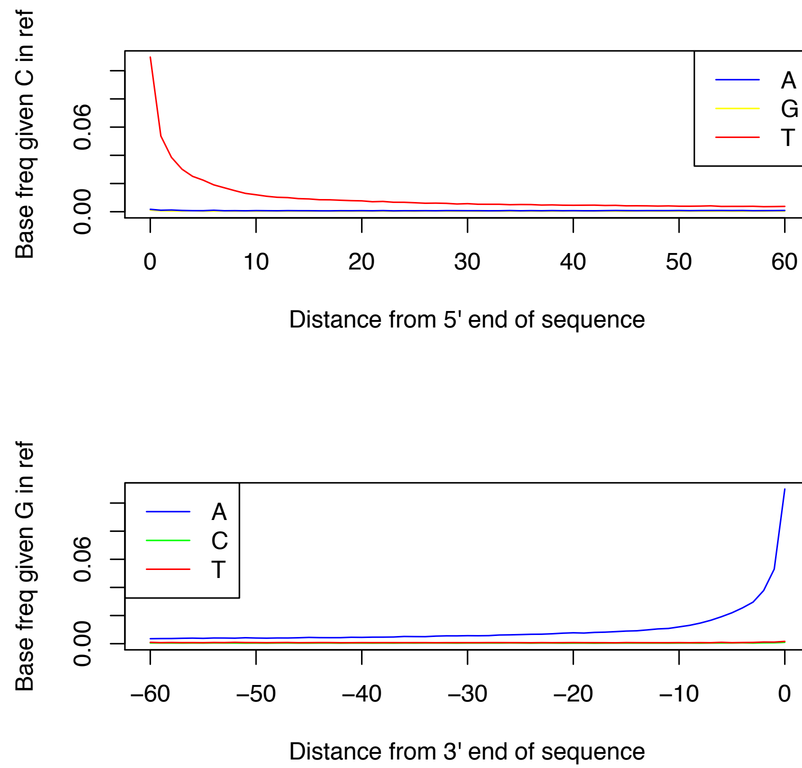


Figure A21: Mitochondrial DNA damage plots of Poulton individuals: A: ASM008, B: ASM010, C: ASM011, D: ASM012 E: ASMO013 F: ASM014 G: ASM015 H: ASM017 I: ASM018