Archaeological Genetics - Approaching Human History through DNA Analysis

EVANGELIA DASKALAKI
Abstract


There are a variety of archaeological questions, which are difficult to assess by traditional archaeological methods. Similarly, there are genetic and population genetic questions about human evolution and migration that are difficult to assess by studying modern day genetic variation. Archaeological genetics can directly study the archaeological remains, allowing human history to be explored by means of genetics, and genetics to be expanded into historical and pre-historical times. Examples of archaeological questions that can be resolved by genetics are determining biological sex on archaeological remains and exploring the kinship or groups buried in close proximity. Another example is one of the most important events in human prehistory – the transition from a hunter-gatherer lifestyle to farming - was driven through the diffusion of ideas or with migrating farmers. Molecular genetics has the potential to contribute in answering all these questions as well as others of similar nature. However, it is essential that the pitfalls of ancient DNA, namely fragmentation, damage and contamination are handled during data collection and data analysis.

Analyses of ancient DNA presented in this thesis are based on both mitochondrial DNA and nuclear DNA through the study of single nuclear polymorphisms (SNPs). I used pyrosequencing assays in order to identify the biological sex of archaeological remains as well as verifying if fragmented remains were human or from animal sources. I used a clonal assay approach in order to retrieve sequences for the HVRI of a small family-like burial constellation from the Viking age. By the use of low coverage shotgun sequencing I retrieved sequence data from 13 crew members from the 17th century Swedish man-of-war Kronan. This data was used to determine the ancestry of the crew, which in some cases was speculated to be of non-Scandinavian or non-European origin. However, I demonstrate that all individuals were of European ancestry. Finally, I retrieved sequence data from a Neolithic farmer from the Iberian Peninsula, which added one more facet of information in exploring the Neolithization process of Europe. The Neolithic Iberian individual was genetically similar to Scandinavian Neolithic farmers, indicating that the genetic variation of prehistoric Europe correlated with subsistence mode rather than with geography.

Keywords: ancient DNA, pyrosequencing, molecular genetics, aDNA, neolithization, evolutionary genetics, mtDNA, viking age, archaeological genetics

Evangelia Daskalaki, Department of Ecology and Genetics, Evolutionary Biology, Norbyväg 18 D, Uppsala University, SE-75236 Uppsala, Sweden.

© Evangelia Daskalaki 2014

ISSN 1651-6214
ISBN 978-91-554-8816-1
urn:nbn:se:uu:diva-211156 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-211156)
“Learn from yesterday, live for today, hope for tomorrow. The important thing is to not stop questioning.”

- Albert Einstein
List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


IV  Sverrisdottir, O.O.*, **Daskalaki, E.***, Skoglund, P.*; Valdiosera, C.E., Carretero, J.M., Ferreras, J.L., Jakobsson, M., Götherström, A. A late Neolithic Iberian farmer exhibits genetic affinity to Neolithic Scandinavian farmers and a Bronze Age central European farmer. *Manuscript*

*These authors contributed equally to the study

Reprints were made with permission from the respective publishers.
Artikel II: En släktskapsstudie från ett vikingatida gravfält. .........................45
Artikel III: En genomisk analys av biogeografiskt ursprung, utförd på delar av besättningen från flaggskeppet Kronan..........................................................46
Artikel IV: En sen-Neolitisk jordbrukare från Portalón, nordöstra Spanien, uppvisar genetiskt släktskap med andra jordbrukare. ........................................47
Acknowledgements.......................................................................................................49
References..................................................................................................................52
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>HVR</td>
<td>hyper-variable region</td>
</tr>
<tr>
<td>AD</td>
<td>anno domine</td>
</tr>
<tr>
<td>aDNA</td>
<td>ancient DNA</td>
</tr>
<tr>
<td>BP</td>
<td>before present</td>
</tr>
<tr>
<td>bp</td>
<td>base-pair</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>GLP</td>
<td>good laboratory practice</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>LHV</td>
<td>London human verification (assay name acronym)</td>
</tr>
<tr>
<td>MRCA</td>
<td>most recent common ancestor</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>mtDNA</td>
<td>mitochondrial DNA</td>
</tr>
<tr>
<td>nDNA</td>
<td>nuclear DNA</td>
</tr>
<tr>
<td>NGS</td>
<td>next generation sequencing</td>
</tr>
<tr>
<td>PCA</td>
<td>principal component analysis</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PMD</td>
<td>post mortem damage</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>SNP</td>
<td>single-nucleotide polymorphism</td>
</tr>
<tr>
<td>ssDNA</td>
<td>single-stranded DNA</td>
</tr>
<tr>
<td>ssstDNA</td>
<td>single-stranded template DNA</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>U</td>
<td>unit (used when measuring enzyme concentrations - U/μl)</td>
</tr>
<tr>
<td>UNG</td>
<td>uracil N-glycosylase</td>
</tr>
</tbody>
</table>
Introduction

The field of ancient DNA (aDNA) is a young area of research that has developed rapidly over the last decade. The publication of mtDNA sequences from a 150-year-old quagga (Higuchi et al. 1984) and a partial mtDNA sequence from a 2430-year-old Egyptian mummy (Pääbo 1985) in the mid-80s marked the starting point of aDNA research. It subsequently developed into an active and important field of research some 10 years ago, as illustrated by the complete sequencing of mitochondrial genomes from moas (Cooper et al. 2001; Haddrath & Baker 2001). Today, it is still a difficult task to work with ancient animal and human remains; however, it is no longer impossible to generate complete human and hominid genomes from ancient remains (Green et al. 2010; Meyer et al. 2012; Rasmussen et al. 2010).

Archaeological genetics or aDNA is essentially any type of genetic study performed on archaeological or historical material. This material can be any type of preserved tissue, with bones and teeth being common material. Other types of material available can be hair, feathers, hoofs, nails, skin (parchment), seeds, pollen, frozen tissue or mummified tissue. Working with these types of specimens differ from working with modern samples since old DNA is generally of poor quality (Pääbo et al. 2004a; Rizzi et al. 2012).

The development of the polymerase chain reaction (PCR) initiated an enthusiastic period as evidenced by early reports of recovered DNA from plant fossils (Golenberg et al. 1990; Soltis et al. 1992), dinosaurs (Woodward et al. 1994) and insects trapped in amber (Cano et al. 1993; DeSalle et al. 1992). However, the majority of these early studies are now viewed with caution and considered to be products of contamination or artefacts (Austin et al. 1997; Hebsgaard et al. 2005; Penney et al. 2013; Pääbo 1989; Pääbo et al. 2004b).

Archaeological genetics has a variety of challenges, methodologically speaking, with issues from DNA preservation and contamination to access to specimens. This has forced the field to become highly method oriented, in order to retrieve as reliable and abundant genetic information as possible from the available specimens. To this end, guidelines for evaluating authenticity have been developed, DNA extraction methods and substrates have been optimized, and new technologies – such as “next generation sequencing” (NGS) – have been embraced and adapted to the specific demands of aDNA. By this it has become possible to address problems that are not well resolved using archaeology or contemporary genetic sources (Willerslev & Cooper 2005). For example, the use of modern population data to reconstruct past demographic events may be associated with the risk of earlier demographic events.
being obscured by more recent ones (Barnes et al. 2007; Leonard et al. 2002). In contrast, the use of models and the coalescent approach with genetic samples collected over a variety of geographical and temporal scales has proved effective in order to understand past population dynamics and evolutionary processes (Ramakrishnan & Hadly 2009).

The field of aDNA has also been a huge asset in studies of the domestication of plants and animals. Significant examples include work by Thalmann et al. (2013) where the authors’ findings support the conclusion that the mitochondrial legacy of dogs derives from wolves of European origin, thus contradicting the leading theories that suggest dogs were domesticated either in the Middle East or in East Asia. Other examples include the work of Larson et al. (2012) that points to an early (~15,000 years ago) domestication of dogs and by Leonard et al. (2002) that concluded that native American dogs originated from old world lineages of dogs that accompanied late Pleistocene humans across the Bering strait. Other examples are the studies on cattle domestication by Anderung et al. (2005), where the effects of contacts between the Iberian Peninsula and Africa on cattle domestication were explored. Furthermore, Svensson et al. (2007) and Telldahl et al. (2011) utilized temporal series of ancient cattle samples to trace genetic change over time, with focus on selective breeding during the domestication process. In the analyses by Larson et al. (2007; 2005), the domestication of pigs were studied through a combination of modern and ancient samples of domestic and wild pigs. The authors concluded that multiple domestication events across Eurasia probably occurred and that the European wild boar is the principal source of European domestic pigs.

Another broad area where studies of aDNA have been valuable is the study of the history and dynamics of natural populations, and evolutionary processes occurring in these. Some examples of such studies are analyses of mitochondrial genetic variability of European rabbits (Hardy et al. 1995), the effect of the fur trade on the genetic variability of sea otters (Larson et al. 2002), and genetic diversity of the grizzlies in Yellowstone (Miller & Waits 2003). Studies of aDNA for endangered species, like the arctic fox, have also helped us to understand how populations respond to climate change (Dalen et al. 2007) and (Teacher et al. 2011).

Human population history and evolution have also been investigated using aDNA techniques. For instance, the Neolithization process in Europe has been a focus area. Most aDNA studies on this subject have used mtDNA and been based on samples from Central Europe, Northern Europe and from the western Mediterranean (Iberia) (Bramanti et al. 2009; Hervella et al. 2012; Lee et al. 2012; Malmstrom et al. 2009). Some studies have also used Y chromosomal haplotypes (Haak et al. 2010; Lacan et al. 2011) or autosomal sequence data (Ermini et al. 2008; Malmstrom et al. 2010; Sánchez-Quinto et al. 2012; Skoglund et al. 2012). Ancient DNA-based work published so far provides a heterogeneous picture of the Neolithic process in Europe. Mitochondrial DNA data from central/northern Europe are consistent with a model of demic diffusion, where farmers formed enclaves surrounded by indigenous hunter-gatherers (Lacan et al. 2013). These pioneers could subsequently have mixed with local populations through contacts around the area of initial settlement,
implying that the Neolithic spread in central Europe was accomplished by local hunter-gatherers with varying degrees of input from the first farmers (Lacan et al. 2013) Studies from finds in Scandinavia show a lack of complete continuity between Neolithic hunter-gatherers and Neolithic farmers (as well as modern-day Scandinavian populations), demonstrating the occurrence of Neolithic, or post-Neolithic gene flow in this area (Bramanti et al. 2009; Malmstrom et al. 2009; Skoglund et al. 2012). Processes in the Iberian region are not as easily described as in the case for central and northern Europe. The composition of mitochondrial haplogroups is distinct from that found in central Europe and Spain (Gamba et al. 2012; Lacan et al. 2013), indicating that the Neolithization process in this region differed from other parts of Europe. However, this was recently contradicted by the findings of Sánchez-Quinto et al. (2012) who analyzed two Mesolithic individuals and failed to find any major deviations from the migratory pattern noted in other parts of Europe, if correct, this would suggest that the Iberian Peninsula was part of the greater European gene-pool in Mesolithic times.

Another area where aDNA has contributed to our understanding of human evolution is the study of archaic humans and their relation to modern humans. Prominent studies here are those dealing with Neanderthals and Denisovans. Landmark studies in this area include the publication of the Neanderthal mitochondrial genome (Green et al. 2008) and the draft nuclear genome (Green et al. 2010). Similarly, for the Denisovans the complete mitochondrial genome was published by Reich et al. (2010) and a draft of the nuclear genome by Meyer et al. (2012). It was initially concluded that Neanderthals were a distinct population, although a minor contribution to present-day ancestry could not be excluded (Noonan et al. 2006). However, later studies demonstrated that while the majority of the ancestry of present-day humans appears to be derived from African populations in the last ~100 kya, a small proportion of genetic material traces its ancestry to the archaic populations, Neanderthals as well as Denisovans, that came to Eurasia much earlier (Krause et al. 2010; Lalueza-Fox & Gilbert 2011; Lowery et al. 2013; Reich et al. 2010; Skoglund & Jakobsson 2011).

Additionally, many human studies have taken a more archaeological/historical perspective, focusing on the identification of historical individuals like the founder of Stockholm (Malmström et al. 2012), the Romanov family members (Gill et al. 1994) and saints such as Lucas the Evangelist (Vernesi et al. 2001) and Birgitta of Sweden (Nilsson et al. 2010). Other examples are the exploration of kinship patterns in burial grounds and sites (Baca et al. 2012; Bouwman et al. 2008), and sex determinations of skeletal remains (Daskalaki et al. 2011; Skoglund et al. 2013; Stone et al. 1996).

For a more detailed background and history of aDNA see for example review articles by Hofreiter et al. (2001b), Lalueza-Fox & Gilbert (2011), Pajimans et al. (2012) and Rizzi et al. (2012).
Ancient DNA (aDNA) and human genetics

Obtaining genetic information from ancient human remains has been a major challenge for aDNA research, for instance, due to the severe risk of contamination modern day humans working with the material. Novel ways to extract authentic DNA from human remains have therefore been developed. Prior to the advent of NGS the majority of aDNA studies focused on mtDNA, however, improved sequencing technologies have enabled broader genomic studies on the complete human genome, i.e. including nuclear DNA (nDNA).

Mitochondrial DNA (mtDNA)

The mammalian mtDNA is a circular double stranded molecule approximately 16.6 kb in length. It contains 13 respiratory chain genes, two ribosomal RNAs (rRNAs) and 22 transfer RNAs (tRNAs). It also includes a 1.2 kb non-coding region with mainly regulatory functions – named the control region or D-loop. The control region contains three hyper-variable regions (HVRs I, II and III) separated by conserved segments.

Mitochondrial DNA has some characteristics that distinguish it from other chromosomes. While each human cell has two copies of its nuclear genome, it can have a large number of copies of its mitochondrial genome. This is because each mitochondrion may contain up to 10 mtDNA molecules and each cell may contain from a few hundred to thousands of mitochondria. Mitochondrial DNA is maternally inherited from mother to offspring without recombination, so mtDNA lineages are commonly referred to as haplotypes (e.g. the sequence itself is synonymous to the haplotype). It is possible to define not only mt-haplotypes (e.g. combination of alleles/DNA sequences that are inherited together), but also haplogroups. A haplogroup is a number of haplotypes that share a common ancestor. Mitochondrial DNA also has a high mutation rate, especially the control region, compared to the mutation rate observed for nuclear DNA.

Phylogeographic analyses of mtDNA variation have played a major role in our knowledge about human evolution. Studies on mtDNA (Cann et al. 1987) corroborated the “out of Africa” theory for modern human origin, initially put forward on the basis of fossil evidence (Bräuer 1984; Stringer & Andrews 1988; White et al. 2003). Subsequent studies of global mtDNA variation provided further evidence for this theory, with the most recent common ancestor (MRCA) of human mtDNA placed in Africa 100,000–200,000 years ago (“the mitochondrial Eve”; Cann et al. in (1987)). This concept was popularized and spread to the greater public by Brian Sykes in his book “The seven daughters of Eve” (Sykes 2002), with each daughter of Eve representing one of the mitochondrial haplogroups. Studies of mtDNA has also contributed to insights into the migration patterns that shaped human populations, such as the settlement of Europe, the peopling of the New World, the colonization of the Pacific as well as New Guinea and Australia (Ho & Gilbert 2010; Pakendorf & Stoneking 2005). However, the facts that the
mitochondrion represents only one locus and has a matrilineal inheritance pose limitations to the use of mtDNA for inferring demographic information.

Nuclear DNA (nDNA)
The human nuclear genome contains approximately 3.2 billion base pairs packaged into the 46 chromosomes, 44 autosomes and the 2 sex chromosomes. The Y-chromosome, like mtDNA, is ideal for phylogeographic studies since it is paternally inherited and does not undergo recombination at meiosis. This means that analyses based on Y-chromosome haplotypes provide information about male mediated gene flow. Combining the two uni-parentally inherited markers, mtDNA and Y chromosome, can therefore reveal different patterns, reflecting evolutionary processes of female and male lineages, respectively. Autosomal markers are inherited bi-parentally and undergo recombination during meiosis. The X-chromosome is inherited bi-parentally for females (and generally undergoes recombination only in females) and uni-parentally for males, e.g. males will always transmit the X-chromosome to daughters.

Single-nucleotide polymorphisms (SNPs) are positions in the genome, which are genetically variable. SNPs originate from the naturally occurring mutational processes and new variants may be passed to offspring and rise in frequency in the population due to, for example, genetic drift or selection. The two “versions” of a segregating site are referred to as alleles. In general, most SNPs are biallelic, meaning that there are only two variants (e.g. A or G) within the population. The less common version in a population is referred to as the minor allele. Nevertheless, a SNP may have multiple alleles if a subsequent mutation hit an already segregating site.

The methodological challenges of aDNA
DNA preservation
A variety of tissues can be used for DNA extraction. Soft tissues are the first to degrade, often leaving only teeth and bones, making them the most favored ones for DNA extractions due to their abundance as an archaeological material. Teeth are good sources for ancient DNA, since their hard enamel surfaces are often well preserved and protect the DNA in the dentine from degradation. Compact bones, for example the femur, the humerus and the mandible also tend to be relatively well preserved, although DNA abundance and preservation can vary locally within a bone sample (Orlando et al. 2011). Other examples of sources for aDNA are coprolites (Gilbert et al. 2008; Jenkins et al. 2012; Poinar et al. 2003; Poinar et al. 1998), hair and leather (Bengtsson et al. 2012; Gilbert et al. 2006b; Gilbert et al. 2004; Rasmussen et al. 2010; Vuissoz et al. 2007), plant remains (Allaby et al. 1999; Kistler 2012; Palmer et al. 2012; Parducci et al. 2013), hoofs (Bengtsson et al.
antlers (Kuehn et al. 2005) or eggshells (Oskam & Bunce 2012; Oskam et al. 2010) as well as ice cores, permafrost and temperate sediments (Hansen et al. 2006; Jorgensen et al. 2012a; Jorgensen et al. 2012b; Parducci et al. 2013; Willerslev et al. 2007). Although aDNA analyses have been conducted on mummified human remains (Hawass et al. 2010; Hekkala et al. 2011; Khairat et al. 2013; Kurushima et al. 2012; Lynnerup 2007; Marchant 2011; Oh et al. 2013; Pääbo 1985), there are still some controversy regarding the credibility of such results, due to the hot environment mummies most commonly originate from (Marota et al. 2002). More credibility has been given to results from cold and stable environments, such as the Tyrolean iceman, Ötzi (Ermini et al. 2008), and Andean mummies (Baca et al. 2012), for reasons outlined below.

A possible assumption for DNA preservation is that the DNA survival correlates negatively with the age of the material. Unfortunately this is a fallacy, age and DNA preservation might, but does not necessarily correlate (Hoss et al. 1996; Poinar et al. 1996; Pääbo 1989). Instead, the surrounding macro- and microenvironment seems to be more important determinants for DNA preservation (Campos et al. 2012; Dabney et al. 2013; Gilbert et al. 2006c; Gilbert et al. 2005b; Hagelberg et al. 1991). One of the initial environmental factors affecting DNA survival post mortem is the extent of microbial activity (Burger et al. 1999; Pääbo 1989). Yet another early environmental factor affecting DNA is the speed of tissue desiccation, with rapid desiccation being favorable for a good DNA survival. In the longer perspective, temperature and humidity, as well as soil properties influence the preservation of DNA. In general cold temperature or temperature with few seasonal fluctuations has been proven favorable for DNA preservation, as has soil with high salt concentration and neutral or slightly alkaline pH (Bollongino et al. 2008; Hagelberg et al. 1991; Lindahl 1993; Tuross 1994).

The most important environmental factor for DNA preservation is temperature. More specifically the temperature a sample has been exposed to during its “lifetime”, e.g. the thermal history of a sample. Fewer temperature fluctuations and low mean annual temperatures favors DNA survival and preservation (Lindahl 1993; Smith et al. 2001; Smith et al. 2003). This correlation is well illustrated by the fact that the oldest reliable aDNA results are mainly from permafrost or cave environments. As seen in the studies performed on 400,000 year-old bear fossils found in the Atapuerca cave system in Spain (Valdiosera et al. 2007). Furthermore, ice Cores from Greenland have generated data from plants and insects that are 450,000 to 800,000 year old (Avila-Arcos et al. 2011; Jorgensen et al. 2012a; Willerslev et al. 2007). Finally, the oldest draft genome published to date is from a horse bone recovered from permafrost (approximately 560–780 thousand years BP) (Orlando et al. 2013).

Another environmental factor that may influence the quality and quantity of DNA surviving in specimen materials is their post excavation treatment (Burger et al. 1999; Pruvost et al. 2007). In general it would be favorable for specimens with potential for use in aDNA studies to be preserved according to the above-mentioned criteria: Cold, dry and non-acidic environment. From a geneticists point of view the
ideal long term storage being in freezers. Most of the archaeological material is
nevertheless stored at room temperature. At the time of sampling for aDNA studies,
the sample may have spent many years at room temperature. It has been shown that
samples stored in a museum environment for many years do manifest a lower
preservation (illustrated by their PCR amplification success rate) compared to
freshly excavated samples (Bollongino et al. 2008; Melchior et al. 2008). The full
mechanism that causes this decrease is not explored. However it is speculated that
the difference is due to the elevated temperatures in storage, or the washing
procedures employed that may alter the sample pH, and scrub of outer protective
layers of the bone rendering it more porous, and thus more prone to be affected by
environmental conditions.

When it comes to DNA preservation within bone samples a within sample
variation has been observed, suggesting that the micro-environment within the
sample affects preservation. Porous bone is less likely to have good preservation,
while compact bone exhibits better preservation. The complete mechanisms and
factors affecting the within sample DNA preservation are unknown, but the
existence of intrasample variation is well illustrated in a study by Orlando et al.
(2011) where the authors report a high variability in the preservation of DNA in
different locations of the same bone fragment, and is also observed in my own
unpublished observations from the laboratory.

In summary, there is no single factor that will determine the DNA preservation
over time. Thus, sampling for aDNA testing is a trial and error process, based on
qualified guesses.

DNA damage
The quality of DNA is of major importance for aDNA studies, since it affects the
downstream processing used in genetic studies. In living organisms, DNA achieves
high stability thanks to the enzymatic maintenance and repair systems present in
living cells. As soon as an organism dies, however, these maintenance systems start
to fail. The DNA in a post mortem tissue will be exposed to cleavage by nucleases
within the cell and to degradation by microorganisms, e.g., bacteria and fungi. The
above factors lead to a fast initial fragmentation of DNA, and to a reduction of the
total amount of DNA available in a sample. In addition, DNA will be exposed to
long-term degradation by chemical reactions and to further chemical processes
altering the DNA, such as hydrolytic cleavage and oxidative damage. These two
latter processes cause the majority of the degradation that affects the possibilities to
amplify DNA from samples, and thus to obtain sequence data from ancient material.
The damage can also cause incorrect amplification, thus affecting the accuracy of
the obtained data. Specifically, the phosphodiester bonds in the phosphate-sugar
backbone (see Figure 1) are susceptible to hydrolytic cleavage generating single-
stranded nicks. Hydrolytic damage also occurs to the glycosidic bonds between
nitrous bases and the sugar backbone (see Figure 2), resulting in abasic sites, e.g.,
sites where the sugar backbone is intact but the base is lost (Hofreiter et al. 2001b;
Abasic sites can lead to mis-incorporations of nucleotides during PCR amplification, or to premature ending of amplification, or they can undergo a chemical rearrangement that promotes occurrence of strand breakage, thus contributing to the fragmentation (Dabney et al. 2013; Hofreiter et al. 2001a; Pääbo 1989).

Figure 1: DNA sites susceptible to chemical damage

Lesions that act as physical blocks to the elongation of DNA strands by the polymerase can also limit the amount of DNA that can be amplified by PCR. Such lesion can occur due to oxidative damage induced by free radicals and UV radiation. Major sites of oxidative attack are the double bonds of both pyrimidines and purines, leading to ring fragmentation. In addition, the chemical bonds of the deoxyribose residues are susceptible to oxidation resulting in fragmentation of the sugar ring (Lindahl 1993). This issue can in some cases be counteracted by treatment with endonuclease III, an enzyme with cleavage specificity for oxidized pyrimidines. Another type of damage that physically blocks the DNA polymerase is crosslinks. Crosslinks are mainly caused by Maillard reactions. Maillard products are formed by condensation reactions between sugars and primary amino-groups in proteins and nucleic acids, and can cause interstrand crosslinks as well as pyrimidine dimers, where two adjacent thymine or uracil nucleotides form dimers; or base adducts, where an chemical component reacts with and covalently bonds to DNA. To counter these types of physical blocks, DNA extracts can be treated with N-phenacylthiazolium bromide, which breaks Maillard products. This strategy has been successfully used, in for example the DNA amplification from ground sloth coprolites (Poinar et al. 1998).

DNA fragmentation and amplification blocking has to be taken into account when designing aDNA assays, since they limit the maximum length of aDNA sequences that can be retrieved. Even with endonuclease and N-phenacylthiazolium bromide treatment, the size of amplifiable fragments usually range from around fifty
to a few hundred base pairs. Hence, when using targeted approaches the assay design has to target short amplicons and overlapping fragments.

Chemical degradation also produces altered bases that can lead to nucleotide mis-incorporation during amplification. This form of damage can be considered more severe than fragmentation and amplification blocking, which leads to “absent” data rather than errors in the DNA sequence. Miscoding lesions are primarily due to hydrolytic deamination of Cytosine to Uracil, 5-methyl-Cytosine to Thymine, Adenine to Hypoxantine and Guanine to Xanthine. The most common form of nucleotide substitutions observed in aDNA datasets are transitions, e.g. substitution of a purine for another purine or a pyrimidine for another pyrimidine. These are classified as Type 1 (A to G or T to C) and Type 2 (C to T or G to A) damages (Hansen et al. 2001). In aDNA data type 2 are predominant, with the majority caused by deamination of C to T and with a smaller fraction of G to A substitutions (Gilbert et al. 2003a; Gilbert et al. 2003b; Stiller et al. 2006). Errors caused by Uracil, the deaminated form of Cytosine, can be minimized prior to amplification, by treating the aDNA extracts with uracil-N-glycosylase (an enzyme that removes uracil) (Hofreiter et al. 2001a; Pääbo 1989).

Damage patterns can also be used as a tool to identify potentially endogenous sequences. From sequence reads generated through next generation sequencing, it has been observed that C to T damage increases toward the 5'-end, and thus also the G to A towards the 3'-end, so this type of damage pattern can be indicative endogenous sequences. Another example is the c-statistic (Helgason et al. 2007), a statistic designed to identify the authentic ancient sequence from a pool of cloned sequences (Malmström et al. 2012). As is the PMD-score (post mortem damage score) launched by Skoglund et al. (2013), which uses the total damage patterns in next-generation sequencing (NGS) generated data to distinguish endogenous ancient sequences from contamination. Another application that uses post-mortem degradation in order to authenticate sequences from NGS datasets is mapDamage 2.0 by Jonsson et al. (2013).

**Figure 2:** DNA damage that can hinder or alter the PCR amplification of aDNA.
DNA contamination

In DNA extracts from modern samples, such as fresh tissue and blood samples, practically all of the DNA molecules are endogenous if good laboratory practice (GLP) is followed. This is generally not the case when working with ancient samples, which can contain high levels of exogenous DNA, owing to microbiological degradation or absorption from the environment. The presence of exogenous DNA is a problem in archaeological genetics studies that was identified early in the history of the field (Higuchi et al. 1984; Higuchi et al. 1987; Hoss et al. 1996), and led to the early establishment of aDNA specific guidelines by the pioneers of the field (Cooper & Poinar 2000; Handt et al. 1996). Especially the publication “Ancient DNA: Do It Right or Not at All” has long been the basic summary of laboratory guidelines in aDNA for PCR based approaches (see textbox 1). Exogenous DNA interfering in studies is summarily referred to as contamination. It can be introduced both prior to the handling within an aDNA laboratory as well as within the laboratory.

Prior to entrance in the laboratory, a specimen can be contaminated by the microbial degradation after death, by burial proximity with other organisms, and/or by boiling/cooking after death (mostly an issue with animal bones). Samples can also be contaminated by human handling during the excavation, during its cleaning and storage in museums, and during morphological investigations (Bouwman et al. 2006; Richards et al. 1995). Human handling during and after excavation is one of the biggest sources of sample contamination, since direct handling and washing procedures employed seem to contaminate both sample surfaces and to some extent, depending on preservation and porosity, also the interior of bones and teeth (Gilbert et al. 2006a; Gilbert et al. 2005b; Malmström et al. 2005; Salamon et al. 2005; Sampietro et al. 2006). Decontamination strategies used to counter this type of contamination are UV irradiation of sampled bones prior to extraction, removing the outer layers of bone before sampling, and bleaching of bone powder after sampling. These precautions are efficient to a great extent, but not completely. Often further analysis after acquiring the genetic information is crucial to verify the authenticity of the results. In addition to decontamination, it can be important to prevent unnecessary contamination. As the field of archaeological genetics has grown, so has the communication between geneticists and archaeologists and anthropologists. This communication has led to a better understanding of the importance of protective gear during sample retrieval, and has slowly resulted in altered working procedures towards better excavation practices that decrease the contamination from excavation and post-exavcation handling.

Contamination can also be introduced at any stage during the laboratory process, especially during the DNA extraction and amplification procedures. Even when working in a clean lab according to GLP, contaminant DNA can be present in small amounts in chemical reagents and on laboratory disposables. This contamination would not cause problems in a modern DNA facility, since the sheer amount of sample material would effectively “drown out” these minute amounts. In the case of
ancient facilities where the sample amount can also be minute – they can overshadow the endogenous DNA. Therefore extra care in ensuring the cleanness of laboratory disposables, and UV radiation of chemicals, in order to crosslink any potential contaminant DNA, are necessary. Furthermore, the inclusion of many blank controls, e.g. mock-samples that are treated as the real samples, in all processing stages is important in order to control for potential in lab contamination.

Even when all precautions are taken, one can periodically observe contamination in negative controls, as seen in several studies of both human and animal remains (Izagirre & de la Rúa 1999; Leonard et al. 2007; Malmström et al. 2005; Yang et al. 2003). Contamination in negative controls is usually derived from few template molecules and is observed sporadically. All reactions in a step with contaminated negative controls (blanks) are discarded. In studies where contamination of blanks

---

**“Ancient DNA: Do It Right or Not at All”**

**Physically isolated work area.** To avoid contamination, it is essential that, prior to the amplification stage, all ancient DNA research is carried out in a dedicated, isolated environment. A building in which large amounts of the target DNA are routinely amplified is obviously undesirable.

**Control amplifications.** Multiple extraction and PCR controls must be performed to detect sporadic or low-copy number contamination, although carrier effects do limit their efficacy. All contaminated results should be reported, and positive controls should generally be avoided, as they provide a contamination risk.

**Appropriate molecular behavior.** PCR amplification strength should be inversely related to product size (large 500- to 1000-base pair products are unusual). Reproducible mitochondrial DNA (mtDNA) results should be obtainable if single-copy nuclear or pathogen DNA is detected. Deviations from these expectations should be justified; e.g., with biochemical data. Sequences should make phylogenetic sense.

**Reproducibility.** Results should be repeatable from the same, and different, DNA extracts of a specimen. Different, overlapping primer pairs should be used to increase the chance of detecting numts or contamination by a PCR product.

**Cloning.** Direct PCR sequences must be verified by cloning amplified products to determine the ratio of endogenous to exogenous sequences, damage-induced errors, and to detect the presence of numts. Overlapping fragments are desirable to confirm that sequence variation is authentic and not the product of errors introduced when PCR amplification starts from a small number of damaged templates.

**Independent replication.** Intra-laboratory contamination can only be discounted when separate samples of a specimen are extracted and sequenced in independent laboratories. This is particularly important with human remains or novel, unexpected results.

**Biocchemical preservation.** Indirect evidence for DNA survival in a specimen can be provided by assessing the total amount, composition, and relative extent of diagenetic change in amino acids and other residues.

**Quantitation.** The copy number of the DNA target should be assessed using competitive PCR. When the number of starting templates is low (<1,000), it may be impossible to exclude the possibility of sporadic contamination, especially for human DNA studies.

**Associated remains.** In studies of human remains where contamination is especially problematic, evidence that similar DNA targets survive in associated faunal material is critical supporting evidence. Faunal remains also make good negative controls for human PCR amplifications.”

---

**Textbox 1: The golden rule – Guidelines for ancient DNA good laboratory practice as stipulated by Cooper and Poinar 2000.**
has occurred, up to 10% of them are reported as contaminated. To complicate things it has been proposed that contamination may be abundant even when no sequences are obtained from the negative controls. This can be caused by the contamination either originating from the sample specimen itself, as seen in several studies (Handt et al. 1994; Kolman & Tuross 2000; Krings et al. 1997b; Malmström et al. 2005; Richards et al. 1995; Vernesi et al. 2004) or from non-visible laboratory contamination. The latter is explained by a “carrier effect” which posits that low concentrations of contaminant DNA is inaccessible for PCR amplification in the blanks through adherence to plastic ware and equipment (Cooper 1992; Handt et al. 1994; Leonard et al. 2007). If this is the case, the contaminant will become amplified in the presence of DNA extracts containing the actual samples but not in the blanks.

Authenticity

All the previously described technical issues of working with aDNA – namely survival, damage and contamination – boil down to affecting the authenticity of the results/sequences acquired. This is an important issue in aDNA studies. It is therefore not surprising that ways of authenticating data has been one of the most studied aspects over the years. Particularly the publication of Cooper and Poinar from 2000 “Ancient DNA: Do it right or not at all” is one of the key guidelines for good laboratory practice and authentication for PCR based approaches (see textbox 1). Nevertheless the issue has been under constant investigation and many researchers have proposed different methods of dealing with the verification of authenticity over the years (some examples: (Binladen et al. 2006; Bouwman et al. 2006; Bunce et al. 2012; Cooper & Poinar 2000; Gilbert & Willerslev 2006; Green et al. 2009; Handt et al. 1996; Krings et al. 1997a; Malmström et al. 2007; Poinar et al. 1996; Pruvost et al. 2005)).

Most of the general guidelines for authentication were proposed before the advent of high throughput sequencing. They are still important, but the variety of sequencing methods on the high throughput platforms means that a more case-to-case approach in ensuring authenticity may be needed. Thus, as Gilbert and colleagues propose in their cognitive and self-critical approach (Gilbert et al. 2005a), each aDNA study should present in detail the methods employed throughout the experimental pipeline, the factors affecting the reliability of each particular study, and the steps taken to ensure the quality and authenticity of the generated data.

Sampling

The most readily available tissues for sampling in human studies are bones and teeth. Both are susceptible to contamination by direct handling and washing procedures (Gilbert et al. 2005b; Salamon et al. 2005; Sampietro et al. 2006). Therefore, freshly excavated specimens that have been handled as little as possible are preferred (Melchior et al. 2008; Pruvost et al. 2007). As the awareness about the
uses of archaeological genetics has grown among archeologists working in the field
greater precautions are being taken to avoid contaminating samples. These
precautions include major modifications in their way of working, including but not
limited to, minimizing handling of the remains, avoid washing the remains, the use
of latex gloves and protective suits during the excavations. Or minor precautions,
such as taking aside a sample of the remains in situ for genetic testing while
processing the rest of the site as usual. However, many samples of interest are in
collections and museums and for these samples, the contamination is potentially a
serious problem.

![Sampling for aDNA studies](image)

**Figure 3: Sampling for aDNA studies.**

A. Example of a poorly preserved bone fragment. The top part of a femur, rich in trabecular bone, and not ideal for aDNA sampling. 
B. Example of a well preserved bone sample, a piece of skull bone (rich in cortical bone) cut by drill. 
C. A lower jaw with teeth attached. The bone is relatively well preserved, but since the teeth are available, these are the better choice. 
D. Ulna, sampled by saw.

Sampling of material for aDNA studies is an invasive/destructive process. In general about 0.05 – 0.2 grams of bone powder is needed for a genetic study. This is a small amount of material compared to other destructive analysis used on archaeological remains. For example in C14 dating and stable isotope analysis 2 - 4 g of bone is consumed. Especially when working with human remains minimizing the sampled amount is crucial, since the skeletal remains are one of the few information sources that exist about our species prehistory. Thus the remains are often valuable, sometimes irreplaceable. Also the parts of the samples that are the most likely to
contain preserved DNA, are among the most informative in osteological analyses (Mays 2010), meaning their morphological integrity should be maintained. The age, rarity and integrity of the remains, and whether morphological studies are important on those particular specimens, are factors that have to be taken into account when deciding what part of the remains to sample in a destructive manner such as the one needed for genetic investigations.

Bone preservation and porosity, especially porosity caused by microbial degradation has been shown to affect contamination of samples. This correlation is likely caused by an increase in permeability of the bones to hydration, whereas successive hydration cycles could carry liquid borne contaminants deep inside specimens (Gilbert et al. 2005b). All bone is porous, but to different extent. Macroscopically there are two types of bone tissue, cortical and trabecular (also known as cancelous or spongy) bone. Cortical bone is the solid, dense part that forms the outer layer of the bones; it is thickest in the shafts of the long bones and around flat and irregular bones (e.g. femur, tibia, humerus, jaw, skull plate). Trabecular bone is less dense than cortical, and has a more porous honeycomb structure, and can be found at the ends of the long bones and in the interior of irregular and flat bones (e.g. skull, vertebra etc) (Cox & Mays 2000). A general rule of thumb, both due to what we know about preservation of DNA and about contamination, is that bones or bone areas with a high amount of trabecular bone, such as joint ball and sockets, vertebra, etc, or heavily degraded bones are to be avoided when sampling. For examples of the appearance of bones encountered when sampling see Figure 3.

Teeth, or more precisely dentine within the teeth, are a very good source of aDNA (Adler et al. 2011). Dentine has been shown to be less susceptible than bone to secondary (exogenous) contamination since the enamel partially protects it from physical contact with contaminants (Drancourt et al. 1998; Gilbert et al. 2006c; Gilbert et al. 2005b; Sampietro et al. 2006). However, the tooth roots are porous and can be contaminated either by direct contact with the environment or through their connection to the jawbone (Gilbert et al. 2005b). Thus teeth sampled directly from the skull or mandible (see Figure 3C and 4A) are preferred to teeth found loose in situ. The major disadvantage of teeth as sample material is that they can have a high informational value in osteological, morphological investigations, and thus too valuable to be subjected to destructive sampling.

In summary, sampling in aDNA studies is governed by availability (preservation), value (rarity) and morphological information content of each specimen to be sampled. The least invasive sampling takes place in cases where a sample can be brought into a dedicated aDNA facility, and there directly drilled for the small amount of bone powder needed for extraction. If this is not possible a small piece of bone can be cut out and transported to the lab (see Figure 3D).
Materials and Methods

A range of different molecular methods and bioinformatic analyses were utilized in this thesis and this section covers the major methodological principles of working with ancient human remains.

Sampling

In this thesis, I have used bone or teeth to extract DNA from ancient human remains. The first phase of DNA extraction involves cleaning the bone/tooth from potential debris or dirt clinging to them with a dry brush when necessary (all tools need to be cleaned with bleach and UV irradiation between usages for different samples). Subsequently the bones were taken into the clean lab, where they were UV irradiated (6 J/cm²) in order to remove surface contamination. Within the clean lab a bench-top UV sterilized PCR hood was dedicated to bone drilling and sampling. Parts of the bone/tooth were pulverized using a dremel drill with dentistry drill-bits. In total, between 0.75mg and 1.5mg bone powder was used for each attempt to extract DNA from the material.

Figure 4: An example of a tooth that has been sampled for aDNA. A. Sampled tooth, drilled from the side to preserve the morphology. B. A Dremel drill that can be used to extract material from teeth and bone.

Extracting aDNA

Over the years, many different extraction methods have been developed and used for accessing DNA from ancient material. Some methods utilize somewhat unusual
reagents, for example Coca Cola (Scholz & Pusch 1998), but the most prevalently used in recent years are the silica based approaches. Silica based extraction methods can be classified into two types, those utilizing silica in suspension as optimized by Rohland and Hofreiter (2007) based on the work of Boom et al. (1990) and those utilizing silica columns based on the Yang 1998 protocol (Anderung et al. 2005; Dabney et al. 2013; Svensson et al. 2007; Yang et al. 1998). In this thesis, I have used a variation of silica column based extraction (Anderung et al. 2005) and it includes the following steps: First, an extraction buffer consisting of EDTA, UREA and proteinase K (0.5M EDTA pH8, 1M UREA and 100µg Proteinase K per ml extraction buffer mix) is used to demineralize the bone powder and release the DNA into the solution. Second, the DNA is bound to the silica matrix of the purification columns though the use of a binding buffer containing guanidine hydrochloride (Gu-HCL) and isopropanol (CH(CH 3)2OH). Third, the bound DNA is washed and desalted by an ethanol (EtOH) and tris-hydrocloride (Tris-HCL) wash buffer, and subsequently eluted with a volume (50-100 µl depending on the downstream application) of a low salt buffer.

Sequencing methods

Pyrosequencing

Pyrosequencing is a real-time sequencing-by-synthesis sequencing method. It is based on the transformation of pyrophosphates (PPi), released during DNA elongation by DNA polymerase, into measurable light (Fakhrai-Rad et al. 2002; Ronaghi et al. 1998). In the pyrosequencing reaction, the released pyrophosphates are rapidly converted by the enzyme Sulfurylase to Adenosine triphosphate, which in turn is utilized by the enzyme Luciferase to produce light. This reaction cascade is accomplished within milliseconds and the amount of light generated is analogous to the number of nucleotides incorporated (Figure 5).

Pyrosequencing can be used for a wide variety of applications, such as single-nucleotide polymorphism (SNP) genotyping, DNA sequencing, loss of heterozygosity analysis, and CpG methylation studies. It is a method ideal for aDNA applications, since short DNA fragments can be targeted and sequenced while directly screening for foreign DNA. The reading of the target sequence starts from the first base after the sequencing primer (in comparison to Sanger sequencing, where the DNA sequence nearest to the primer is unreliable). Additionally the method is fast, versatile yet robust, relatively low cost (compared to Sanger sequencing) and allows for a strict control of the complete process, minimizing the risk for laboratory contamination. Some examples of studies where traditional pyrosequencing have been used are studies on ancient human kinship (Malmström et al. 2012), a lactose persistence variant (Malmstrom et al. 2010) and sex determination. Other examples include studies about horses (Pruvost et al. 2012), woolly mammoth (Workman et al. 2011), cattle (Anderung et al. 2005; Svensson et
al. 2007; Telldahl et al. 2011), in degradation studies (Stiller et al. 2006), and in forensic studies (Tschentscher et al. 2008; Daskalaki et al. 2011; Li et al. 2012; Madi et al. 2012). For paper I – the investigation of sex determination on 19th century remains from the Spitalfields cemetery – pyrosequencing was used to obtain information from the ancient material.

Figure 5: The principle of pyrosequencing™. The height of the peak indicates the number of nucleotides of the same type incorporated in a row. In the example pyrogram, single and double incorporation are illustrated. (Illustration provided by, and used with permission from, QIAGEN – © QIAGEN, all rights reserved).

NGS - Next generation sequencing

The development of high-throughput DNA sequencing platforms, commonly known as next generation sequencing, enabled aDNA researchers to produce great amounts of data. NGS can be used to sequence DNA without targeting specific regions (e.g. shotgun sequencing) – used in paper III and IV – or by targeting a particular region (similar to conventional molecular cloning) as used in paper I. The shotgun approach simply sequences (a large number of) DNA molecules that exist in the extract. In contrast, the second approach focuses on particular (multiple) regions by first PCR amplifying DNA fragments that are sequenced in parallel.

The most commonly used NGS instruments are at present: the Roche/454 FLX, the Illumina/Solexa Genome Analyzer, the Applied Biosystems SOLiD™ System (Mardis 2008; Shendure & Ji 2008) and the Helicos Heliscope™ (Thompson & Steinmann 2001) instruments.

Most NGS approaches are based on constructing libraries from template genomic material. During library construction, adapters (with known sequences of nucleotides) are ligated to template DNA fragments (Mardis 2008). The library construction workflow for aDNA applications varies slightly from the standard protocols that are developed for ideal conditions (e.g. DNA extracted from fresh
blood). The fragmentation step that is usually needed for modern genomic applications is often omitted since the DNA molecules are already fragmented. Additional modifications include optimization of cycling conditions during library amplification in order to improve sequence efficiency and to avoid amplifying some (few) DNA fragments much beyond other fragments (which would lead to inefficient sequencing) and the use of sensitive polymerases that can read through specific aDNA modifications. In this thesis, I use the Roche/454 FLX and Illumina/Solexa Genome Analyzer and these sequencing technologies are described in more detail below.

**Roche/454 FLX**

The sequencing chemistry of the Roche/454 FLX is a further development from traditional pyrosequencing. As such, it shares the benefit of the short read length, ideal for aDNA, and it can generate substantial amounts of endogenous sequence data, even if a large part of the DNA has an exogenous origin (Noonan et al. 2005). The Roche/454 sequencer amplifies single-stranded DNA copies from a library constructed from fragmented genomic DNA or directly from aDNA extracts, on DNA-capture beads. A mixture of DNA fragments with beads containing oligonucleotides complementary to the adapters at the fragment ends are mixed in an approximately 1:1 ratio. The mixture is emulsified, e.g. encapsulated by vigorous vortexing into aqueous micelles that contain PCR reactants surrounded by oil. The micelles, e.g. the micro-reactors, are pipetted into a 96-well microtiter plate for PCR amplification. The PCR amplification within the micro-reactors results in each bead being covered with approximately 1 million copies of the original single-stranded fragment. For the sequencing each bead is placed into a ~44 μm well on a PicoTiterPlate, together with enzyme beads containing a mix of the enzymes needed for pyrosequencing, e.g. DNA polymerase, ATP sulfurylase, Luciferase and Apyrase.

**Illumina/Solexa Genome Analyzer**

The Illumina system utilizes a "bridged" amplification reaction that occurs on the surface of the flow cell. The flow cell surface has a coating of single stranded oligonucleotides that correspond to the sequences of the adapters used during the library preparation. Single-stranded template DNA fragments (sstDNA) are bound to the surface of the flow cell, and their free/distal end "bridges" to a nearby complementary adapter oligonucleotide on the surface that acts as a primer for elongation. Repeated cycles of denaturation and extension results in a localized amplification of single molecules creating a unique cluster for each sequence on the flow cell surface.

The sequencing chemistry is based on fluorescently labeled 3'-OH blocked nucleotides. During each cycle one nucleotide is incorporated at each strand cluster, the leftover reagents washed off and the nucleotide incorporated at each cluster identified by its fluorescent label. The cycle is concluded with the cleavage of the
fluorescent labels and the removal of the 3’-OH blocking group, thus rendering the cluster strands ready for another round of fluorescent nucleotide incorporation.

Data analysis
Analyzing ancient genetic data entails added challenges compared to modern data, due to the three big challenges of aDNA: DNA fragmentation, postmortem DNA damage and the risk of contamination.

SNP analysis from pyrosequencing data
Pyrosequencing has been used to analyze SNPs in aDNA studies. SNPs can be assayed in “genetic testing mode”, where one receives purely the haplotype as output (for an example of this type of output, see Figure 9) or in the “allele quantification mode”, where one receives a quantification of the alleles present (see Figure 6 for an illustration of this type of output). Investigating SNPs in archaeological samples is particularly challenging because of the low quality of the DNA. For example, in a particular DNA extract, it is possible that only one gene-copy of a (potentially heterozygous site) exists, a phenomena known as “allelic dropout”. Hence, heterozygous individuals can be mistyped as homozygous due to allelic dropout and in order to determine the true genotype of a site, it has become common practice to repeat the genotyping several times. The allelic dropout rate will likely be greater in less well preserved samples due to degradation and fragmentation. The probability of calling a false homozygote (given that it truly is a heterozygote and given the number of replicate genotyping assays) can be computed using the equation of Gagneux et al. (1997):

$$P \text{ (false homozygotes)} = K \times (K/2)^{(n-1)},$$

where $n$ is the number of replicates and $K$ is the observed number of allelic dropouts divided by the total number of observations from heterozygous individuals.
Figure 6: Example of a pyrosequencing pyrogram when used to quantify alleles. The grey area highlights the SNP position, e.g. the area where the heights of the pyrogram peaks will vary due to the presence of a polymorphic site. The allele frequencies are calculated from the peak heights at the SNP position. A. Theoretical results of raw data and allele frequencies for an assay sequencing a bi-allelic SNP (sequence TATATCRAC). B. Actual pyrogram of a sample containing 84% of the G allele and 16% of the A allele.

Mitochondrial DNA (mtDNA) analysis

Ancient mitochondrial genomic data have been obtained through a variety of methods over time, and a detailed overview of all different methods can be found in the reviews by Ho and Gilbert (2010) and Paijmans et al. (2012). In paper II, part of the mtDNA HVR1 region was sequenced using tagged primer PCR that targets multiple fragments for each individual (each with a uniquely tagged primer).
The c-statistic is used to rank the probability of a sequence being ancient based on the amount of type 2 damages present in the cloned sequences. It accomplishes this by creating a median joining network, describing the mutational pathways connecting all the cloned sequences to each other. The length of each branch is proportional to the number of mutational differences between sequences, the numbers on the branch represent the nucleotide positions that differ from the reference sequence and the darker/thicker lines indicate which substitution patterns that are damage-like and that contribute to the c score of the reference sequence. The median joining network shown in this illustration originates from the analysis of one of the mtDNA fragments of individual T3 (Paper II). Seq 1 – Seq n: n number of cloned sequences obtained for a specific fragment. The darker and lighter grey colors in certain positions of the sequences represent a number of inconsistencies in the sequence alignment (usually type 2 nucleotide substitutions (C to T or G to A) and/or contaminant sequences). The sequence designated as the likely authentic ancient sequence (here the fictional Seq 3) is the one in the center of the median joining network.

**Figure 7:** Concept illustration for using the c-statistic used to identify authentic ancient sequences.
The sequence data received through this process is subsequently sorted based on the unique tag-primer combinations identifying each fragment as belonging to a specific individual. The sequences are checked for authenticity (e.g., using the c-statistic), aligned and assembled to consensus HVR I sequences, from which each individual can be assigned to a mitochondrial haplogroup. In the case that the data consists of a large number of sequences in a manner that mimic cloning and Sanger sequencing, the c-statistic developed by Helgasson et al. (2007) is a very useful tool for identifying authentic ancient sequences (Hofreiter et al. 2001a). The c-statistic uses the post mortem damage patterns in the sequence pool in order to construct a median-joining network and calculate a c value and probability of a sequence to be authentic. Even if no undamaged DNA fragments remain in the sample, the median-joining algorithm and the c-statistic may help reconstructing the original undamaged sequence provided there are enough sequences with post mortem damage. The authentication of the sequence data in paper II used the c-statistic and program PhyloNet (PhyloNet v.5, unpublished program - developed by Agnar Helgason at deCODE Genetics in Iceland). The c-statistic concept is illustrated in Figure 7.

Whole genome shotgun sequencing

The data acquired from NGS presents a major bioinformatic challenge, both because of its sheer amount of data, and due to the nature of aDNA (Green et al. 2006; Green et al. 2008; Schubert et al. 2012). For example, the fragment sizes of aDNA are generally shorter than the fragments usually sequenced (e.g., modern DNA). Furthermore, the sequenced fragments will typically include the adapters ligated during the library build, which adds an extra step of removing the adapter sequences, i.e., adapter trimming, using programs such as AdapterRemoval ver. 1.1 (Lindgreen 2012).

After the preprocessing of the raw data (base calling, merging of paired ends if paired end sequencing has been employed and adapter trimming), the data has to be analyzed to ascertain its authenticity. The sequences are mapped to the human genome in order to determine if the sequences are endogenous. The mapping can use standard alignments algorithms (e.g., Li & Homer 2010) that allows for parameter alteration to match the needs of aDNA data. In the specific studies of this thesis (Paper III and IV) the Burrows-Wheeler Alignment tool (BWA) by Li & Durbin (2009) was utilized in order to align the aDNA sequences with the human and chimpanzee genome. Then we extracted haploid variants for the positions that overlapped with SNPs in a reference data set consisting of ~500k SNPs that have all been genotyped in the HGDP panel (Li et al. 2008), HapMap (Altshuler et al. 2010; Surakka et al. 2010) and FINMAP (Surakka et al. 2010). We randomly sampled a single haploid variant from each modern individual to match the haploid state of the low-coverage aDNA data (Skoglund & Jakobsson 2011). Subsequently we performed principal component analysis (PCA) (Patterson et al. 2006) on various population subsets for each historical individual separately. For the PCA analysis,
all transition SNPs where the historical individual displayed a T or an A, as well as triallelic SNPs, were excluded.

PCA has played a prominent part in the population genetic analyses included in this thesis. Principal component analysis is a statistical technique for finding patterns in data of high dimension that has found application in many fields, such as face recognition, image compression, as well as population genetics. Essentially it operates through a mathematical algorithm that reduces the dimensionality of the data, while retaining most of the variation in the data. It reduces the dimensions by identifying the principal components (PCs), e.g. the directions along which the variation in the data is maximal. By using the PCs each sample can be represented by relatively few numbers instead of by values for thousands of variables. Samples can then be visualized as plots, making it possible to visually assess similarities and differences between samples and determine whether samples can be grouped (Jackson 2005).
Research aims

The main objective of this thesis has been to use DNA analysis to approach archaeological and population genetic questions. In order to answer these different types of assays and experimental designs, from single locus markers to genomic approaches, were used according to the needs of each study. Specific aims were to:

I  Develop a fast, simple and reliable sex identification assay of archaeological material that can be used when traditional methods, based on osteological morphometrics, cannot be utilized.

II  To confirm or disprove the direct maternal kinship within a suspected Viking age family constellation from a Swedish proto-Christian burial site.

III  The investigation of the biogeographic ancestry of a selected number of crewmembers from the 300 year-old Swedish man-of-war, Kronan.

IV  To test hypotheses about population structure during the Neolithic transition in Iberia using ancient genome sequencing and investigating their impact on modern populations.
Summary of papers

In the following papers different applications of archaeological genetics are explored. In paper I a pyrosequencing based method for determination of biological sex is tested and validated for use in an archaeological context. Paper II is a study of biological kinship performed on archaeological samples. Both these studies fall into the context of studies containing methods used in aiding to resolve common archaeological considerations. Papers III and IV are of another nature. In these papers, the power of the great amount of modern data available is combined with the shotgun sequencing of aDNA in order to address biogeographical origins of the crew from the man-of-war Kronan (paper III), and the population structure in Neolithic Spain (paper IV).

Paper I: Further developments in molecular sex assignment: A blind test of 18th and 19th century human skeletons

An important question within archaeological investigations involving human remains is the identification of biological sex. That piece of knowledge is crucial in order to understand and form hypotheses around the social and biological structure of past societies, as well as in reconstructing past population demographic events.

Figure 8: Analyzed X- and Y-chromosomal sequences of the human amelogenin gene; underlined - PCR primers; boxed - pyrosequencing primer

Sex determination in an archaeological context is usually based on morphological traits of the skeletons. Although all of the skeleton should be taken into account, key areas of sexual dimorphism are the pelvis and the skull. In particularly the pelvis is the single most reliable area for identification of sex in adults, since the differences there are directly related to functional differences between the sexes (childbirth). Despite the reliability of sex determination using the sexual dimorphism present in male and female skeletons, it has the drawback that it requires well preserved
remains. In addition most reliable methods do not apply to juveniles, since skeletal sexual dimorphism is vague before puberty.

In cases where morphological methods cannot be used, or are ambiguous, methods of molecular sexing systems are an alternative. In this methodological study we tested and validated the accuracy and usefulness of a molecular sexing method based on the amelogenin gene, using pyrosequencing assays. We did this in a double blind study of documented 18th and 19th century human remains from the burial site belonging to the burial area of the Augustinian priory and hospital of St Mary Spital, e.g. Spitalfields London. The study was designed to mimicking a realistic situation where a molecular sexing might be needed, e.g. a burial site with incomplete and highly fragmented skeletal remains. This meant that the national history museum staff provided the lab with small samples from skeletons of known, to them, sex and origin (human or animal) but unknown to the experimenter. The aDNA lab received 13 human and animal samples of unknown sex.

Figure 9: Typical pyrograms for the sex determination assay with signal intensity on the y-axis and nucleotide dispensation order on the x-axis. (A) Female individual, (B) male individual and (C) allelic dropout. Asterisks indicate spacer nucleotides, e.g. nucleotides added to the dispensation order to make the assay pyrograms easier to read.

We used two assays in our experimental design, an amelogenin sex determination assay developed for modern genetic testing by (Tschentscher et al. 2008), and a second one targeting a 91 bp mitochondrial DNA (mtDNA) fragment designed to allow us to determine whether a bone fragment was of human origin or animal (referred to as LHV assay). The amelogenins are a family of genes involved in the
development of the tooth enamel matrix. In humans, there is only one gene located on the XY homologous regions. The sex determination assay is based on the fact that the two homologous copies of the gene (AMELX and AMELY for X and Y chromosome respectively) exhibit numerous polymorphisms. This particular assay used a 3 bp insertion/deletion (see Figure 8 and Figure 9). The animal sequences included in the experimental design of the LHV (the assay designed to test whether a fragment was of human origin or not) were cow, dog, horse, goat and sheep, as these are most likely to be mixed up with human material.

All of the study specimens (n = 13) were consistently typed as human except for one sample, the one designated as nr 215, that did not contain enough DNA. Out of the 13 specimens, 11 could be reliably sex determined and all tests agreed with the sex from the burial records. Allelic dropout (e.g. the false typing of a heterozygote as a homozygote) was observed in all male samples at a frequency of 63%. However with a minimum of 8 replications per sample the risk of mistyping a male as female was less than 0.01% (P = 0.00014). The probability of mistyping a heterozygote as homozygote for one of the alleles (false homozygote) was calculated as follows: P (false homozygotes) = K × (K/2)n−1, where n is the number of replicates and K is the observed number of allelic dropouts divided by the total number of observations from heterozygous individuals (Gagneux et al. 1997). Additionally, no significant difference in dropout frequency could be found between the two sex-linked alleles, as ascertained by a Chi-squared test of all heterozygote observations (p = 0.297, 1 df).

Both our assays were successful and performed well in this double blind study. Especially the pyrosequencing based sex identification assay proved to be fast, robust and reliable in identifying the sex of the sampled individuals and could potentially be used as a standard sexing method for individuals that cannot, or are ambivalently assigned through morphological techniques for sex determination, such as juvenile samples and/or highly fragmented material.

Paper II: Mitochondrial DNA reveals a lack of kin relations in a suspected Viking age family constellation from a Swedish proto-Christian burial site

In the study in Paper I, we approached the archaeological question of biological sex. Another issue often raised in this context, is the one of biological kinship or recovered remains. The individuals featuring in this study were recovered from a proto-Christian cemetery from 1000 AD with no indications of continuity from pagan periods, in Broby (Täby) in the Stockholm area of Sweden. The remains of an older woman and two adolescents were buried in a close proximity that implies a family relation. The woman is believed to be Estrid Sigfastsdotter (Figure 10), the grandmother of Jarlabanke and the matron of a wealthy and powerful Viking age kin. The Jarlabanke kin (Jarlabankeätten in Swedish) actively participated in
transforming the Scandinavian society to a Christian one as illustrated by the runic stones left behind. Estrid’s hierarchical position is manifested by the runic stones as well; she is mentioned in 6 of the rune stones in the area (Andersson 2011; Andersson & Boije-Backe 1999). As far as we know, she is the only high status female on the burial site and also unique in that we know exactly which remains are hers. The remains of two children were found close to Estrid’s burial, so it is reasonable to believe that Estrid and those two children represent parts of a Viking age family. Especially since one of the two children, a boy of approx 10 years of age, seems likely to have been Estrid’s firstborn son – Gag – with her husband Östen. He died young and a runestone nearby is dedicated to him and Östen (Andersson 2011).

In order to establish a possible parent and child biological relation between the two children and the older woman the mitochondrial HVR1 region was sequenced using tagged primer PCR and high throughput 454 FLX/Roche sequencing according to the methodology developed by Malmström et al. (2009; 2012) based on the work of Binladen et al. (2006). A total of 341 bp was sequenced in 7 overlapping fragments of sizes ranging between 85 and 127 bp. Using the unique tag-primer combinations the sequences belonging to each fragment and each individual, the complete 341 bp sequences were identified, aligned and assembled. In the case of alignments containing ambiguities, indicating DNA damage and/or contamination, the authentic aDNA sequences were identified by using the c-statistic (Helgason et al. 2007) through the program PhyloNet (see e.g. Figure 7).

Based on the sequences of the control region the three individuals were assigned to haplogroups. Estrid belong to haplogroup J (16051-16391; 16069T, 16126C, 16189C), the youngest of the children to haplogroup K1a2a1 (16051-16391; 16189C, 16224C, 16311C), and the older child (the one potentially being Gag) to haplogroup K (16051-16391; 16224C, 16311C).

Figure 10: Estrid Sigfastsdotter. Forensic reconstruction in wax made for, and displayed at, Lånmuseet in Stockholm. Photograph provided by Lånmuseet Stockholm ©.
This excludes both children from being related in a direct maternal line to Estrid. The older boy cannot be Gag, but there are other children belonging to the Jarlabanke family that could have been buried with the matron of the clan according to the runic inscriptions. For example Jarlabankes first son, and Estrids grandchild, Sven (Andersson & Boije-Backe 1999), or the children may be fostered charges, and not biologically related. In addition all these 3 haplogroups occur in Europe at reasonable frequencies today. Thus, there is no indication of these children, or Estrid, to be far away migrants.

Paper III: Genomic analysis of biogeographic ancestry of 15 crew members from the 300 year-old Swedish man-of-war Kronan

The royal Swedish warship Kronan capsized, exploded and sank in the Baltic Sea in June 1676, before the beginning of a battle between the Swedish and the allied Danish-Dutch fleet. The majority of her crew, numbering 850 persons, perished. Due to its size, and composition according to historical sources, this population would be representative of a 17th century Swedish society (Einarsson 2001).

According to the historical record, the majority of the crew was recruited from eastern Sweden, from places like Öland and Åland, but also Västerbotten (During 1997). However, an osteological study of morphometrics, performed by Sholts et al assigned a number of individuals from the shipwreck to non-European origins. The study indicated area of origin, such as Patagonia, China and Egypt. The putative presence of non-Scandinavian crew members, and especially from outside of Europe, raised questions concerning the crew composition. Is it possible that the Swedish navy recruited personnel from parts of the world other than Europe during the 17th century?

We used low coverage genome sequencing of the archaeological remains to obtain genotypes for between 90 and 14,000 autosomal genetic markers from a selected number of individuals. Thirteen out of the selected 15 crew members could be sequenced successfully and all, including the putative non-Swedish crewmembers, could be confidently inferred to be males of European ancestry.

In more detail, the principal component analysis (PCA) performed comparing the ancient samples with modern populations, placed individual K1 centrally in the European (EU) cluster but in proximity to the overlapping regions of the EU and Central/Southern Asian (C/S Asia) clusters, and well separated from the American, Oceanian and East Asian (E Asia) clusters. Similarly crewmember K2 and K4 are found in the center or the EU cluster and within the EU and C/S Asia overlapping region. For K3, K6, K9 and K18 the analysis resulted in a lower resolution between

---

1 The crew samples were assigned sample names in the format of K (Kronan) and a number. The numbers range from 1-15 (omitting numbers K5, K10, K15)
the population subsets included, with a combined EU and C/S Asia cluster and separated from American, Oceanian and E Asia where all 4 samples were placed within the combined EU and C/S Asia cluster. For the remaining individuals (K11, K12, K13, K14, K16 and K17) all reference clusters were well resolved, and all these six crewmembers were placed centrally in the EU cluster. (See Figure 11).

For the crewmembers with the greatest amount of sequence data, K1 and K11, we could further investigate the genetic affinities among European reference groups. Both were consistent with Northern European ancestry, falling close to the Russian, Orcadian, and Finnish samples (the other comparison groups being French, Italian, Sardinian, Adygei and Palestinian).

![Figure 11](image-url)

**Figure 11:** Illustration of four representative PC analyses (PCA) of the crew members together with a modern day reference group that spans across the world with PC1 and PC2 displayed. **A.** crewmember K1; **B.** crewmember K4; **C.** crewmember K11 and **D.** crewmember K13. Color coding: Grey – Europe, Red – West Asia, Green - Oceania, Pink – East Asia, and Purple – America.

We could verify the European origins of all samples, including K17 and K18 that had been indicated as non-European (K17 – S America and K18 – E Asia) – see Figure 12. For two samples, the ones with the highest sequencing coverage – K1 and K11, the genetic data points to a Northern European ancestry.
Increasing the amount of sequence data may increase the resolution, allowing for example assigning the ancestry of these individuals to a particular part of Europe or maybe even a particular part of Sweden.

![Figure 12](image)

**Figure 12:** PC analysis (PCA) of the 2 crewmembers suspected non-European ancestry individuals, with PC1 and PC2 displayed. A. K17 and B. K18, suspected origin South America and East Asia respectively. Color coding: Grey – Europe, Red – West Asia, Green - Oceania, Pink – East Asia, and Purple – America.

**Paper IV:** A late Neolithic Iberian farmer exhibits genetic affinity to Neolithic Scandinavian farmers and a Bronze Age central European farmer.

Farming began approximately 10,000 years ago in the near east. The spread of the farming lifestyle, known as the neolithization process, swept over Europe from its local point of origin – the near East – and had reached most of Europe by 5,000 BP. The transition was swift, and comprehensive, containing more than just the act of farming itself, making the description of the Neolithic cultural package an often used term. However the mode of transition and its impact on the demographic patterns of Europe remains largely unknown and under debate. In this study we used shotgun low coverage genome sequencing on Neolithic farmers from the site of El Portalón, the present day entrance to the Cueva Mayor cave complex, in the Sierra de Atapuerca, 15 km east of Burgos, Spain. One individual yielded 66,476,944 bp of DNA sequences, and his genetic composition was compared to modern populations as well as previously sequenced ancient individuals. Three Scandinavian Neolithic hunter-gatherers, a Scandinavia Neolithic farmer, 2 Iberian Mesolithic hunter-gatherers and a South European farmer from the Chalcolithic (see Figure 13A).

There are major genetic similarities between the Scandinavian farmer, the Iberian farmer, and Ötzi, who all cluster with contemporary southern Europeans (see Figure 13B). Similarly, the Scandinavian Neolithic hunter-gatherers and Mesolithic Iberian
all have genetic affinities towards contemporary Northern Europeans. The farmers can be considered to form a “farmer” cluster separated from the “hunter-gatherer” cluster, e.g. the gene-pool of hunter-gatherers. The fact that the Iberian farmer clusters closely with contemporary southern Europeans in contrast to Iberian Mesolithic individuals suggest an early colonization of Iberia and (at least one) later distinct migration event. Nevertheless, while the genomic data of the farmer indicates the possibility of a population discontinuity between the hunter gatherers of the region (Mesolithic Iberians) and farmers, the mitochondrial haplotype found, U5b1b, is a haplotype usually associated with hunter-gatherers. This underlines findings of previous mtDNA studies – that on the mitochondrial level, the farmer populations on the Iberian Peninsula is different from the central European and Scandinavian mtDNA gene-pool. This discrepancy is not a surprise, since mtDNA is only one locus, and thereby only conveys a limited amount of information about population ancestry.

Figure 13: Combined principal component analysis through procrustes transformation. Subfigure A. Spatial and temporal relationships of ancient individuals. B. Illustration of the affinities of all ancient individuals compared to modern populations and each other.

A sample size of one may be small, but since genomic data captures many features of the population by tracing ancestry to an increasing number of individuals (backwards in time), it can still be a powerful piece of information on the population history. Especially considering most studies of the human population during the neolithization thus far, including the studies focusing on the Iberian Peninsula, have been based on single marker studies (mainly mtDNA). Perhaps the largest caveat would be that one cannot know if this specific individual is representative of this population, or an outlier of the population.

There is need for more Neolithic and pre-Neolithic genomic data from Europe and the Iberian Peninsula, but the data from this study shows gene-flow occurring during the Iberian neolithization in a similar fashion that has been shown for Scandinavia.
Svensk sammanfattning (Swedish summary)

Det finns en mängd arkeologiska frågor, som inte kan bedömas endast genom traditionella arkeologiska metoder. På liknande sätt finns det genetiska frågor, t.ex. om evolutionära processer, arters förflyttningar och annan populationsgenetik som inte kan få svar genom enbart genetiska tester på moderna prover. Just i det gränslandet, mellan genetik och arkeologi har arkeogenetik funnit sin nisch.


I många fall så kan dessa frågeställningar besvaras med hjälp av genetiska analyser på arkeologiskt material. Neolitiseringsfrågan är ur biologiskt perspektiv en fråga om genetisk kontinuitet eller om genflöde. Att identifiera släktskap är en fråga om att testa för genetiska likheter eller olikheter (det finns olika markörer, SNPar, mikrosatelliter, eller D-loopen) och att identifera biologiskt kön om att testa ifall individens kromosomuppsättning består av XX eller XY. Även att studera biogeografiskt ursprung är en fråga om att studera genetiska likheter och olikheter mellan olika grupper och individer, eftersom den genetiska diversiteten som en population uppvizar kan ge ledtrådar kring de bakomliggande demografiska processerna i tid och rum.

Arkeogenetik, eller aDNA som fältet oftare benämns (från engelskans ”ancient DNA”), kan definieras som all typ av molekyläriobiologiskt arbete som utförs på arkeologiska material (även om termerna ofta används som synonymer så är de

---

2 fördelningen av arter genom tid och geografisk plats.
3 utbyte av gener mellan populationer av samma art.
egentligen inte totalt överlappande, då arkeogenetik kan omfatta modernt DNA och aDNA kan användas i paleoekologiska studier). Material som kan komma ifråga till den här typen av analyser kan variera beroende på frågeställningen. Exempel på material som använts som prover i aDNA studier är ben, tänder, mumiefierade vävnader, hår, fjädrar, naglar, hovar eller horn. Även växtfrön, delar av papyrus och pergament och sediment från grönlandisar och sjöbottnar har använts.

Det finns dock ett stort abort när det gäller att använda äldre material som studieobjekt. När man använder sig av gamla prover är det DNA de innehåller mycket lite och i dåligt skick (degraderat och fragmenterat) vilket medför extra tekniska komplikationer. Det är även troligt att det är kontaminerat med exogen DNA, dvs. DNA som inte tillhör provet, och ger felaktiga resultat. Vidtar man de åtgärder som krävs för att kompensera de extra problem som degraderingen, fragmenteringen och den potentiella kontaminationen skapar så är arkeologisk genetik ett användbart verktyg.

Artikel I: Vidareutveckling av genetiska könsbestämningar: Ett blindtest utfört på mänskliga kvarlevor från 1700- och 1800-talet.

En av de första detaljerna arkeologer är intresserade av när de gräver ut ett nytt material är könstillhörigheten. Den är en av markörerna som är viktiga för vår förståelse av, och utvecklingen av hypoteser kring, de sociala och biologiska strukturer som förekommit i forntiden.


När de traditionella osteologiska metoderna inte kan användas eller inte är tillförlitliga kan man ta hjälp av molekylärbioologiska metoder. I den här studien utvärderar vi just en sådan molekylärobisk metod, baserad på pyrosekvensering.

För att kunna utvärdera metoden så simulerar vi ett fall då osteologisk könsbestämning inte är möjlig. Namnligen ett fall där kvarlevorna är fragmenterade, och inte identifierbara som vare sig mänskliga eller tillhörande män eller kvinnor. Vårt laboratorium tog emot 13 prover (ett litet benfragment per prov) beskrivna av ett nummer vardera. Uppgiften var att bestämma ifall det var mänskliga kvarlevor,
och om så var fallet, deras biologiska kön. I verkligheten kom proverna från väldokumenterade gravar från gravfältet Spitafields i London.

Två olika experimentella analysprocesser användes för ändamålet. Den ena utnyttjade skillnader i mtDNA mellan människa och de vanligaste djursorter vars ben har rätt storlek och morfologi för att kunna misstas för mänskliga om de är fragmenterade. Den andra utnyttjade de skillnader som finns mellan den version av den homologa genen amelogenin som finns på X och Y kromosomen för att bestämma om varje individ hade endast X kromosomer (kvinnan), eller X och Y (män). När man arbetar med gammalt material är risken att man av slump missar en av de två olika versionerna (alleler) och misstar en heterozygot (XY) för en homozygot (XX), problemet benämns allelic dropout. För att minimera risken att vi bestämt fel kön pga allelic dropout använde vi oss av ett statistiskt mått, sannolikheten för falsk bestämning av heterozygot som homozygot. Detta mått avgör hur många repetitioner av analyserna som krävs för tillförlitlig könsbestämning.

Av de 13 proverna kunde 12 identifieras som mänskliga, och korrekt kön kunde härledas för 11 av 13. Båda analysprocesserna visade sig vara snabba, kostnadseffektiva och tillförlitliga.

Artikel II: En släktskapsstudie från ett vikingatida gravfält.


förstfödde, Gag. Gag dog ung och en runsten i närheten är tillämnad honom och Östen, Estrids första make.

För att se ifall det finns ett biologiskt släktskap på mödernet, mellan de två barnen och den äldre kvinnan så sekvenserades kontrollregionen, eller HVR1 regionen som den även kallas, på mitokondrien m.h.a. en metod utvecklad an H.Malmström. Vi använde även c-statistik, för att utvärdera autentiska gamla sekvenser i de fall de krävdes. Den del av kontrollregionen som sekvenserades är en 341 baspar lång sekvens, från vilken man kan utläsa tillräcklig variation för att bestämma individernas mitokondriella haplogrupp. Vi fann att Estrid tillhör haplogrupp J, den yngsta av barnen haplogrupp K1a2a1, och det äldre barnet (det som skulle kunna vara Gag) till haplogrupp K. Alla dessa haplogrupper förekommer i Europa så det finns således inget i resultaten tyder på ett avlägset ursprung utanför Europa för dessa människor. Men det kanske viktigaste resultatet är att ingen av de 3 personerna delade sekvensmotiv.


Artikel III: En genomisk analys av biogeografiskt ursprung, utförd på delar av besättningen från flaggskeppet Kronan.

Flaggskeppet Kronan, svenska flottans stolthet, förliste i juni 1676 utanför Öland. Ombord fanns en besättning på ca 850 man, och de flesta omkom i fartygshaveriet vilket gör Kronans förlösning till en av de största enkilda fartygskatastroferna i Östersjön. Besättningen på ett fartyg av denna storlek innehåller män från alla klasser, och ett bredt spann av åldrar finns representerade (från skeppspsojkar till officerare ur adeln). Det gör att besättningen är ett representativt tvärnitt av den tidens samhälle. Därmed är Kronans besättning ett intressant studieobjekt. Dessutom innebär kvarlevornas ringa ålder (ca 300 år) och att de bärgats från östersjöbottnen (där de befunnet sig i ca 4-gradig glacial-lera) att de är mycket välbevarade – och idealiska att använda i genetiska studier.

Mysteriet kring Kronans besättning är att de historiska källor som finns, säger att besättningen bestod av män från Öland, Åland och Västerbotten. Morfologiska studier som utfördes på skeletter indikerade dock att vissa besättningsmän
härtammar från fjärran platser. Exempelvis andra delar av Europa, Sydamerika och Östasien.

För att undersöka robustheten i detta antagande använde vi oss av en lågupplöst helgenomskevsering på 18 utvalda individer. De resulterande sekvenserna för varje individ jämfördes sedan med genomiska data för moderna populationer för att bestämma med vilka den delar flest gemensamma särdrag. Resultatet åskådliggörs med hjälp av en Principalkomponentanalys\(^4\), och principalkomponentsgrafer där man kan se med vilka populationer man kan gruppera varje individ. Vi fick nog med data ifrån 15 av de 18 besättningsmännen för att utföra analysen, och alla kunde med säkerhet härledas till ett europeiskt geografiskt ursprung. I framtiden vore det särdeles intressant att öka känsligheten på jämförelserna för att se om man kan härleda besättningsmännen till specifika länder, och kanske även regioner. Ett sådant resultat skulle vara spännande att jämföra med de skriftliga källorna om varifrån soldater och båtsmän rekryterats.

Artikel IV: En sen-Neolitisk jordbrukare från Portalón, nordöstra Spanien, uppvisar genetiskt släktskap med andra jordbrukare.

Ursprunget av jordbruk och boskapshållning som livsstil i Europa kan härledas till mellanöstern, för ca 10 300 år sedan. Och spröd sig relativt snabbt över hela kontinenten. De ungefärliga tidpunkterna för när den Neolitiseringen nådde olika delar av Europa är känt, men inte de exakta mekanismerna för spridningen, och de konsekvenser den kan ha medfört för de europeiska populationerna. Inom ramen för denna studie använde vi oss av helgenomskevsering för att generera lågupplösta sekvensdata från en Neolitisk jordbrukare från Portalón, ett område ca 15 km Öst om staden Burgos i Spanien.


---

\(^4\) metod som transformerar en given uppsättning av \( n \) variabler \((x_1, \ldots, x_n)\), till \( n \) nya variabler (principalkomponenter) som är linjära funktioner av de givna \( x_i \), och parvis okorrelerade.
kan vi konstatera att den genetiska variationen inom det forntida Europa tycks följa kulturella mönster (jordbrukare vs jägare-samlare) snarare än geografiska.

Dessa slutsatser är på intet sätt definitiva, eftersom de baseras på få individer, men de tillför ännu en pusselbit till lösningen av gåtan: Hur spreds neolitiseringen i Europa. Fortsatta studier på fler individer, från flera geografiska områden kommer förhoppningsvis att sprida mer ljus över frågan.
Acknowledgements

When I got started on this PhD I had the naïve notion that I knew what I was getting into… if only I knew. I am deeply grateful for the people I’ve had in my life during these past years, whether they fall into the category of friends, family or colleagues. I would never have made it without all of you!

First of all I’d like to thank my supervisor – Anders Götherström for employing me as a PhD in his group and for everything he taught me. Hanna Johannesson and Hans Ellegren for making the department of Evolutionary biology the scientific environment it is. I would also like to thank my co-supervisor Mattias Jakobsson for all his help and support, inspiring discussions and advice. Thanks for picking up the slack.

I owe a lot of gratitude Gunilla Kärf, thank you for being the bright and generous person that you are. Thank you also for all the help and for everything you’ve taught me – in the lab and outside of it. Most of all I am grateful for all the laughs we’ve shared.

The nexus of craziness, the aDNA lab and group… Thank you to all past and present members of the aDNA group, permanent and visiting researchers, Emma Svensson, Helena Malmström, Oddny Sverrisdottir, Pontus Skoglund, Maja Krzewinska, Linus Girdland-Flink, Irene Ureña, Christina Valdiosera, Amalia Callado, Frida Johnsson, Ayça Omrak and Carolin Johansson. Cia Anderung, thank you for being there for me, as a friend and as a researcher, when I was lost and needed help and guidance. (By the way… Irene, Cris and Amalia, - Less Shakira in the lab and more rock please!) Sen, my officemate and fellow traveler through all our the office moves around EBC – thank you for the dry humor hiding behind your silent demeanor and sorry for freaking you out with bringing old bones to our office. Reto – that espresso machine saved our lives didn’t it? Best idea ever. Rob – my crazy British office mate… you may very well be the most accident-prone person I have ever met! But your humor and happy comments have made me smile more times than I can remember. Nina, stitch and bitch rules, and I’m forever envious of your greaves and mittens of academic power. Agnes, your quick wit is a delight. Jennifer, the master of self sarcasm – I will forever strive to match you. Ioanna, your kindness and sweetness is amazing!
There are so many others I should thank from the my years spent at EBC. Science is not like any eight to five job. Science is passion and creativity. Its long hours and brain-numbing work. Its failures and successes (ok in aDNA more failure than success). And we soldier through it all thanks to the people around us. So a big THANKS to all my office mates and fellow researchers over the years. To anyone whom I met, shared a smile, a laugh or a joke. Sometimes maybe even a complaint. To all with whom I’ve bounced ideas and thoughts with over a conference or pub table. Thank you. And sorry for everybody I should have mentioned by name and forgot… In these final hours of this marathon I plead temporary memory loss.

Special thanks to Agnar Helgasson for allowing me to use his program, PhyloNet and for discussions about its use. To the staff of the «Εφορεία Παλαιοανθρωπολογίας-Σπηλαιολογίας Νότιας Ελλάδος» in Athens, and especially to Νίνα Κυπαρίσση-Αποστολίκα and Ελένη Ι. Στραβοπόδη, for hosting me. And to Ludovic Orlando, Tom Gilbert, Manaasa Rhagavan and Cris Valdiosera for hosting me in Denmark.

I am also grateful to Mattias Jakobsson, Helena Malmström, Hans Ellegren, Kiril Goguev, Annete Herson and Lena Daskalaki for their comments and suggestions on the kappa of this thesis. Also, a big thank you to Karin Olsson – for helping me with transcribing my handwritten scribblings when I couldn’t type.

Also I wish to acknowledge the following foundations and institutions that provided funding for the work presented in this thesis:
The Royal Swedish Academy of Sciences, the Sven and Lilly Lawski foundation, the Helge Ax:son foundation, the Nilsson-Ehle foundation, the Lars Hierta foundation and the Anna Maria Lundin foundation. I am also grateful at the graduate school on genomes and phenotypes for providing the means and opportunity for traveling, attending courses as well as for organizing new courses for its students.

My dear friends and family…

The fosko-crowd. From our shared time at the forskarskolan of Stockholm, and onward, we’ve been companions in the perils and joys of research. Your steady friendship, and our dinners, soup and games, parties and coffee breaks have made these years so much easier and fun! Karin, my crazy bohemian friend. You always have time and room in your heart to take care of us all. Johan, my brother is spirit. You are amazing, a rock in the storm – and I hope I have been the same to you a few times. Christine, always humble and kind. You told me once that I bring out the “worst” in you in the best way. Back at ya girl! Jeanette, your passion and dedication has always amazed me and inspired me. Annelie – quiet, sweet and always positive. You balance the rest of us. Masako. Always adventurous, a citizen
of the world and gourmand. I miss you and hope to see you soon again. Anh Thu, always chic and elegant, even while feeding cells. And Jingwen, I hope you’re happy in Switzerland. You deserve it.

Then there’s the Uppsala crowd. Danai, sister of my heart and pillar of support during these years. I love your frankness and playful attitude. Also thank you for the drawing of the damage figure in the thesis. You’re the best. Maria and Argyro (the Erasmus duo) you’re amazing girls, and never ever forget that. Kiril, your grumpy attitude hides a golden heart… you don’t fool anyone anymore. Rebecka, a ray of sunshine is the only way to describe you, and I’ve missed your sun at EBC ever since you defended.

And last but not least… from back home in Crete. Inga Siden Kiamos – thank you for showing me the joy of science. All this… is in a way your fault. Evi and Tina – My oldest friends, thank you for putting up with me. And my family, especially my parents for always supporting me.

Πατέρα μου γλυκιέ – το πάθος και την φλόγα τα πήρα απο σένα. Αγαπημένη Μαμά – είσαι πάντα η φωνή της λογικής και της ισορροπίας. Σε σας χρωστάω τα πάντα. Στην αγάπη σας. (Translation: My dear Father – you gave me the fire and the passion. My lovely Mother – the voice of reason and the provider of balance. I owe everything to you and your love.)


Teacher AG, Thomas JA, Barnes I (2011) Modern and ancient red fox (Vulpes vulpes) in Europe show an unusual lack of geographical and temporal structuring, and differing responses within the carnivores to historical climatic change. *BMC Evol Biol* 11, 214.


A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology.