Application of Mitochondrial DNA Analysis in Contemporary and Historical Samples

MARIA LEMBRING
Dissertation presented at Uppsala University to be publicly examined in Auditorium Minus, Gustavianum, Akademigatan 3, 75310 Uppsala, Saturday, 14 December 2013 at 13:00 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Niels Morling (Faculty of Health and Medical Sciences, University of Copenhagen University of Copenhagen).

Abstract

The mitochondrion is a tiny organelle that is the power supplier of the cell and vital to the functioning of the body organs. Additionally it contains a small circular genome of about 16 kb, present in many copies which makes the mitochondrial DNA more viable than nuclear DNA. Mitochondrial DNA is also maternally inherited and thus provides a direct link to maternal relatives. These two properties are of particular use for forensic samples, which only contain limited or degraded amounts of DNA, and for historical samples (ancient DNA). This thesis presents work on the mitochondrial DNA in the hypervariable regions (HV) I and II, in both contemporary and historical samples. Forensic genetics makes use of mitochondrial DNA analysis in court as circumstantial evidence, and population databases are used for the calculation of evidence value. Population samples (299) across Sweden have been analysed in order to enrich the EDNAP mtDNA database (EMPOP) (paper I). The application of mitochondrial DNA analysis allowed for analysis of historical skeletal remains: Copernicus, 1473-1543 (paper II), Karin Göring, 1888-1931 (paper III) and Medieval bones, 880-1000 AD, from a mass grave found in Sigtuna, Sweden (paper IV). The thesis also includes analyses of bones and teeth from the shipwrecked crew of the Vasa warship, 1628, samples from the Vasa museum, Stockholm, Sweden (paper V). Overall, the varying age of the samples and the different conservation environments (soil and water) accounted for variations in quality, but still allowed for successful DNA analysis.

Keywords: Forensic genetics, Mitochondrial DNA, HVI/HVII, Population database, Haplotype, Haplogroup, Ancient DNA, Historical DNA samples, skeletal remains, Vasa museum, Medieval samples, Copernicus, Göring

Maria Lembring, , Department of Immunology, Genetics and Pathology, Rudbecklaboratoriet, Uppsala University, SE-751 85 Uppsala, Sweden.

© Maria Lembring 2013

ISSN 1651-6206
urn:nbn:se:uu:diva-209970 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-209970)
To my family
List of Papers

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


Reprints were made with permission from the respective publishers.
Contents

Introduction .......................................................................................................................... 11
DNA profiling ......................................................................................................................... 13
  Autosomal DNA ................................................................................................................ 13
  Low Copy Number ............................................................................................................ 15
Mitochondrial DNA ............................................................................................................. 15
Mitochondrial DNA analysis .............................................................................................. 17
Heteroplasmy ....................................................................................................................... 18
Mitochondrial DNA analysis of crime scene samples ......................................................... 18
Mitochondrial DNA in population genetics ......................................................................... 20
  Mitochondrial Eve ........................................................................................................... 20
  Mitochondrial DNA polymorphisms in population genetics ........................................... 20
  Haplogroups ..................................................................................................................... 21
  DNA databases ............................................................................................................... 23
    National DNA databases ............................................................................................... 23
    Population DNA databases ......................................................................................... 24
Ancient DNA ........................................................................................................................ 26
  The introduction of ancient DNA studies and guidelines ................................................. 26
  Contamination prevention ............................................................................................... 26
  Sources of ancient DNA ................................................................................................. 27
DNA extraction ................................................................................................................... 28
  DNA extraction from bone ............................................................................................... 28
DNA survival and degradation ......................................................................................... 29
Degradation types ............................................................................................................... 29
  Hydrolytic and oxidative damage ................................................................................... 30
Targeting ancient DNA by mtDNA analysis ....................................................................... 32
Authenticity ......................................................................................................................... 33
  Other markers in ancient or degraded samples ............................................................... 33
Applied techniques ............................................................................................................. 35
  PCR ................................................................................................................................. 35
  PCR inhibition ............................................................................................................... 35
  Sanger sequencing ......................................................................................................... 35
  Pyrosequencing .............................................................................................................. 36
  Sex determination based on Amelogenin ...................................................................... 37
Next generation sequencing and ancient DNA.................................37

Present investigations........................................................................39
  Paper I – Mitochondrial DNA analysis of Swedish population samples .39
    Background ..................................................................................39
    Material and methods ....................................................................39
    Results and discussion ..................................................................39
  Paper II – Genetic identification of the putative remains of the famous
    astronomer Nicolaus Copernicus......................................................40
    Background ..................................................................................40
    Material and methods ....................................................................40
    Results and discussion ..................................................................41
  Paper III - An analysis of the alleged skeletal remains of Carin Göring ..41
    Background ..................................................................................41
    Material and methods ....................................................................42
    Results and discussion ..................................................................42
  Paper IV - Mitochondrial DNA analysis of a Swedish Medieval mass
    grave ..............................................................................................43
    Background ..................................................................................43
    Material and methods ....................................................................43
    Results and discussion ..................................................................43
  Paper V- Mitochondrial DNA analysis of the human remains found on the
    Vasa warship.................................................................................44
    Background ..................................................................................44
    Material and methods ....................................................................44
    Results and discussion ..................................................................45

Future perspectives and concluding remarks ..................................46

Acknowledgements ............................................................................47

References ..........................................................................................50
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>aDNA</td>
<td>Ancient DNA</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>D-loop</td>
<td>Displacement loop</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxynucleotide triphosphate</td>
</tr>
<tr>
<td>ddNTP</td>
<td>Dideoxynucleotide triphosphate</td>
</tr>
<tr>
<td>EMPOP</td>
<td>EDNAP mitochondrial DNA population</td>
</tr>
<tr>
<td>HGP</td>
<td>Human genome project</td>
</tr>
<tr>
<td>HV I</td>
<td>Hypervariable region I</td>
</tr>
<tr>
<td>HV II</td>
<td>Hypervariable region II</td>
</tr>
<tr>
<td>LCN</td>
<td>Low copy number</td>
</tr>
<tr>
<td>LT</td>
<td>Low template</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>OXPHOS</td>
<td>Oxidative phosphorylation</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PPI</td>
<td>Pyrophosphate</td>
</tr>
<tr>
<td>RFLP</td>
<td>Restricted fragment length polymorphism</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal RNA</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>STR</td>
<td>Short tandem repeat</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
<tr>
<td>UNG</td>
<td>Uracil-N-glycosylase</td>
</tr>
<tr>
<td>VNTR</td>
<td>Variable number of tandem repeat</td>
</tr>
</tbody>
</table>
Introduction

The DNA profile is unique for each individual, and most people know that it is now possible to relate a person to a certain crime, solely by a microscopic stain from the crime scene, due to specific individual length variations of DNA stretches in the genome. Even cold cases and historical problems can be resolved, but then the DNA that is analysed is mostly mitochondrial DNA (mtDNA), present in many copies in the cell and more persistent than nuclear DNA (nDNA), which is only present in two copies. In this thesis I present the challenges of ancient DNA research and its limitations, but also its potential for analysing certain problematic cases. Firstly, a brief overview of three applications of mtDNA: ancient DNA, forensic genetics and population genetics.

Since 1984 it has been possible to study historical biological samples, either animal or human, using mtDNA. Ancient DNA could first be isolated in an animal study concerning the quagga, a zebra subspecies, extinct in 1883 [1]. As for studies on humans, DNA extraction was first performed in 1985, on samples from an ancient Egyptian mummy [2]. The introduction of the Polymerase Chain Reaction (PCR) technique was revolutionary in that it duplicates the number of copies obtained in each and every cycle [3,4]. Subsequently many studies of ancient human DNA have been successful and have provided deeper insight into historical events or understanding of population migration history in past times.

Recently, in September 2012, human remains were found under a car park in Leicester England. Five months later the University of Leicester announced that the remains had been identified as belonging to the last member of the Plantagenet dynasty, king Richard the III [5,6]. During the process of identification the mtDNA profile from the remains was compared to the mtDNA profile of a now living distant maternal relative of the king [7]. As the mitochondrial DNA is submitted unchanged from mother to all of her offspring (unless a mutation occurs) [8] it provides a direct link between maternal relatives and offers the possibility of consolidating findings in archaeology and population genetics. MtDNA therefore also has a crucial role in phylogenetics (evolutionary history).

Another important field of application is crime scene investigations, in forensic genetics, where sometimes just a few single hairs can be recovered. When mtDNA is used as circumstantial evidence in court the evidence value is significantly lower than the evidence value of nuclear DNA, as a number
of people, those of the same maternal lineage, carry the same mtDNA profile. Population studies can target a cross-section of the current population and then possibly use genetic ancestry for stratification [9,10]. Either way the mtDNA profiles can be divided into haplogroups that can be used for tracing human migration and human ancestry.

Increased knowledge of population composition constitutes a platform that provides a better estimate of the frequency of a certain profile in a given population group. Growing population databases now can make it possible to give a broad indication of the geographical origin of a sample. The European-based database EMPOP [11] currently presents one of the largest collections of mtDNA data. Since mtDNA databases are the sole foundation for frequency estimates of mtDNA sequences, many forensic groups are now working to extend the genetic data of these databases. To cover most of the worldwide populations, a vast amount of work must be done, and all contributions to the databases are important for future improved frequency estimates.

A challenge in both fields, forensic genetics and ancient DNA, is to be able to analyse limited amounts of DNA. The samples are often limited, either because they are old or because they originate from a microscopic stain or trace DNA (touch DNA). The DNA is also affected by natural environmental conditions. The samples can also be affected by more uncommon events, such as a mass disaster, flooding or fire. This often leads to a degradation of the DNA, which means that the DNA segments residing in the sample are fragmented and a challenge to work with. For some of those samples, the only way to process them is by mitochondrial DNA analysis. MtDNA analysis results in a higher sensitivity level and also allows for kinship investigation due to its maternal inheritance pattern. In addition, some geographical information may be obtained, depending on the range of mtDNA genome that is covered by the analysis. MtDNA is therefore a useful marker for very old samples and has the potential to resolve historical questions [5,12,13,14].

This thesis centres on mtDNA analysis, both in contemporary samples (Paper I) and historical samples (Paper II-V).
DNA profiling

In forensic genetics two types of DNA can be utilised for investigations of human biological samples or traces, either nuclear DNA (nDNA) or mitochondrial DNA (mtDNA). An ideal sample process should incorporate questions concerning where the sample was found, the presumed age of the sample and the environmental factors affecting the sample. When sufficient amounts of DNA are present the sample can be processed by analysing nDNA, which is the preferred analysis due to high discriminatory power. However, when the source of DNA is limited or old, mtDNA analysis allows for analysis of minute amounts of DNA or degraded DNA.

Autosomal DNA

Surprisingly few genes were found when the human genome was first drafted within the Human Genome Project (HGP) 2001 [15,16]. The complete human genome was finalised in 2003 [17]. The human genome consists of 98 % non-coding regions [16]. A large number of repetitive segments are abundantly distributed across the genome. These repetitive segments can be classified in different groups according to their length, among them Short Tandem Repeats (STRs) and Variable Number of Tandem Repeats (VNTRs) [18], which in forensic genetics can be used to investigate individual variants.

Genetic fingerprinting or DNA typing was first introduced in forensic genetics in 1985 as a technique for identifying individual human beings [19]. In the early establishment of DNA usage in forensic analysis investigation of the VNTRs, also referred to as minisatellites, was the primary procedure. To allow the analysis of VNTRs, Restriction Fragment Length Polymorphism (RFLP) was performed. Genomic DNA was cut using restriction enzymes creating differently sized fragments, that were separated using gel electrophoresis. The fragments were then transferred to a nylon membrane by Southern blotting. The pattern of the specific variable locations was detected by probes labelled with a radioactive isotope, that was hybridised to the single-strand DNA and visualised with an X-ray film [19]. The technique was first used in court in the UK in 1985 [20]. The patterns produced by RFLP analysis were highly variable between individuals, but very complicated to analyse, requiring large amounts of DNA [21]. Because of the com-
plicity of the RFLP analysis, the new STR markers quickly replaced the VNTRs as the preferred markers in forensics, and they are still the golden standard today. STRs are short repeat units consisting of 3-6 bp in length and can be found throughout the human genome. One key feature of these markers is their high variability, which thereby also provides a high evidentiary value [22]. Another important progress was the invention of the Polymerase Chain Reaction (PCR) technique that allowed for DNA to be multiplied in a much more efficient way than previously [3,4]. Several common sets of STR markers were proposed for forensic use and FBI implemented 13 core loci (CODIS markers, see Figure 1) in 1997, which since then have been the basis for further development and adaptation of the STRs used for DNA identification [23].

![13 CODIS Core STR Loci with Chromosomal Positions](image)

*Figure 1. An overview of the 13 CODIS core STRs. The CODIS core loci D5S818 and CSF1PO are located on the same chromosome, but they are separated by 26.3 Mb with no linkage between them. (Figure courtesy of John Butler, National Institute of Health, Maryland.)*

Multiplex kits have been commercially developed, generally containing 16 markers, including a sex specific marker, amelogenin. As these kits amplify many markers simultaneously, considerably less DNA is used in the process, which makes further analysis an option. In addition to saving DNA, multiplex kits also provide for a much-improved evidentiary value [22]. Today common sets of markers allow for sharing of STR profiles that can be entered in databases and thus have the potential to solve both national and international crimes.
Low Copy Number

Low copy number (LCN), also called low template (LT) DNA, is the approach where lower amounts of DNA can be analysed with STR markers. When we touch objects epithelial cells are shed from the skin and leave latent fingerprints with traces of DNA. These minute amounts of DNA are also called ‘touch DNA’. The first study on DNA in fingerprints, taken from various commonly used objects (cups, gloves, pens etc.) was published in 1997 [24]. When a sufficient number of cells are transferred the DNA can be successfully genotyped. This progress allowed collection on the crime scene of touched evidence materials that previously had not been considered. Swabbed DNA is extracted and many markers are amplified simultaneously, resulting in full or partial nuclear DNA profiles. It has been demonstrated that as little as 100 pg can produce results [25]. In the LCN procedure the number of PCR cycles are increased from 28-30 in the standard STR protocol to 34 cycles [26]. With increased sensitivity the risk for contamination also increases and stochastic effects become more prevalent. However, the improved sensitivity can also successfully generate DNA profiles, where previously no useful results could be obtained.

Mitochondrial DNA

Elements of trait were discovered to be transmitted via the cytoplasm, and these were the first hints of something extranucleus. The mitochondrion is a small but vital organelle and can be referred to as the powerhouse of the cell as it supplies the cell with the energy to uphold the cell’s function. Many mitochondria are present in each cell, and the average number of mtDNA copies per cell is 500 [27]. Every mitochondrion contains extranuclear DNA, and consequently one cell contains many copies of the mitochondrial genome. Two specific characteristics of mtDNA are that it is maternally inherited and has a high mutation rate, approximately 10 times higher than nuclear DNA [28]. The high mutation rate can be explained by the lack of an efficient DNA repair system [29] and the lack of protective histones [30,31].

The complete mitochondrial DNA genome was published 1981 and revealed a double-stranded molecule of 16569 base pairs (bp) in length with the nucleotides distributed unevenly, creating a light strand and a heavy strand. The basis that constitutes the mtDNA is labelled from the light strand from nucleotide number 1 to 16569 [32]. The DNA is composed of a large coding region and a smaller non-coding region. The coding section houses 37 genes that encode for 13 proteins, 22 transfer RNAs (tRNAs) and two ribosomal RNAs (rRNAs). The 13 mtDNA encoded genes are associated with oxidative phosphorylation [33]. The two strands of the DNA molecule, H-strand and L-strand, have their own origin of replication. The heavy strand
(the outer strand) contains the origin of replication $O_H$ in the non-coding
region and the light strand contains the origin of replication $O_L$ in the coding
region [32], see Figure 2.

**Figure 2.** The human mitochondrial genome, 16569 bp. Enlarged at the top is the
control region (D-loop) with the two hypervariable regions, HVI and HVII, that are
commonly used in forensic DNA analysis.

MtDNA replication was discovered in the early 1970’s, and the strand dis-
placement theory was then proposed as a possible mechanism of replication.
In this theory the mtDNA replication was assumed to be initiated by RNA
primers inducing RNA synthesis in the presence of the transcription factor
mtTFA at $O_H$. A DNA polymerase synthesised the new template strand cre-
ating a D-shaped loop via displacement of the parental strand. When the
displacement strand passes the site of the origin of the replication on the L-
strand ($O_L$), corresponding to 2/3 of the replication length of the H-strand,
the replication of the L-strand starts in the opposite direction [34]. However,
recent studies have proposed other models, the strand-coupled mechanism
[35] or the RITOLS (RNA incorporation throughout the lagging strand)
model of DNA replication [36]. It has also been suggested that different
mechanisms could be active simultaneously with different intensity [35].
The 13 protein-coding genes play a major role for the cell’s energy production. Oxidative phosphorylation pathway (OXPHOS) is responsible for the energy production of the cell and is embedded as complexes I-V and two proton carriers, ubiquinone and cytochrome c, in the inner membrane of the mitochondrion. By oxidation of macronutrients, such as carbohydrates, fats and proteins, electrons are released and produce energy via the electron transport chain (complexes I-IV). The energy is stored as a proton gradient, and the protons are utilised by ATP synthase (complex V) to generate ATP [37,38]. The mitochondrion is also part of many other cellular processes, such as apoptosis [39], ageing [40] and disease [41]. The non-coding region is called the displacement loop (D-loop), as this segment of 1 200 base pairs is non-coding for gene products. There are three hypervariable regions in the D-loop (HVI, HVII and HVIII) and these are important and analysed in many research areas, as population genetics, evolutionary studies, medical genetics and forensic analysis.

Mitochondrial DNA analysis

Analysis of the short variable control region in mtDNA analysis has been robust and commonly used since its introduction. Mitochondrial DNA analysis can be performed using restriction site polymorphisms, SNP typing, pyrosequencing or Sanger sequencing. The mutation rate in the mtgenome is generally high but is not uniform across the mtDNA molecule. The non-coding regions do not contain genes, and mutations accumulate with a 10-times-higher mutation rate than in the coding regions [42,43] and therefore are the most variable regions between people. On average per hundred bp, the D-loop displays 1-3 base differences amongst non-related individuals [44].

The HVI and HVII regions are the most thoroughly analysed regions up to the present and span from respectively 16024-16365 and 73-340. HVIII is a less variable region but is increasingly being analysed, as the analysis of the complete control region is now recommended. HVIII spans from 438-574. What is monitored in the mtDNA genome is single nucleotide polymorphisms (SNPs), single base changes, insertions or deletions. The first published mtDNA genome in revised condition, the revised Cambridge reference sequence (rCRS) [32,45], is still used to annotate differences between individuals. This is now the sequence to which all other sequences under analysis are compared. The differences are recorded and result in a personal mtDNA profile. To find common ground for the analysis there are recommended guidelines that should be followed [46].
Heteroplasmy

Reporting an mtDNA profile is often straightforward, and for contemporary samples the sequences are normally easily interpreted, but mixed positions do occasionally occur.

The mitochondria and mtDNAs are maternally transmitted through the oocyte cytoplasm. The sperm’s mitochondrial DNA is degenerated by ubiquitin [47], allowing only maternal inheritance. In general an individual is considered to be homoplastic, i.e. only harbouring one mtDNA population. However, it is inevitable that a mutation may arise at some point among the thousand of copies of mtDNA in segregation, which may lead to a population mixture, heteroplasmy, consisting of a wild type and a mutant mtDNA variant within the same individual. Heteroplasmy can be present in a certain tissue or even in a cell [48,49,50]. Mitochondria undergo replicative segregation during cell division, and mtDNA variants have been observed to have a rapid segregation between generations. The concept of bottleneck, reduction of available mitochondria, has been proposed and can explain rapid segregation of a mutant population [51,52], though the underlying mechanism and how it occurs are still under debate [53,54,55,56]. Heteroplasmy has been associated with disease but is also often seen among healthy individuals [57].

Mitochondrial DNA analysis of crime scene samples

Just as autosomal DNA analysis of crime scene samples can exonerate or tie a person to a crime scene, mitochondrial DNA can also be used for the same purpose. Nuclear DNA is the ideal option, but, when possible, mtDNA profiles should also be considered, even though the evidentiary value is much lower and thus less discriminating for individual identification. MtDNA can be used either as a sole circumstantial evidence or as a complement to nuclear DNA evidence in court [58], as very small amounts can be analysed. At times, when the biological sample only contains minute amounts of DNA, STR analysis may fail to yield a nuclear DNA profile. Then the higher sensitivity of mtDNA is very valuable and may give a useful mtDNA profile. Even very short pieces of hair (0.5 cm) often allow for analysis [59]. When Low Copy Number (LCN) analysis fails, trace DNA has successfully been typed using mtDNA [60]. Studies have also shown that even 10 mtDNA copies equivalent to 30 fg can produce DNA profiles [61].

However, specific properties of mtDNA, i.e. strict maternal inheritance and lack of recombination, paternal leakage [62], heteroplasmies within families [63,64] may cause problems – and these infrequent scenarios should always be taken into consideration during evaluation of results. Crime scene samples subjected to mitochondrial DNA are very often hairs. However hair
is also a tissue that has a higher reported frequency of heteroplasmy, compared to other tissues within the same individual [65]. This property has been reported to be due to the nature of the hair follicle and the rapid growth phase of the hair [66]. Heteroplasmy encountered in reference samples that show an inconsistent nucleotide difference can whenever possible entail the taking of supplementary reference hair samples [67]. If a suspect’s profile matches the profile of the evidence material from a crime scene, it is reported as an inclusion. In cases where the samples do not match completely, there are two possible outcomes. When there are at least two nucleotide differences between the two profiles, an exclusion is reported. This means that it can be excluded that the samples originate from the same person. A non-conclusive result is obtained when only one difference is observed between the two samples, which is not sufficient for either an inclusion or an exclusion [68]. When mtDNA is used in casework analysis it is extremely important that specific properties that may complicate the analysis are correctly monitored and reported. Therefore recommended guidelines should be followed for uniform nomenclature and consistent interpretation of the analyses [46].
Mitochondrial DNA in population genetics

Mitochondrial Eve

In 1987 a paper by Cann et al. [69] presented a phylogenetic tree based on mitochondrial DNA restriction site polymorphisms of 147 humans from different parts of the world. This tree was short rooted in Africa, and it clustered African Americans in one group and all other geographic origins in another one. The branches were relatively short and suggested a female common ancestor who lived in Africa approximately 200 000 years ago [69]. The conclusion of the study was that humans evolved in Africa 150 000 years ago and started spreading out of Africa 10 000 years ago. The study was generally considered correct, except for errors in the application of the method of maximum parsimony. Nevertheless, that paper coined the phrase Mitochondrial Eve.

Mitochondrial DNA polymorphisms in population genetics

MtDNA polymorphisms are accumulated on the mtDNA lineages and are geographically dependent, with a high mutation rate. They are also haploid markers, which means that they do not recombine and that polymorphisms are specific for a certain maternal lineage. MtDNA is therefore a very suitable marker, along with Y-STRs, when the aim is to look into human expansion events, migration and population origins. This is done by studying and comparing the polymorphisms in different populations. Phylogenetic analysis then makes it possible to explore human evolution.

Early mtDNA population studies were often limited to the sequencing to HVI, as it is the most variable mtDNA region, in comparison with HVII. Today the whole mtDNA genome is often sequenced, as it yields a higher resolution of phylogenetic trees [70].
Haplogroups

Neutral mtDNA mutations accumulate over time and have established different, geographically dependent, mtDNA variants within the world population. These variants are maternally inherited and make up mtDNA haplotypes. Related mtDNA haplotypes can be classified into haplogroups by a specific set of single nucleotide polymorphisms (SNPs) [71]. The mtDNA variants that make up the haplogroups are defined by specific mutations, also referred to as haplogroup-defining mutations, that need to be present for classifying a mtDNA profile into a specific haplogroup [72].

Haplogroup classification can be based on restriction fragment length polymorphism (RFLP) analysis of the mtDNA [73] or by analysis of the control region (CR) polymorphisms of the non-coding region [69,74,75]. The control region has a higher mutation rate and thus a higher mtDNA variation. Therefore it is commonly used in establishing sequence variation between individuals. However, in some instances the sole use of the non-coding region is a constraint, and a combination of both coding and non-coding region is at times recommended [76]. The different haplogroups are named alphabetically in order of first discovery (see Figure 3), such as A, B, C, D [77]. The relatedness of haplogroups can be visualised in phylogenetic trees and specific methods, parsimony and neighbour-joining, can be used for clustering. The phylogenetic trees place the African lineages as the most ancient haplogroups L0–L7 [78]. The L3-branch is the descendant of haplogroups M and N, first derived during or after the migration out of Africa [79] and haplogroups M and N include all mtDNA variants outside Africa [78,80,81].
Figure 3. A schematic overview of the worldwide spread of the mtDNA haplogroups. (Reproduced with kind permission of Nature Publishing Group from Schriever & Kittles, 2004.)

A nomenclature of the haplogroups and sub-haplogroups has previously been proposed [82], a nomenclature that remains a work in progress. Recently, updated, recommended guidelines have been put forward on how to handle mtDNA haplogrouping [83] and a regulated updated mtDNA haplogroup phylogenetic tree can be used as an aid in the classification process (http://www.phylotree.org) [76] (see Figure 4). So far haplogroup information has not been used extensively in the forensic community, but the development of the haplogroup classification can perhaps bring about a change in this matter.
DNA databases

There are two types of DNA databases, national forensic DNA databases (restricted to a country and based on samples associated with a crime) and population DNA databases (random selection of individuals representing a population). There are several different types of population databases, collecting data for frequency estimates of Y-STR [84], mtDNA [11] and genomic SNP [85] profiles.

National DNA databases

When a trace of evidence is secured from a crime scene, it will be processed in the forensic laboratory, and unidentified nDNA profiles will be stored in the national DNA database as well as nDNA profiles of convicted offenders. Depending on the country, profiles from missing people can also be entered. The latter type of data can be of use in the event of mass disasters, when nDNA profiles may be used for identification [18]. The collection of DNA profiles is based on STR markers produced by different multiplex kits. Different countries use different multiplex kits, but most markers are common and can thus be of use to fight international crime.

United States is basing all nDNA profiles on 13 core markers [23] that are currently also used in many other countries. The European Union proposed seven common markers as a European standard, far fewer than the 13 core markers in US. This proposal was opposed by the forensic community, as few markers may result in false-positive matches [86]. The European stand-
ard was finally set to 12 STR markers (five added) [87]. There are also different legislations in different countries concerning the types of crime scene samples, suspects and convicts that are entered into the database and thus searchable in the system.

The random match probability of an STR profile is calculated using the product rule, with consideration of the STRs independent segregation and assuming the population to be in Hardy Weinberg equilibrium [88].

Population DNA databases

Lineage markers, Y-STRs or mtDNA polymorphisms are also used in forensics, for human identification or in casework. When there is a matching mtDNA profile between a sample from a crime scene and a reference sample, they might originate from the same individual. However, the match can also be random, due to the inheritance pattern. Therefore the random match probability is a necessary and important measurement in the analysis. Random match probability is an estimate of the rarity of a certain profile in a population, based on the profile’s frequency in the database [89,90]. Therefore, parameters such as the size and the composition of the database are important as they influence the outcome of the frequency estimate. In addition to the size of the database, the geographic representation and the distribution of the samples are important.

In the EDNAP mitochondrial DNA population (EMPOP) database the number of profiles available for comparison is constantly expanding, with 33,195 mtDNA haplotypes included as of September 2013 (http://www.empop.org) [11]. Although the geographic coverage is growing rapidly, some regions or origins are not yet covered in EMPOP, and uncommon profiles in certain populations might not be represented at all. For all these reasons, the accuracy of an obtained frequency in a database search, and the resulting random match probability, is dependent on the representativity of the samples in relation to the geographic regions of origin and the overall number of samples [91,92].

In a population database for lineage markers, the counting method is used for calculation of the random match probability. The number of hits (matches) in the database is divided by the total number of profiles in the database, or in a specific population of the database, and the value obtained will then be the frequency estimate.

A population database is vulnerable in the sense that it is based on typed data from forensic laboratories and literature studies. To control the quality of the samples entered in the database, EMPOP routinely performs quality control checks via a software-based format utilising phylogenetic methods [90].

Today the majority of the sequences entered in the EMPOP database consist of two short fragments, 16024-16365 (HVI) and 73-340 (HVII) within
the control region, though the number of studies reporting complete mitochondrial genomes is now gradually increasing. In most cases and as presented in paper I, generation of population data is performed using Sanger sequencing of PCR products, a time-consuming and costly process. New and more efficient procedures as Next Generation Sequencing (NGS), are now emerging, although at this stage NGS does not yet seem to attain forensic quality requirements [93].
Ancient DNA

Ancient DNA analysis is strongly dependent on the sample’s quality and quantity. In this respect the polymerase chain reaction (PCR) technique has constituted an important improvement and enabled investigation of even a small number of molecules. Guidelines and precautions were soon introduced as the high sensitivity of PCR also increases the risk of contamination of the template molecules. Since the introduction of PCR the number of studies of ancient DNA in animal, human and plant samples has been multiplied. Just as the PCR revolutionised ancient DNA research, the implementation of next generation sequencing (NGS) will be the next additional extension of the field.

The introduction of ancient DNA studies and guidelines

When ancient DNA was introduced as a new discipline in science, it was followed with a great interest. Early on, studies of museum specimens and mummified tissue were published [1,2], followed by a more questionable study on dinosaur bone fragments [94]. There were also other studies of 30 million years old DNA from amber that were published [95,96], and later criticised, which led to a demand for more controls in order to achieve more reliable aDNA publications [97,98]. The guidelines proposed for ancient DNA research [99] and the criteria in those guidelines are generally still followed today, along with updated guidelines [100]. More recently research has established that DNA in samples found in temperate climates does not generally survive for more than 100 000 years [101,102], although the DNA in samples found in permafrost environments may survive for up to 800 000 years [103]. Extremely old samples (over a million years old), however, have not been successfully reproduced so far [104,105].

Contamination prevention

Older bones, used as a source of DNA, have either been recently excavated or can have been stored as part of museum collections. These samples have often been handled by a number of persons involved in the process of recovering the bones from their original location. Preferably excavations should
be performed under contamination control conditions, by archaeologists/osteologists [106]. Older specimens have been shown to be very sensitive to contamination, and therefore treatments to reduce the amount of contamination are an important step in the processing of older DNA samples [99,100]. Several methods are used for removing outside contaminants from bones and teeth, for example UV-radiation, removal of the bone surface or ethanol and bleach washes. It has been suggested that outer contaminants can be efficiently removed with sodium hypochlorite (bleach) [107], but it has also been shown that sodium hypochlorite in itself may reduce the amount of authentic ancient DNA [108]. Some researchers claim that 81-99% of contaminants generally can be eliminated [109]. However, in samples handled in many consecutive steps the contamination can be more severe and difficult to remove with the currently used decontamination methods [106]. One study tested soaking samples in 6% sodium hypochlorite for 21 hours, and found that the endogenous DNA is very stable [107]. Pilli et al. compared so-called “virgin samples”, collected under controlled contami-nated conditions and directly sent to the DNA laboratory, with “lab samples”, handled according to normal procedures in the osteological laboratory. They concluded that teeth are the best source of ancient DNA, as they are less sensitive to contamination than bones, and that bones with known taphonomics are also good candidates for ancient DNA analysis [106]. These studies show the sensitivity of ancient DNA and the importance of following guidelines. Stages where extragenous DNA can be introduced are during the excavation, the osteological analysis and the analysis in the ancient DNA laboratory. Studies have suggested that washing might move surface contaminants into the bone [110]. In the ancient DNA laboratory cross contamination of samples can occur and plastics or reagents may also introduce contaminants. Preferably the taphonomics should be established [106], but this has not always been possible for older samples. In a forensic laboratory, however, the DNA profiles of the personnel handling the sample could be recorded, in order to exclude contamination from staff. As a rule, protective clothing, facemarks, shoe covers, hair nets and disposable gloves are used. Laboratory surfaces are cleaned thoroughly and plastics, reagents and surfaces are UV radiated to minimise introduction of contaminants. In conclusion, strict guidelines should be followed to ensure ancient DNA authenticity [99,100].

Sources of ancient DNA

Ancient DNA has been extracted predominantly from teeth and longer bones (femur and tibia) as they are less porous and more intact, compared to rib bones and vertebrae bones [111]. More compact bones contain more bone cells (osteocytes) [112] that hold the DNA used for DNA extraction.
Teeth, especially, have been used in many studies, as the enamel protects the DNA, which is therefore less accessible to contamination [113,114,115] and also protected from degradation [115,116,117]. When possible, teeth and/or long bones should therefore be the first choice for sampling of material, as they are less affected by negative environmental factors.

DNA extraction

DNA extraction is a very important step in forensic genetics, especially from forensic samples that contain limited amounts of DNA, as well as for older samples (referred to as ancient DNA from here on). The efficiency of the DNA extraction method and the purity of the final extract will play a crucial role for the success rate of analysis of the sample further downstream in the laboratory. The protocols used are often modified and adapted to different materials (for example, hair, bone, saliva) in order to maximise the DNA output. The DNA extraction itself consists of three major steps: lysis of the cell membrane, followed by protein denaturation and finally the separation and collection of DNA from the cell debris. This can be achieved by slightly different methods. DNA can be isolated using e.g. silica-based, Chelex-based or phenolic extraction procedures.

DNA extraction from bone

Bones are the material most often studied in ancient DNA research. Bones can be cleaned and surfaces can be removed to minimise the risk of contamination from modern DNA (see section on contamination). DNA extraction from bone is a particular process, in the sense that bone has a hard surface, a mineral barrier – hydroxylapatite, that first must be broken down. This process requires additions of EDTA, which chelates with the calcium ions from the bone, and subsequently of proteinase K to release the DNA from the proteins [118]. The phenolic extraction procedure was introduced in 1956 [119] as the earliest method for RNA (DNA) purification and is still used today, especially for DNA extraction from bones. Phenol is added to precipitate the DNA from proteins, and the DNA is then separated from the sample by centrifugation and finally purified with ethanol.

The main problems, apart from damage, when working with ancient DNA samples are inhibitors in different forms; humic acid from soil is one example. Inhibitors can often, but not always, be detected by a brown-coloured extract [120,121]. One way to remove inhibitors that have been co-extracted with the DNA is to use column purification. The sample can then be further treated during the PCR set up by different additives (see the section on PCR inhibitors).
DNA survival and degradation

Recently, in investigations of fossils of extinct moa species in New Zealand, the half-life of nDNA was estimated at 521 years. The calculation of mtDNA survival in the different bone samples, between 600 and 8 000 years old, was made under identical conditions of preservation at a temperature of +13.1 °C and the half-life of the mtDNA was approximated to be twice that of the nDNA [122]. However, this estimate is a crude approximation that needs to be refined, but it nevertheless constitutes a baseline for the decay of nDNA and mtDNA. It has also been argued that a great number of parameters interplay in the process, primarily environmental conditions, and further investigation is therefore needed. In the study above there was the possibility of selecting bones from a cave, where the temperature is somewhat cooler and the samples more protected, compared to open-air conditions. In that case the DNA decay occurred in natural environments and was then measured at the collection stage [122].

The environment at the location of the sample origin cannot be controlled and must always be taken into consideration. Later, in the laboratory, conditions can be more closely monitored however, and DNA samples can be stored to minimise the continuing degradation after excavation (storage in freezers or refrigerators).

Temperature, humidity and pH value are some important factors that influence DNA degradation, the pH value is varying between different soil compositions. In fact, environmental conditions at the location where a sample was retrieved are reported to be of greater importance than the sample’s actual age [101,102,123]. There is also a presumed negative correlation between the number of surviving molecules and the degree of fragmentation observed in the DNA [102,108,124].

Degradation types

The living body naturally presents defence mechanisms against DNA damage. However, as soon as death occurs, the degradation of the body via autolysis is activated by enzymes present in the normal functioning of the body, but which after death are released and degrade the DNA. The body is also exposed to attacks from microorganisms and free nucleases that contribute to the fragmentation of the DNA [125,126]. The chemical processes that cause strand breaks in the DNA backbone are accelerated by unfavourable environmental conditions. The environmental factors driving the process further are temperature, humidity and pH value [125,127]. The chemical processes that affect the degradation are introduced either as hydrolytic or oxidative damage [125].
Hydrolytic and oxidative damage

There are several different types of damage that may occur in the DNA, but here only a few common types of damage will be mentioned. The most vulnerable bond in the DNA molecule is the glycosidic sugar-base bond [128]. Hydrolytic breaks that occur at phosphodiester and glycosidic bonds cause single stranded nicks in the DNA. The breakage of β-N-glycosidic bonds between the sugar and the base of the nucleotide results in the base being released, leaving an abasic site (AP site). Eventually the sugar left at the abasic site is susceptible to chemical rearrangement, which leads to strand breaks [129]. Over time the DNA retrieved from an ancient sample will be composed of short fragments.

Miscoding lesions is a major problem in ancient DNA analysis. The lesions are set off by hydrolytic processes, such as deamination (loss of an aminogroup – NH₂) and depurination (loss of the purine bases). The deamination of the cytosine (C) generates uracil. Cytosine is the most common deaminated base [130]. These types of damages can be classified into different groups, type 1: (A→G/T→C) and type 2: (C→T/G→A) [131,132]. These miscoding lesions will introduce sequence deviations in the PCR leading to difficulties in the interpretation of a following sequence analysis. In the PCR the uracil pairs with adenine (A), and adenine (A) will then pair with thymine (T), which leads to C→T miscoding lesions. Also occurring, but less common, is the deamination of adenine (A) to hypoxanthine (Hx), see Figure 5. Since hypoxanthine mostly forms base pairs with cytosine and since cytosine pairs with guanine (G), miscoding lesions arise as A→G [131,132].
Moreover, oxidation of ancient DNA is introduced by free radicals via radiation and is the major contributor of blocking lesions that hinder the PCR amplification process of the DNA [133]. At times, when the DNA is affected by damages it is nevertheless possible to amplify longer fragments. This can, however, be the result of “jumping PCR”, template switching that generates longer fragments. Jumping PCR can thereby result in long fragments, a chimera [134] between authentic ancient DNA and contaminated DNA.

DNA damages introduced in the ancient DNA can, as described above, largely complicate the interpretation of the ancient DNA sequencing results, and the specific interpretation problems arising from miscoding lesions and jumping PCR has to be kept in mind. On the other hand, signs of degradation can also be interpreted as an indication of authenticity. There are several ways to repair the DNA, and the DNA damage introduced by the hydrolytic deamination of cytosines. As an example, DNA can be treated with uracil-N-glycosylase (UNG), where the uracils are excised, reducing the number of misincorporations [135].

*Figure 5. Deamination of adenine to hypoxanthine and deamination of cytosine to uracil.*
Targeting ancient DNA by mtDNA analysis

The amount of DNA needed to perform a routine STR typing is ranging from 0.25 to 1 ng [136,137]. Ancient DNA analysis often concerns minute amounts of DNA, and therefore other markers or methods are used, for example mtDNA, whose multicopy feature allows for analysing amounts as small as 30 fg, corresponding to 0.00003 ng [61]. Thus, only a few copies of DNA may be sufficient for analysing ancient DNA. The negative correlation between surviving DNA in samples and fragmented molecules also means that the amplification is more successful if shorter fragments are used, i.e. 60-100 bp. Larger regions can still be analysed by using overlapping primer pairs. So far mtDNA studies have presented data predominantly using Sanger sequencing of the hypervariable regions, HVI and HVII, of the mitochondrial genome. Analysing ancient DNA is not a straightforward process, and many properties that are specific to ancient DNA must be considered. The sequencing method itself can introduce problems when small amounts of DNA are Sanger sequenced. The most common phenomenon is base modifications, induced by chemical changes in the DNA. These changes are referred to as miscoding lesions and are predominantly introduced by deamination products of C (uracil) 5-methyl-cytosine (thymine) and A (hypoxanthine) (see Figure 5). The base modifications are visualised in sequencing results from ancient DNA. They are often shown as mixtures of either C and T or A and G nucleotides [132]. They are a challenge for the analysis, and repeated PCRs are necessary for confirmation and to detect deviating results [100]. It should also be kept in mind that ancient DNA is usually fragmented, but fragments normally remain within the range of 100-500 bp [102,138], while longer fragments are uncommon. There are reasons to be suspicious if very long fragments can be amplified and the sequencing results have no signs of degradation. The extraction of a sample should also be repeated to verify the sequence results [100]. Possible chimeric DNA sequences (jumping PCR) and errors introduced by the DNA polymerase should also be considered. Profiles of mtDNA can be tested in order to determine whether the DNA profile seems to be phylogenetically correct [139]. The original results are further strengthened if an independent confirmatory test is performed in a second laboratory. Moreover, all persons that have handled the samples should preferably be known and their mtDNA profiles should be compared to the project samples. Ancient DNA should be studied following the recommended guidelines [99,100] in addition to the guidelines recommended for forensic mtDNA analysis [46].
Authenticity

Pääbo et al. 2004 suggest that despite following the guidelines recommended for ancient DNA analysis, there is no guarantee that what has been amplified really is ancient DNA [100]. Nevertheless it is recommended to follow the guidelines proposed, as they can at least minimise the risk of not analysing authentic ancient DNA. Visible signs of damage (changes from C to T or from A to G) in a sequence can be helpful as this can be interpreted as a possible sign of authenticity [140,141]. Others emphasise that descriptive information of the sample history and potential problems within a specific study should be addressed [142]. According to the optimal scenario the sample should be processed as thoroughly as possible, taking into account the background information at hand.

Ancient DNA research is confronted with numerous challenges, as presented in this thesis. Nevertheless a vast amount of successful research projects have been accomplished in the field of ancient DNA over the years, answering many interesting questions and driving the field forward.

Other markers in ancient or degraded samples

Although the mitochondrial DNA genome is of special interest as a marker, due to its multi-copy feature, other markers may also be used for less degraded samples to render genetic data, for example mini-STRs, Y-STRs and SNPs and indels.

Mini-STRs

Stochastic effects, allelic drop-outs [26] and high background noise are often problems encountered when processing older or degraded samples with conventional STR analysis. However, shorter targets can render successful amplification even from severely degraded samples [143] and can therefore also be useful in ancient DNA studies [144]. By reducing the target size, placing the primer binding sites closer to the STR region of interest, shorter amplicons of nSTR markers are created, so called mini-STRs. When comparing conventional STRs in commercial multiplex kits with the reduced sizes of the multiplex kit MiniFiler, the latter produced an increase of the amount of information [144,145,146].

Single Nucleotide Polymorphisms

Single nucleotide polymorphisms are the most abundant variation in the human genome, occurring approximately once in every 1000 bp [16]. These single base changes are, in contrast to the multiallelic STRs, biallelic markers (two alleles per loci), which reduces the power of discrimination. Nevertheless they are robust markers that can beneficially be used for identification purposes in highly degraded samples, as they have a much lower muta-
tion rate (1 in $10^{-8}$ to $10^{-7}$) [147], compared to STRs, which have a higher mutation rate ($10^{-3}$) [148]. Thus they are stable and valuable markers for kinship analysis [147]. For identification, an analysis of 52 SNPs will obtain approximately the same evidentiary value as an STR multiplex [149]. SNPs can also provide an insight into phenotypic traits, eye colour and hair colour, as well as ancestry information[150,151].

**Y-STRs**

The Y-chromosome is one of the two sex-determining chromosomes in humans (the second being the X chromosome). It contains 78 protein-coding genes and also harbours the male specific region, MSY, needed for the development of male characteristics. Just as mitochondrial DNA is maternally inherited, the Y-chromosome is the equivalent on the paternal side. The Y-chromosome is a small chromosome, only 60 Mb long, that contains a large number of highly repetitive sequences, found in the non-recombining region, NRY, and covering 95% of the Y-chromosome [152]. The Y-chromosome has been scanned extensively in order to locate STRs of use for forensic genetics. The Y-STR markers can be used in sexual assault cases, paternity testing and in deficient paternity cases, where the father is not available for testing. Y-STRs are transmitted unchanged from father to son and can therefore also be beneficially applied in historical investigations [13,153]. However, when evaluating data for paternity testing, it should be considered that mutations within the paternal lineage cannot be excluded. Reliable knowledge of the mutation rates among the markers currently used for paternity testing (e.g. Yfiler) is therefore important [154].

**Indels**

Analysing indels, insertion-deletion polymorphisms found in the human genome, is another way to assist in forensic investigations. The indel markers are on average shorter than conventional STRs, which makes indel markers good candidates when targeting degraded samples [155]. Indels have been evaluated in paternity testing and show low paternity exclusion values [156]. However, the low mutation rate of indels is an advantage that may provide additional useful information to STR data [157].

34
Applied techniques

PCR
The Polymerase Chain Reaction (PCR) technique was invented in 1983 and is an in vitro process of DNA replication, capable of producing vast amounts of DNA copies [3,4]. The PCR reaction needs DNA polymerase, primers and nucleotides to function. The PCR process starts with denaturation (95°C) of the strands from double- to single-stranded DNA. In the annealing step (50-60°C), the primers attach to the complementary sequence. In the extension step (72°C) the DNA polymerase, e.g. Taq DNA polymerase, builds a copy of the strand on the flanking segment between the primer sites. These steps are normally repeated for 28-30 cycles. The number of DNA copies increases exponentially.

PCR inhibition
The PCR is an invaluable technique, robust and fairly straightforward, used in many research fields. However, when it comes to forensic samples, and especially ancient DNA samples, inhibition of PCR is a problem. When the extraction process is carried out, inhibitors can be coextracted along with the DNA residing in the material. PCR inhibitors are encountered in various materials as humic acid soil [158], dyes in clothes [159], calcium ions in bone [160], melanin in hair [161], hemoglobin in bloodstains [162] to name a few. One way to circumvent inhibitors is by dilution, thereby reducing the amount of inhibitors in the PCR. Bovine serum albumin (BSA) and glycerol are sometimes added in the PCR [163]. Increasing the amount of DNA polymerase has also been shown to enhance the PCR process [164].

Sanger sequencing
In 1975 the Sanger sequencing was invented with the aim of determining the order of the nucleotides in a DNA sequence [165]. The PCR products are purified, and then a Sanger sequencing reaction is carried out. The components of the reaction are a DNA primer, a DNA polymerase, deoxynucleotide triphosphates (dNTP) and dideoxynucleotide triphosphates (ddNTP). As
ddNTP is missing a hydroxyl group, which it needs for binding with the next nucleotide, this results in the termination of the DNA strand elongation, whenever a ddNTP is added. At the end of the reaction there are several DNA fragments of different lengths, which can be separated by Capillary Electrophoresis (CE). The different ddNTPs are labelled by different fluorescent dyes and separated by length, thus displaying the base order of the DNA sequence.

**Pyrosequencing**

The pyrosequencing technique was developed in 1996 by Ronaghi and Ny-rén at the Royal Institute of technology (KTH). Pyrosequencing is carried out in real time and is based on enzymatic processes for sequence synthesis [166]. The reaction starts with the release of a pyrophosphate (PPi), and the enzyme ATP sulfurylase converts PPi to ATP, which in turn is used by the enzyme luciferase to generate a light cascade when the complementary nucleotide is incorporated in the sequence. The light intensity is proportional to the number of incorporated nucleotides [167]. In the final step of the enzymatic reaction another enzyme, apurase, will degrade the so far non-incorporated nucleotides before a new cycles, see Figure 6. The incorporation of the nucleotides can either be cyclic or sequence directed, depending on the sequence to be analysed.
Figure 6. Illustration of the real-time enzymatic process of pyrosequencing.

Pyrosequencing produces shorter sequence reads of about 150 bp. However, the sequencing results are produced faster, compared to Sanger sequencing. Pyrosequencing is now also the foundation of the 454 technology for Next Generation Sequencing (NSG), with read lengths extended to about 500 bp.

**Sex determination based on Amelogenin**

The amelogenin is a gene that codes for a protein, found in tooth enamel. The Y-chromosome and the X-chromosome both contain this gene, but the X-chromosome has a deletion of 6 bp, with the resulting product 106 bp, compared to the Y-chromosome product length of 112 bp [168]. This feature is very useful, allowing sex determination of archaeological bones as well as forensic intelligence information [59,169].

**Next generation sequencing and ancient DNA**

In recent years several ancient DNA studies have been performed with next generation sequencing (NGS). This technique is currently used in many
fields and allows generation of sequence data from whole genomes in only a few weeks, at low cost. A large number of targeted areas can be analysed simultaneously, and in the medical field complete exome sequencing is a possible strategy. NGS broadly includes three different approaches: sequencing by ligation, pyrosequencing and cyclic reversible termination.

The preparatory steps of the three approaches are relatively similar and generate vast amounts of short sequence reads (generally not longer than 500 bp) and can cover a complete genome. Here the description presented is based on the 454 method by Margulies et al. First, a so-called library of short DNA fragments is created, and these fragments are linked to adaptors in both ends. The adaptors contain a sequencing primer and a region for amplification. The adaptors are in turn attached to beads, placed in water droplets in an oil emulsion, where the amplification takes place. The short DNA fragments are simultaneously amplified on the beads, with a subsequent sequencing step transferring the beads to a well plate and placing one bead in each well. The sequencing process in the 454 method is based on the sequencing-by-synthesis principle, which is also used in pyrosequencing [170] (for further details on this method, see section on pyrosequencing). Ancient DNA is already fragmented, which makes NGS a suitable approach, as short fragments and whole extracts can be used in the library preparation. The first ancient DNA study using the NGS method was published 2006 on the genome of the woolly mammoth [124], and since then the NGS method has been used for a study of a 38 000-year-old Neanderthal genome [171] and led to further insights into human evolution and the relationship between Homo sapiens and the Neanderthals. As mentioned earlier in this thesis, ancient DNA survival and degradation are strongly dependent on the impact of environmental factors. One example of this came in 2010, when the complete genome was obtained from a 4000-year-old paleo-Eskimo. In that case 79% of the genome could be sequenced, as the DNA extracted from hair had been very well preserved in the permafrost [172], whereas only 45.4% could be sequenced for the woolly mammoth, also found in permafrost. The high throughput of NGS and the rapid progress of different additional techniques will enable numerous analyses that might not otherwise have been attempted. However, there is room for refinements, crucial especially for ancient DNA and in forensic genetics, where scarcity of material is still a problem [93].
Present investigations

Paper I – Mitochondrial DNA analysis of Swedish population samples

Background
Knowledge of the genetic population structure can be of crucial importance in certain situations, for example when interpreting the evidentiary value of DNA profiles for forensic purposes. When routine STR analysis fails to generate a nuclear DNA profile, as in the case of a limited evidence source or when the DNA is old and/or degraded, mtDNA can be successfully used. To achieve reliable estimates in forensic casework analysis, it is important to know the frequency of occurrence of an mtDNA profile in a specific population. To gain more insight into the population composition and structure of the Swedish population, by means of mtDNA profiles, we conducted a study of 299 samples.

Material and methods
Volunteering blood donors (299 individuals), with self-assigned Swedish origin of at least two generations on the maternal line, were randomly selected for the study. The donors’ blood samples were collected at Uppsala University Hospital. Götaland, Svealand and Norrland, all the three major geographical areas of Sweden, were represented. The hypervariable regions HVI 16024-16365 and HVII 73-340 was targeted by PCR and Sanger sequencing. The genetic diversity and the random match probability were calculated, and the mtDNA profiles were classified into haplogroups and compared to other European populations.

Results and discussion
The sequence data was divided into haplogroups employing both Haplogrep (http://haplogrep.uibk.ac.at) and manual classification, using Phylotree (http://www.phylotree.org). Our sequence data revealed a homogeneous Swedish population where 46.5% carry the haplogroup HV. We found 8 additional major haplogroups I, J, R1, T, U, K, V and X. The genetic diversi-
ty was 0.9895 (+/- 0.0023), using the software Arlequin, and the random match probability was calculated to be 1.39%. Our results are in agreement with previous studies, e.g. Tillmar et al. 2010 who also studied the Swedish population, and with other European populations. All retrieved data was sent to the Innsbruck Institute of Legal Medicine, Austria. After revision the data was fed into EMPOP database (EDNAP forensic mtDNA population database). The 299 mtDNA profiles are our contribution to an ever-growing database.

Paper II – Genetic identification of the putative remains of the famous astronomer Nicolaus Copernicus

Background

Born in Thorn, Poland, in 1473, Nicolaus Copernicus was an astronomer known for his heliocentric model of the universe, placing the sun in the centre and the planets orbiting around the sun. In addition to his astronomical research, Copernicus had duties as a canon in the Frombork cathedral. After his death in 1543 the track of his burial place was lost. It was later suggested that a possible location of his grave could be under the St Cross Altar in the Frombork cathedral, but the exact spot was not known. A number of unsuccessful excavations were made to find the grave, but in 2004 a Polish research team launched a new search. They found skeletal remains from a male, 70 years of age and with a visible cut above the nose, according to osteological analysis. A subsequent skeletal reconstruction resulted in many features that were in agreement with known facts about Copernicus (e.g. a war wound above the eye). There was no living relative that could be used as a reference. However, a calendar that had belonged to Copernicus for many years and is currently a part of the collections of Museum Gustavianum, Uppsala University, turned out to be useful. It was suggested that hairs found in the calendar could be used as a possible reference.

Material and methods

This study was performed in collaboration with the laboratories of the Museum and Institute of Zoology, the Polish Academy of Sciences in Warsaw and the Institute of Forensic Research in Krakow, Poland. The teams agreed on which samples from the skeletal remains should undergo DNA analysis, namely three teeth and a femur bone. As a first step mtDNA from the hyper-variable regions HVI and HVII and coding region mtDNA SNPs were analysed. Additionally amplification of 15 STRs was attempted with the Identifiler kit (Applied Biosystems) and amplification of 16 Y-STRs was per-
formed using the Y-filer kit (Promega). Moreover, eye colour was also predicted by analysing the SNP position rs12913832 of the HERC2 gene. Hairs found in the calendar in Museum Gustavianum, Uppsala, were analysed and the mtDNA sequence was compared to the mtDNA profile obtained from the bones. All obtained profiles were checked for their occurrence in respective databases.

Results and discussion

Independently all three laboratories obtained a matching mtDNA profile from the tested femur and teeth. The estimated occurrence of this mtDNA profile was 1 out of 483 individuals (as searched in EMPOP, 2009). MtDNA SNP results assigned the Copernicus mtDNA profile to haplogroup H. The Y-STR haplotype was searched for in the YHRD database, but there was no match among 2,595 complete haplotypes. Six out of 15 STRs produced a genotype, i.e. the shorter STRs, while longer fragments did not amplify, which is a clear sign of degradation. Sex determination using the amelogenin gene marker was included in the Identifiler kit and revealed a male profile. The SNP analysis of the HER2 gene gave a C/C homozygote genotype, indicating a blue eye colour with 83% probability. A facial reconstruction, based on the skull measurements, was made at the forensic laboratory of the Warsaw police. All these data combined strongly indicates that the excavated remains were in fact those of Nicolaus Copernicus. The remains were therefore reburied in Frombork Cathedral in 2010, in a grave with his name, a late recognition of Poland’s famous scientist.

Paper III - An analysis of the alleged skeletal remains of Carin Göring

Background

Carin Göring, born 1888 in Stockholm, is known for her connection to the Nazi leaders of Germany in the 1920’s and early 1930’s. Carin Göring married her first husband, Nils von Kantzov in 1910 and the couple had a son, Thomas von Kantzov. Her second husband was Hermann Göring, a man with a leading position in the Nationalist German workers’ party, NSDAP. In 1931, after a long period of illness, Carin’s condition worsened and while visiting Sweden she died of heart failure. Her family had her buried in the family tomb at Lovón, an island near Stockholm. Herman Göring built a hunting lodge in the outskirts of Berlin and named it Carinhall in her memory. He moved her remains from Sweden, reburied them at this estate.
and had the original coffin encased in a zinc coffin and then in an additional tin coffin.

Herman Göring destroyed Carinhall 1945, to prevent the Russians from getting hold of his property, but the mausoleum was left more or less intact. In 1950 skeletal remains of a human torso were found near Carinhall. As they were supposed to be the remains of Carin Göring, a new burial in her name took place in Lovön, Sweden. Much later, in 1991, the Swedish National Board of Forensic Medicine in Linköping received skeletal remains from a zinc coffin, accompanied by an amateur film showing grave hunters digging in the vicinity of Carinhall. These remains were sent to Uppsala University for analysis.

Material and methods
The skeletal remains first underwent thorough osteological analysis. Bone pieces from a cranium and the ulna bone were then selected from the skeletal remains and subjected to DNA extraction. A paraffin-embedded tissue sample from Carin’s son, Thomas von Kantzov, was used as a reference. To avoid contamination during the extraction and the PCR set-ups, recommended guidelines for mtDNA analysis were followed.

Amplifications of the mtDNA sections of the displacement loop, the D-loop, i.e. the two hypervariable regions, HVI (16128-16340) and HVII (45-287) were performed, and amplification of longer fragments of 400 bp and 600 bp was also attempted. A sex determination (the amelogenin gene marker) was performed by pyrosequencing, and genotyping by pyrosequencing was also attempted for five nSTRs: TPOX, D5S818, TH01, D7S820, D8S1179.

Results and discussion
The remains and the paraffin-embedded tissue showed identical mtDNA profiles, a common profile among Europeans. A fragment length of about 200 bp in the two hypervariable regions, HVI (16128-16348) and HVII (45-287), was amplified for both the samples from Carin Göring and the reference sample from Thomas von Kantzov. To evaluate the degree of degradation and indicate the authenticity of the results, longer fragments of 400 bp (15971-16410) were also targeted, but only amplified partially. Fragments of 600 bp failed to amplify, indicating older bones. A sex determination of both samples revealed female sex for the remains and male sex in the case of the reference sample.

The samples were further investigated for five nuclear STRs with shorter fragment size, as shorter fragments allow for easier amplification in degraded samples. Three out of the five STR markers were amplified.
The three STRs (TH01, D7S820 and D8S1179) show at least one identical allele in the remains and in the reference sample, indicating a possible relationship. DNA profiles from the analysts were compared to the obtained mtDNA profile and the nuclear STR genotyping to exclude contamination. The osteological data displayed signs of disease and indicated an estimate of age and sex, in agreement with reported facts and presumptions surrounding Carin Göring’s life. Osteological findings in combination with the mtDNA and nSTR analysis constitute pieces of evidence for the identification of the remains as being the skeletal remains of Carin Göring.

**Paper IV - Mitochondrial DNA analysis of a Swedish Medieval mass grave**

**Background**

In 1998 a mass grave was found during excavations in Sigtuna, in an area known as St Laurence’s church yard. The remains were dated to 880-1000 A.D. and estimated to a total of 19 individuals, of different sexes and ages. Cuts from bladed weapons were also observed on the bones. To gain further insight and complement the osteological findings the bones were subjected to mtDNA analysis.

**Material and methods**

Out of the 21 bones, mostly femur bones, pieces of approximately 1 cm² were used for the DNA extraction. Extraction was performed in a dedicated laboratory for ancient DNA (aDNA) studies, and recommended guidelines for aDNA analysis were followed. Several shorter fragments of 220 bp and 130 bp in the mitochondrial hypervariable regions were amplified and analysed by Sanger sequencing. The samples were amplified in original concentration and in dilutions 1:10 and 1:20 to counteract inhibitors.

**Results and discussion**

Approximately one half of the samples amplified successfully for 220 bp and the other half allowed amplification of 130 bp. The higher success in amplifying shorter fragments is an indication of degradation of the DNA in the samples. Nevertheless, the obtained sequence data showed overall good sequence quality, though some damaged sites were observed in some samples. In order to remove and thereby reduce the amount of visible damage, UNG treatment of the samples was used prior to PCR amplification. The degrada-
tion is not surprising, as DNA damage and fragmentation occurs over time. The analysis also reveals that some of the individuals can be maternally related. Further analyses will be performed to confirm the results and, if possible, a sex determination using the amelogenin gene will also be performed. Improved protocols now make it possible to retrieve high quality sequence data that can yield a more complete picture of past events and also contribute to the understanding of the Swedish population structure, culture and habits of the time.

Paper V- Mitochondrial DNA analysis of the human remains found on the Vasa warship

Background
In early August 1628, the Vasa warship sailed the calm waters of Stockholm harbour, Sweden, on its maiden voyage. Family members of the seamen were on board to take farewell of the sailors. Without warning the ship tilted twice in a mild wind. The second time the ship was not able to straighten, and the magnificent ship sank in front of all the spectators that had gathered along the shore to witness its send-off.

After the salvage of Vasa an osteological analysis revealed an estimate of remains from 15 individuals who perished in the accident, and among them possible relatives. Today the skeletal remains are in the possession of the Vasa Museum, where some are on display to visitors. To acquire genetic information and to investigate possible relationships between individuals, pieces of bones from all individuals were collected. Using mtDNA analysis, the present study aims to resolve some of the questions that have persisted since Vasa’s salvaging in 1961.

Material and methods
Teeth were sampled from all individuals, that were to be analysed. Bones and teeth were also collected from samples with unknown or questioned origin. Analysis of mtDNA was performed as it is maternally inherited and allows for kinship analysis. Moreover mtDNA is sensitive enough for analysis of even highly degraded samples, due to the many copies present in each cell. The hypervariable segments, in the D-loop of the mitochondrion were amplified and analysed. Samples showing signs of damage were treated with uracil-N-glycosylase (UNG) to reduce the amount of damaged molecules by removal of DNA strands containing uracil.
Results and discussion

Despite more than 300 years under water, all samples have been successfully amplified and sequenced. Nevertheless, the bones analysed up to this point were relatively damaged, as revealed by the sequencing results. However, subsequent UNG-treatment reduced visible damage substantially and thus made interpretation easier due to improved sequence quality. All but two samples had a unique mtDNA profile, revealing that the individuals are not maternally related. Two teeth, have an identical mtDNA profile, although a few sites show some damage. The samples are a part of the collections of the Vasa museum, and there is a story to be told from these bones and about the people who perished with Vasa warship and their relationship to each other.
Future perspectives and concluding remarks

The discipline of forensic genetics is in many ways similar to that of ancient DNA research and also faces similar problems and limitations. The development of more sensitive techniques is carried out within the laboratories, both in ancient DNA research and in forensic genetics, and is advancing in tandem. Even so, the implementation of new techniques in a legal casework setting is somewhat slower, as rigorous testing and maybe even new legislation are needed before they can be applied.

In this thesis the non-coding SNPs have so far been discussed for their potential value in distinguishing between individuals and as a complement to STRs. However, the power of the coding SNPs opens up a new area of research with great potential. Within the coding region of the genome, SNPs within or close to genes can be employed to investigate visual traits. Analysis of informative SNPs have made estimation possible of specific traits as eye colour and hair colour, with high AUC (Area Under the Curve) values [150,173]. Research has also been aimed towards other phenotypic traits like skin pigmentation [174,175] and height [176]. These latter studies give an indication of future possibilities. As for eye and hair colour investigations, the implementation of the tests is already on its way to be used as investigative leads. It will be interesting to see, in the not so distant future, the information that may be retrieved from a single stain encountered at a crime scene.

In this thesis the historical samples have in most cases been subjected to analysis of segments of the hypervariable regions of mtDNA, with the addition of other genetic markers when allowed. Ancient DNA genetics has now embarked on a new era, where Next Generation Sequencing (NGS) presents a myriad of opportunities to expand the research field into territories not yet explored. One of the steps in NGS is to prepare a library containing sequences of interest, using the sequence capture technique. Many ancient DNA studies are now based on NGS, a future possibility also for the sample extracts presented in this thesis. By targeting visual traits, for example, it may be possible to retrieve extended information from the best-preserved samples, beyond the phylogenetic information and maternal relationships. This would make history come alive, visually and literally.
Acknowledgements

**Professor Marie Allen:** I don’t know where to start – but THANK YOU! I have had a great time working with you and learning under your wing. Thank you for trusting me with historical samples and for letting me be part of your group!

I would like to thank my lovely colleagues in the forensic group (in no particular order):

**Mia Bjerke,** for your poise and tips and talks on how to proceed in an academic career!

**Magdalena Bus,** I have enjoyed all the months you have been in Sweden. It was nice to have company when working weekends at the lab and I have so much appreciated your “we’ll fix it attitude” and enthusiasm.

**Joakim Grånemo,** it was so great that you came to the group and shared my responsibilities! Thank you for always sharing your endless knowledge about anything!

**Malin Abramsson,** thank you for your enthusiasm and for making me laugh!

**Hanna Edlund,** it’s great to have you still very close by, so that we can meet up and discuss things in general and ancient or forensic genetics in particular! Not forgetting the dinosaurs😊...Thank you for reading this thesis.

**Martina Nilsson,** finally we worked together after all these years feeling like we already had! Thank you for always including me in “non-work activities” and also for keeping me up to date in work related “police” things. But, most of all for being a good friend (and thank you for reading this thesis)!

**Maria Montelius,** the best and kindest tutor of aDNA I could ask for! So very meticulous and also teaching me how to behave in the aDNA lab. Yeah, finally can meet without discussing mtDNA?

**Anne-Christine Lindström,** good luck with your PhD thesis and let’s stay in touch!

More co-workers and students not to be forgotten **Viktoria A, Anika N, Stina N, Ann S, Iran.**

And other “forensic ladies” **Paula Å, Sara F, Hanna A, Anna-Maria, Charlotte,** thank you for lovely dinners!
Ammar Zaghlood - how long have we “studied together”, both Master degrees and PhD degrees? It has been really good to have you co-studying with me all these years. Thanks for always having time for me and, during these last months, also calming my pre-PhD nerves!

All the present and former PhD students at the IGP department, and especially to the PhD students in the IGP PhD Student Council – thank you for the nice times! I will miss you!

I would also like to express my gratitude to Uppsala Genome Centre, Inger Jonasson and her co-workers, current and former.

Other very important colleagues: Sara Lind, Hongxing and Ludmila.
Mårten Lindén – your computer skills should really be acknowledged, as you have given me a lot of help! We have also a common experience of the daily working life of a PhD student.

My office mates in the past: Angelica D-V, Karolina E, Sara L, Johanna M, Monica L and Magnus - thank you for nice chats in the office.

My wonderful IGP-BMC family – All the colleagues on the 4:floor! It has been great working next to you every day!
Sara E – thank you for all your encouragements and kind words the last months, they have meant a lot. You are truly a great person.
Liza L – I really like working in the same place as you! Just like I did when we were students, working together in different projects.😊
Linda A – Your enthusiasm and energy is compelling!
Karin G – I have enjoyed very much organising things with you, and I think we did a good job, too.
Tonge, Yanhong, Johan V, Rongqin, Sara K for being excellent lab assistants in the forensic Master program.
I don’t want to forget anyone else back in the Rudbeck building – it’s been a great number of years at Rudbeck!
I would also like to thank the wonderful people at the administration offices. Thank you for your time and energy - you are all doing a great job, always appreciated.

Professor Ulf Gyllensten and Professor Ulf Pettersson, thank you for being my second supervisor and my examiner respectively!

A shout out to the Department of Forensic Molecular Biology (FMB) in Rotterdam – all former and present members – thank you for everything and my best wishes to all of you! Especially Miriam Goedbloed for giving me a solid work foundation and for being so kind.
Mannis van Oven – when I was in Rotterdam I didn’t know that we would collaborate in the future. I feel really lucky to have been able to work with you on population genetics, your expertise. I really enjoyed it, thank you again!

Professor Manfred Kayser – thank you for introducing me to the world of forensic genetics and for letting me be a part of your research! Without the very instructive months at your laboratory, I wouldn’t be where I am today.

Financial support was provided by Kjell and Märta Beijer Foundation.

Släkt och vänner – tack för att ni finns!

Kusin Jonte, Jonathan – en gång Uppsalastudent, alltid Uppsalastudent, inte sant? Jag nämner en kusin, men hälsar till alla er andra – ni är bäst!

Kära familjen:

Anna (supersyster), tack för att du är så underbar! Jag är säker på att du blir en succé bland musikeleverna!

Mamma och pappa – ni är de bästa som finns! TACK för att ni alltid ställer upp och stöttar i vått och torrt!
References

Academy of Sciences of the United States of America 106: 12279-12282.


and a Pre-Screening Method. The Open Forensic Science Journal 5: 13-22.


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 950

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, 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 Medicine.