UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 606

Sensitive Identification Tools in Forensic DNA Analysis

HANNA EDLUND

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2010

ISSN 1651-6206
ISBN 978-91-554-7915-2
urn:nbn:se:uu:diva-131904

Dissertation presented at Uppsala University to be publicly examined in Rudbecksalen, Rudbeck Laboratory, Dag Hammarskjölds väg 20, Uppsala, Friday, November 26, 2010 at 09:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English.

Abstract
Edlund, H. 2010. Sensitive Identification Tools in Forensic DNA Analysis. Acta Universitatis Upsaliensis. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 606. 49 pp. Uppsala. ISBN 978-91-554-7915-2.

DNA as forensic evidence is valuable in criminal investigations. Implementation of new, sensitive and fast technologies is an important part of forensic genetic research. This thesis aims to evaluate new sensitive methods to apply in forensic DNA analysis including analysis of old skeletal remains.

In Paper I and II, two novel systems for analysis of STRs, based on the Pyrosequencing technology, are presented. In Paper I, Y chromosomal STRs are analysed. Markers on the male specific Y chromosome are especially useful in analysis of DNA mixtures. In Paper II, ten autosomal STRs are genotyped. The systems are based on sequencing of STR loci instead of size determination of STR fragments as in routine analysis. This provides a higher resolution since sequence variants within the repeats can be detected. Determination of alleles is based on a termination recognition base. This is the base in the template strand that is excluded from the dispensation order in the sequencing of the complementary strand and therefore terminates the reaction. Furthermore, skeletal remains are often difficult to analyse, due to damaging effects from the surrounding environment on the DNA and the high risk of exogenous contamination. Analysis of mitochondrial DNA is useful on degraded samples and in Paper III, mtDNA analysis of 700 years old skeletal remains is performed to investigate a maternal relationship. The quantity and quality of DNA are essential in forensic genetics. In Paper IV the efficiency of DNA isolation is investigated. Soaking skeletal remains in bleach is efficient for decontamination but result in a lower DNA yield, especially on pulverised skull samples.

In conclusion, this thesis presents novel sequencing systems for accurate and fast analysis of STR loci that can be useful in evaluation of new loci and database assembly as well as the utility of mtDNA in forensic genetics.

Keywords: forensic genetics, STRs, Y-chromosome, Pyrosequencing, mitochondrial DNA, skeletal remains, DNA extraction

Hanna Edlund, Uppsala University, Box 256, Uppsala University, SE-75105 Uppsala, Sweden. Department of Genetics and Pathology, Rudbecklaboratoriet, Uppsala University, SE-75185 Uppsala, Sweden.

ISSN 1651-6206
ISBN 978-91-554-7915-2
urn:nbn:se:uu:diva-131904 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-131904)

Till min familj

List of Papers

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

I **Edlund, H.**, Allen, M. (2009) Y chromosomal STR analysis using Pyrosequencing technology. *Forensic Science International: Genetics*, 3(2):119-124

II Divne, A-M*., **Edlund, H*.**, Allen, M. (2010) Forensic analysis of autosomal STR markers using Pyrosequencing. *Forensic Science International: Genetics,* 4(2):122-129

III Nilsson, M., Possnert, G., **Edlund, H.**, Budowle, B., Kjellström, A., Allen, M. (2010) Analysis of the putative remains of a European patron saint- St. Birgitta. *PLoS One,* 5(2), e8986

IV **Edlund, H.**, Nilsson, M., Lembring, M., Allen, M. DNA extraction and analysis of skeletal remains. *Manuscript*

*The authors have contributed equally to the work

Related papers

i Andreasson, H., Nilsson, M., **Styrman, H**., Pettersson, U., Allen, M. (2007) Forensic mitochondrial coding region analysis for increased discrimination using Pyrosequencing technology. *Forensic Science International: Genetics,* 1(1): 35-43

ii **Styrman, H**., Divne A-M., Nilsson, M., Allen, M. (2006) STR sequence variants revealed by Pyrosequencing technology. *Progress in Forensic Genetics, International Congress Series 1288:669-671*

iii **Edlund, H.**, Allen, M. (2008) SNP typing using molecular inversion probes. *Forensic Science International: Genetics Supplement Series 1(1):473-475*

Contents

Abbreviations

ATP	Adenosine triphosphate
bp	Base pair
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
HV1	Hypervariable region 1
HV2	Hypervariable region 2
LCN	Low copy number
mtDNA	Mitochondrial DNA
NGS	Next-generation sequencing
NRY	Non recombining region of the Y chromosome
PCR	Polymerase chain reaction
PPi	Pyrophosphate
rCRS	Revised Cambridge reference sequence
RFLP	Restriction fragment length polymorphism
SNP	Single nucleotide polymorphism
STR	Short tandem repeat
VNTR	Variable number tandem repeat

Introduction

In the recent years, the importance of DNA as forensic evidence in criminal investigations has increased rapidly worldwide. The ability to tie an individual to a crime or crime scene or to exonerate the innocently accused or convicted, by comparing a genetic profile to a reference has become an indispensable tool for the police. In a mid-sized Swedish county, 115 matches (hits) were made between trace and individuals in the Swedish national DNA database in 2008. In 2009, the figure had risen to 150 matches, an increase with 30% [1]. According to European Network of Forensic Institutes (ENFSI), the total number of DNA profiles in the Swedish national DNA database increased from 12 435 to 77 191 profiles between the 2006 and 2009, emphasising the importance of DNA as forensic evidence in criminal investigations.

During the past decade, there has been a rapid development in technologies for DNA analysis in many fields. Minute amounts of DNA can be analysed using next-generation sequencing (NGS) technologies, which enable genome-wide sequencing in less than a few weeks [2]. In forensic genetics, the increase in sensitivity, i.e. improvements in the ability to analyse minute amounts of DNA, has made it possible to analyse a broader spectrum of materials found at crime scenes and also allowed so called cold cases (old unsolved cases) to be reopened and solved. The importance of DNA analysis for individual identification was further demonstrated in the aftermath of the World Trade Center attack in 2001 and the Tsunami catastrophe in 2004 [3,4]. Furthermore, biological evidence found at crime scenes moreover shares several characteristics with ancient DNA and therefore use the same safety precautions, methods and technologies when performing a DNA analysis. For instance, if a DNA extraction method or the use of specific set Polymerase Chain Reaction (PCR) primers gives successful results on an ancient sample, it is likely that the same methods will be successfully applied on forensic samples. The work presented in this thesis is focusing on evaluation of new and highly sensitive methods useful in several areas in forensic genetics including analysis of old bone samples in historical investigations.

Forensic science

Forensic science is the collection of disciplines that scientifically contribute to the legal system, as for instance pathology, odontology, anthropology, chemistry, toxicology and genetics. Forensic genetics is the area in forensic science where DNA analysis is used for molecular identification of biological material found at crime scenes. DNA analysis is also useful for individual identification after mass disasters, paternity testing as well as identification of missing persons.

Challenges in forensic genetics

The condition of the biological evidence found on a crime scene is often not ideal for a molecular analysis. There are several factors affecting the DNA molecule that result in a combination of challenges the forensic scientist has to deal with.

Degradation and chemical modifications: Evidence material is often exposed to a very harsh environment, in which microorganisms, high temperatures and UV radiation can cause fragmentation and degradation of the DNA. In addition to these environmental factors that physically affect the DNA, the material may also be subjected to chemical modifications. When an organism dies, nucleases attack the DNA, which also results in fragmentation of the DNA molecule. Moreover, hydrolytic- and oxidative damage are the two major contributors to DNA damage, which can lead to preventions in the amplification process and cause base modifications [5-7]. There is an inverse relationship between fragment length and successful amplification, i.e. smaller PCR fragments, <300 base pairs (bp), are more likely to be amplified in samples that contain degraded DNA [8].

Inhibitors: The amplification of DNA in evidence samples from various crime scenes can be complicated by the presence of inhibitors in the samples. Inhibitors can be present in the soil, in blood or in textile dyes and they negatively affect the DNA extraction by interfering with the cell lysis step but also inhibit the activity of the polymerase in the PCR reaction [9].

Contamination: The risk of contamination (i.e. introduction of exogenous DNA to a forensic sample) is another major challenge in forensic genetics. Clean laboratories are crucial and to minimise the risk of contamination from the surrounding environment, reagents, laboratory supplies and the analyst, extreme safety precautions have to be followed [10,11]. Benches and equipment should be repeatedly treated with bleach and UV light. Disposable gloves should frequently be changed, and protective clothes, facial masks, hair- and shoe covers should be worn at all times. Pre- and post PCR laboratories have to be physically separated and negative controls should be included at every stage in the analysis procedure. However, even when these safety precau-

tions are taken, there is still a high risk of contamination during the collection of the biological evidence prior to the DNA analysis. Therefore, it is important that all individuals involved in the handling of crime scene evidence are aware of the risk of contamination and take efforts to minimise it.

Mixtures: An additional issue with biological evidence from a crime scene is the risk that the samples may contain mixtures of DNA from different individuals. Mixtures are especially common in sexual assault cases in which samples may contain DNA from both male and female individuals. Often the female's DNA is present in large excess.

DNA quantity: Forensic evidence samples often contain low numbers of DNA molecules, which can result in drop-outs and partial DNA profiles in the typing. In samples with few DNA templates (<100 pg or approximately 17 diploid cells or 34 genome equivalents), the PCR primers have difficulties to hybridise properly to all of the DNA molecules present in the sample, which results in unequal amplification of the alleles. These stochastic effects can result in allele drop-out, heterozygote peak imbalances and increased stutter, i.e. signals that differ in length from the original allele by one repeat unit due to replication slippage [12,13]. One solution for improving sensitivity when analysing low template DNA samples is to increase of the number of PCR cycles (from the standard 28 cycles to 34 cycles). This approach is referred to as low copy number (LCN) analysis [12-15]. However, using LCN analysis does not eliminate the problems associated with amplifying low template DNA samples and furthermore presents an increased risk of contamination. Gill *et al.* have proposed several guidelines regarding the interpretation of results from LCN analyses [12,14]. Moreover, Budowle *et al.* argue that LCN analysis requires extensive evaluation from the scientific community, and that if LCN analysis is proposed for use in an investigation, the scientists should inform all involved parties of the limitations and difficulties associated with the technique and its results [13].

Implementation of new technologies

The process of implementing a new technology in forensic genetics is not straightforward. There are a wide variety of available technologies for DNA analysis based on different chemistries and detection methods, but not all technologies are suitable for forensic DNA analysis. The cost of using new methods must be considered. Moreover, the technology must be able to process samples that contain very limited amounts of DNA and its robustness, accuracy and reproducibility must be extensively validated. These factors mean that the adoption of new technologies in forensic DNA analysis is a slow process that can take several years. Nevertheless, progress in forensic genetics is dependant on the uptake of new technologies. For instance, both larger and more sensitive multiplexes that would reduce the consumption of

precious evidence samples and methods that can analyse highly degraded samples in a short period of time are needed. Furthermore, high throughput technologies for the evaluation of new markers and fast database compilation are also significant for the progression of the field.

History of forensic DNA analysis

In the early 20th century, Karl Landsteiner discovered the human ABO blood groups, which were the first biological markers used to distinguish individuals. The system was based on four groups (A, B, AB and O) and a useful tool for excluding the possibility of a given individual having been the contributor of a particular sample found at a crime scene. In 1985, Alec Jeffreys described DNA fingerprinting [16]. DNA fingerprinting was based on restriction fragment length polymorphism (RFLP) and on the analysis of variable non-coding stretches in loci that may be up to 1000 bp in length, known as variable number tandem repeats (VNTRs) [16,17]. In the initial version of the method, restriction enzyme cleavage was used to produce fragments that were then size separated by gel electrophoresis and detected by southern blotting. Because the number of repeats differed at each locus, a unique pattern for each individual was detected on the blot. RFLP analysis was extremely discriminative, however, the method required relatively large amounts of DNA, between 50-500 ng, for successful analysis [18]. Moreover, RFLP was difficult to interpret because a single probe detected multiple VNTR loci resulting in a highly complex pattern. These multiple locus probes (MLP) were replaced with single locus probes (SLP), which only detected one or two alleles (i.e. bands on the membrane corresponding to a homozygote or heterozygote). The results were easier to interpret and less DNA was needed for successful analysis. As previously discussed, DNA found at crime scenes is exposed to various environmental factors and chemical agents that degrade and fragment the DNA. This makes it difficult to amplify longer fragments, such as VNTRs. The introduction of PCR [19] along with analysis of short tandem repeats (STRs) revolutionised forensic DNA analysis. Minute amounts of DNA could be amplified using locus specific primers, and the sensitivity of the method made it possible to analyse more degraded material found at crime scenes. Today, multiplex PCR amplification of STRs has completely replaced DNA fingerprinting in forensic genetics.

Genetic markers in forensic DNA analysis

In 1953, James Watson and Francis Crick discovered the double helix structure of DNA [20]. Almost 50 years later, the first draft of the entire human

genome sequence was published [21,22]. The human genome contains approximately 3 billion bases and around 20 000-25 000 protein-coding genes. Approximately 99.9 % of the DNA sequence is exactly the same between two individuals [23]. Nevertheless, significant genetic variation exists between different individuals and between populations. Genetic variation can influence an individual's susceptibility to certain diseases, behaviour and physical appearance. However, most genetic variations have no known effects on humans.

Genetic variations in the human genome can be divided into different types. Copy number variation (CNV) are larger variations, where a segment of DNA, ranging from one kilobase (kb) up to several megabases (Mb), can vary in copy number in comparison to a reference genome [24]. Other types of variations are VNTRs, STRs or single nucleotide polymorphisms (SNPs). In forensic DNA typing, highly polymorphic STRs are most commonly genotyped in order to distinguish between individuals, to tie an individual to a crime or crime scene, or to exonerate the innocent. SNPs have also gained interest in the forensic community, since short PCR fragments can be designed and thereby facilitate analysis of degraded materials. Mitochondrial DNA (mtDNA) analysis is based on detection of base substitutions, which are compared to a reference sequence (discussed in the section on mtDNA analysis).

STRs

The human genome contains a large proportion of repetitive sequences [25,26]. STRs, which are also known as microsatellites or simple sequence repeats (SSRs), are among the most variable DNA sequences in the human genome and are therefore very suitable for DNA analysis in criminal investigations. STRs were identified in the early 1980s [27,28]. They consist of mono-, di-, tri-, tetra-, penta- and hexanucleotide repeat and are spread over the entire human genome with the majority being located in non-coding regions, either in introns or intergenic sequences [29]. More than 100 000 regions of the human genome contain STRs [30]. An individual is either homozygous (having identical alleles, i.e. the same number of repeats) or heterozygous (having different number of repeats) at a particular locus. The mutation rate of STRs differs between different loci and is dependant on repeat number, repeat type or the composition of the STR. The main proposed mechanism for new mutations in STRs is replication slippage [31,32].

STRs found in coding regions are generally trinucleotide repeats but hexanucleotide repeats can also be found [29]. A subset of these trinucleotide repeats play an important role in some human neurodegenerative disorders such as Huntington's disease [33] and also in some human cancers [34]. Huntington's disease is a dominant inherited monogenic disease caused by the abnormal expansion of a CAG repeat in the coding region of the Huntingtin

gene. The cause of this abnormal expansion is not completely understood but could be explained by strand slippage or multiple events of recombination [35].

STRs can be divided into different groups depending on their repeat structure. Simple or perfect repeats have the same length and sequence in every repeat unit (e.g. TATC). Compound or imperfect repeats contain stretches of two or more different repeat types (e.g. TCTATCTG). Complex or interrupted repeats have several blocks of repeats with different unit lengths but also contain intervening sequences [25].

STRs in forensic DNA typing

Not all STRs are optimal for forensic DNA analysis and there are some criteria the markers have to meet to be considered suitable. The STRs should be inherited independently of other markers that are analysed and should preferably be located on separate chromosomes. The loci should be highly polymorphic with a high degree of heterozygosity (>70%) and demonstrate low stutter characteristics [18,36]. Tetranucleotide repeats are preferred for genotyping in forensic DNA analysis because they are less prone to the formation of stutter products. Short dinucleotide repeats are very prone to slippage during PCR amplification [36,37]. This can give rise to peaks that are two, four or six bases shorter than the original allele, which in turn creates difficulties in interpreting the data, especially if DNA from multiple individuals is present. For tetranucleotide repeats, there is often only one shorter allele of four bases observed [37]. The use of STRs for forensic DNA typing was reported in the early 1990s [38,39]. The Forensic Science Service (FSS) in the UK and the Federal Bureau of Investigation (FBI) in the U.S. have been the leading countries in the development of STR typing systems. In 1997, FBI presented a database named CODIS (Combined DNA Index System) that consisted of 13 autosomal core loci suitable for forensic STR typing [40] (Table 1). These 13 loci are highly polymorphic, found in non-coding regions and located on different chromosomes, with the exception of CSF1PO and D5S818, which are positioned on chromosome 5. However, since approximately 26.3 Mb separates the two markers, no linkage between the loci is found [41]. The database includes DNA profiles of convicted offenders, from crime scene samples, missing persons and allele frequency data from different U.S. population groups [40]. In Sweden and most European countries, ten loci overlapping with the CODIS core loci are analysed. The loci used in forensic DNA analysis do not encode any proteins, no human characteristics can be distinguished by the profile and there is no linkage to any diseases. However, it has been suggested that the locus TH01 may be linked to schizophrenia [42]. However, this finding was not confirmed by a follow-up study [43]. It should be noted that many of the core loci used in forensic STR typing have been used in linkage studies of human diseases and that the findings of such studies are often tentative [41].

A DNA profile refers to the genotype (i.e. the number of repeats in each allele of the analysed STR marker) of the suspect, the victim or the crime scene sample.

Table 1. *The 13 core set of STR loci included in CODIS with chromosomal location and repeat motif.*

Marker	Chromosomal location	Repeat motif
CSF1PO	5q33.1	TAGA
FGA	4q31.3	CTTT
TH01	11p15.5	TCAT
TPOX	2p25.3	GAAT
vWA	12p13.31	[TCTG][TCTA]
D3S1358	3p21.31	[TCTG][TCTA]
D5S818	5q23.2	AGAT
D7S820	7q21.11	GATA
D8S1179	8q24.13	[TCTA][TCTG]
D13S317	13q31.1	TATC
D16S539	16q24.1	GATA
D18S51	18q21.33	AGAA
D21S11	21q21.1	[TCTA][TCTG]

miniSTRs

Larger PCR fragments, >300 bp, can be difficult to amplify if the DNA is highly degraded and fragmented. Allelic drop-out or complete loss of signal are often observed for larger sized PCR products [44]. One solution to this problem is to reduce the amplicon sizes (<150 bp) by moving the PCR primers closer to the repeat region, creating so-called miniSTRs [8,45-47]. An advantage using miniSTRs is that the obtained DNA profiles can be compared to existing convicted offender profiles in national DNA databases. One disadvantage is the multiplexing capacity. Since smaller amplicons are used, there are fewer possibilities for adjusting the sizes of the amplicons labelled with the same fluorescent dye than in routine STR typing. Therefore fewer loci can be amplified simultaneously [8]. MiniSTRs were successfully used in the victim identification of the World trade center attack in 2001 [3]. The remains were extremely degraded due to the intense heat from the fire and several samples were collected months after the attack. Moreover, the DNA had been degraded and fragmented by bacteria and other environmental factors.

SNPs

SNPs are the most common genetic variation found in the human genome [22,48]. In the NCBI SNP database more than 23 million SNPs are reported (www.ncbi.nlm.nih.gov/SNP/; build 131).

In forensic genetics, autosomal SNP typing is useful for typing severely degraded samples and can be applied as an alternative to STR typing in complicated cases. The main advantage is that short amplicons (between 50-100 bp) can be created since the primers can be located close to the poly-

morphic site. However, SNPs are not as informative as STRs, due to their bi-allelic nature. Around 50-75 SNPs, have to be analysed in order to equal the high discrimination power achieved by an STR analysis of multiple loci [49,50]. Moreover, the selection of SNPs is important in forensic genetics. The European SNP*for*ID consortium has therefore developed a multiplex of 52 unlinked polymorphic nuclear SNPs suitable for identification of individuals with different population origin [51].

Routine forensic DNA analysis

Forensic DNA analysis is based on multiplex PCR amplification of 10-17 STRs after which the fragments are size separated and detected using capillary electrophoresis (CE) [52-55]. This method is commonly referred to as fragment analysis. The size of the fragments, seen as peaks in electropherograms, can be determined using an internal size standard that is added to each sample. Fluorescently labelled PCR primers are used to facilitate multiplexing and simultaneous detection of multiple markers. Furthermore, the length of the amplicons is adjusted to avoid size overlap between different STRs that are amplified using primers labelled with the same dye [39] (*Figure 1*). Commonly, four fluorescent dyes are used with the internal size standard being labelled with a specific dye in order to distinguish it from the other STRs to be analysed. The alleles are determined using an allelic ladder, which contains all of the known alleles for each of the analysed STR marker. Several commercial kits are available for forensic DNA analysis. These include loci that overlap with the CODIS core loci and the European Standard Set (ESS) (TH01, FGA, vWA, D3S1358, D8S1179, D18S51 and D21S11). Two recently released kits are the AmpFℓSTR® Identifiler® Plus PCR Amplification Kit (Applied Biosystems), which analyses 15 loci plus the amelogenin locus for sex determination and the PowerPlex® 16 HS system (Promega) which analyses 16 loci simultaneously including the amelogenin locus [56,57].

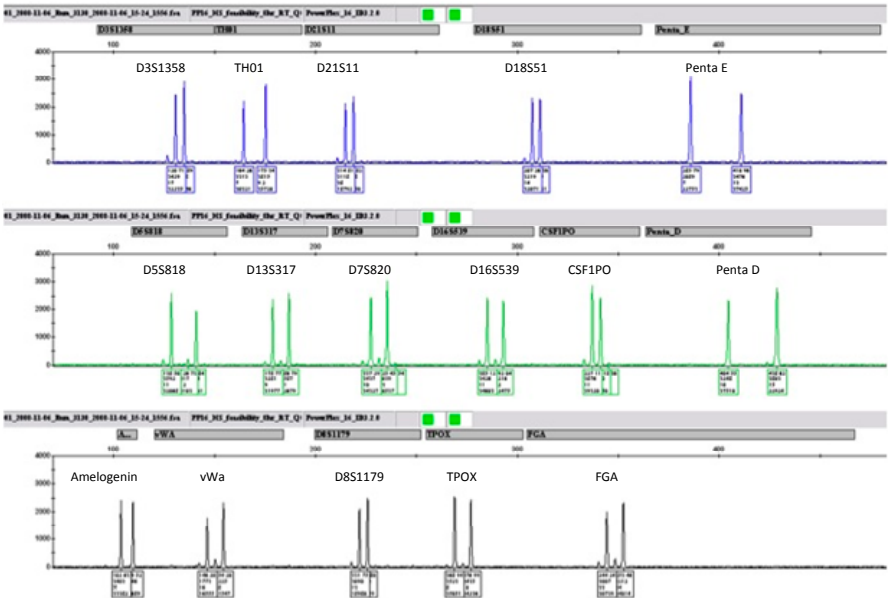

Figure 1. Electropherogram showing results of STR typing using the PowerPlex® 16 System. The y-axis shows the fluorescence intensity and the x-axis the size in bp. A total of 16 STR loci including the amelogenin locus used for sex determination are simultaneously amplified and detected by capillary electrophoresis. Three different dyes are used to distinguish the alleles. The electropherogram was provided by Promega Corporation (www.promega.com).

Match probability

When there is a match between DNA profiles from an evidence sample and a reference sample, the frequency with which that particular profile occurs in the population has to be estimated. The match probability is the probability that an unrelated, randomly selected individual in a population will have the exact same genotype observed in the sample. Since, STRs segregate independently during meiosis, the product of expected genotype frequencies (of all analysed loci) in a population can be calculated using the product rule [41]. The expected genotype frequencies are based on allele frequencies in a population using Hardy-Weinberg equilibrium principles.

National DNA databases

Many European countries have national DNA databases. These databases contain DNA profiles (STR-profiles) from convicted offenders (an intelligence database) as well as profiles from crime scene samples. The criteria for the entry and removal of profiles vary between different countries. In Sweden, new legislation was introduced in 2006 that allows the police to sample DNA from all individuals that are suspected on reasonable grounds of a crime that can lead to a term of imprisonment. If the individual is con-

victed, the profile is retained for ten years after release from prison. Before 2006, DNA could only be sampled from individuals that were suspected on reasonable grounds of a crime with a minimum sentence of two years. In the UK, all individuals that are arrested for a recordable offence (which includes most crimes aside from traffic offences) are sampled and registered regardless of the procedural outcome. The profiles are never removed from the database. The aim in establishing this large database was that it should cover the entire active criminal population of the UK [58]. The UK's DNA database is the largest in Europe, containing over 4 million DNA profiles in December 2009 (corresponding to 9% of the population) and ~350 000 crime scene sample (Table 2).

Table 2. *DNA profiles in the national DNA databases of Sweden, the UK, Denmark and Germany in December 2009 (www.enfsi.eu)*

Country	**Total number of individual profiles**	**% of the population**	**Crime scene sample profiles**	**Matches individual to crime scene sample**
Sweden	77 191	0.9	19 929	23 936
UK (England and Wales)	4 856 902	9.0	354 132	957 638
Denmark	56 323	1.0	34 068	13 672
Germany	668 721	0.8	166 554	73 078

Special forensic DNA analysis of challenging samples

Autosomal STR analysis is highly discriminative and sensitive. Using the 13 CODIS core loci the match probability for the theoretically most common profile is around 6.3×10^{-12} or 1 in 160 billion (among U.S. Caucasians) [59]. However, there are certain cases in which the STR analysis fails or is difficult to interpret due to the presence of mixtures of DNA or severe degradation of the sample. In these cases, special analyses based on markers on the Y chromosome or mtDNA can be performed.

The human Y chromosome

The human genome contains 23 chromosome pairs. Chromosomes 1 to 22 are referred to as the autosomes and the remaining are the sex-determining X and Y chromosomes. The sex chromosomes are unique in the sense that females have two copies of the X chromosome (XX) and males have one

copy of the X chromosome and one copy of the Y chromosome (XY). It is assumed that the sex chromosomes evolved from a pair of autosomes and evolution has made them genetically very different [60]. The X chromosome consists of ~155 Mb of DNA and contains around 1100 protein-coding genes while the Y chromosome is one of the smallest chromosomes in the human genome (~60 Mb) and contains only 78 protein-coding genes [61,62] (*Figure 2*). The repression in gene content and recombination are distinct features of Y chromosome evolution. Early in mammalian evolution, one of the autosomes obtained a male sex-determination function followed by accumulation of genes advantageous for males [63]. Selection for male specific alleles eventually resulted in the suppression of recombination and the non-recombining region of the Y chromosome (NRY) now covers 95% of the chromosome (it is also known as the male specific region (MSY)). The NRY contains 27 protein-coding genes, which are involved in sex determination (SRY) and spermatogenesis [61,64]. The remaining part of the Y-chromosome, the pseudoautosomal region (PAR) is located in the telomeric regions of the chromosome and recombines with its sex-specific counterpart during male meiosis [65]. Since almost the entire Y chromosome does not undergo recombination, mutations are the only force resulting in diversity. Otherwise, the chromosome is passed on from generation to generation in a paternal lineage. Therefore, the haploid Y chromosome is very useful in both evolutionary studies but also in forensic genetics [66-69].

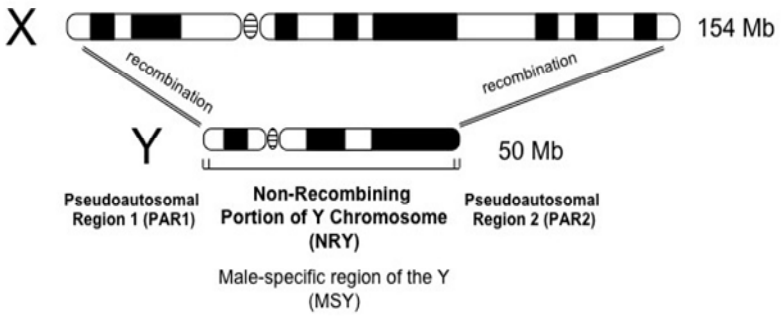

Figure 2. Overview of the X and Y chromosomes. Reprinted with permission from Elsevier Academic Press (Forensic DNA typing 2nd Edition – Biology, Technology and Genetics of STR markers, 2005).

Y-STR analysis

The Y chromosome is unique in the sense that it is male specific and this makes it a valuable tool in certain criminal investigations. Most sexual assault cases involve a male perpetrator. Therefore by using markers on the Y

chromosome, DNA mixtures containing high levels of female DNA and a minor proportion of male DNA can more easily be resolved [70,71]. Y chromosome analysis can also be used successfully used on azoospermic semen samples [72].

The Y chromosome contains a large proportion of repetitive sequences, convenient for forensic DNA analysis [63,69]. In forensic genetics a core set of eight Y-STR loci are analysed. This marker set is named the minimal haplotype and include DYS19, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS393 and DYS385ab [69]. Because autosomal markers segregate independently, the product rule can be employed when estimating the match probability for a certain DNA profile. However, the paternal inheritance and the lack of recombination in the NRY prevent the use of the product rule, since the markers are linked on the same chromosome as a haplotype. Y-STR analysis is therefore not as discriminative as autosomal STR typing and it is necessary to use a reference database to estimate the frequency of a certain Y-STR haplotype in a population. The largest Y-STR database today is the Y chromosome haplotype reference database (YHRD), which contains around 89 200 haplotypes from 693 populations (www.yhrd.org; release 34) [73,74]. Moreover, as with autosomal markers, commercial Y-STR kits are available which include the minimal haplotype loci as well as two loci (DYS438 and DYS439) recommended by the Scientific Working Group on DNA analysis methods (SWGDAM) [75-77].

The mitochondrial genome

The human mitochondrial genome is a circular double-stranded molecule comprising 16 569 bp of DNA and is located in the mitochondria in the cytoplasm. The mitochondrial genome was sequenced in 1981 [78] but was revised in 1999 due to some sequence errors and the current sequence is referred to as the revised Cambridge Reference Sequence (rCRS) [79]. The genome contains one heavy strand and one light strand with different base compositions. The genome is furthermore divided into the coding region and the non coding control region. The coding region contains 13 protein-coding genes, involved in oxidative phosphorylation (OXPHOS) as well as two rRNAs and 22 tRNAs essential for the translation of mtDNA-encoded mRNAs [78,80]. The oxidative phosphorylation pathway produces ATP, which is essential in various cellular processes. The control region, which is also known as the Displacement (D) loop, constitutes 7 % of the genome (around 1100 bp) and contains the origin of replication of the heavy strand and promoters for transcription.

The mitochondrion demonstrates a uniparental inheritance pattern analogous to the Y chromosome. However, mitochondria are maternally inherited [81]. Sperms contain around 100 copies of mtDNA, whereas the oozyte contains more than 100 000 mtDNA molecules [82]. One proposed mechanism for

maternal inheritance in mammals is that during fertilization the protein ubiquitin target paternal mitochondria, which result in degradation of sperm mitochondria inside the oozyte cytoplasm [83,84]. Due to this uniparental inheritance, there is no recombination between maternal and paternal genomes. Therefore, similar to the NRY on the Y chromosome, genetic variation in mtDNA is primarily due to mutations. The mutation rate of human mtDNA is about five to ten times higher compared to the nuclear genome [85]. This high mutation rate can be explained by the absence of protective histone proteins, but also by the lack of an efficient DNA repair system or by free radicals generated during the OXPHOS process in the mitochondria [80,85]. Another characteristic of mtDNA is the high copy number per cell. A mitochondrion contains 2-10 mtDNA molecules and each cell may contain up to 1000 mitochondria [86]. However, the number of mitochondria and mtDNA copies in a cell is dependant on the tissue type. This high copy number per cell is especially advantageous in forensic genetics as well as in ancient DNA analysis [10,87-91].

Individual haplotypes can be classified into different haplogroups that is based on specific sequence differences, more or less common between populations. The most common haplogroup in the Caucasian population is haplogroup H [92].

Forensic mtDNA analysis

Some forensic evidence found at crime scenes or in missing person investigations, such as shed hairs, fingerprints, severely burned samples and bone fragments, contain minute amounts of DNA. In these cases where both the quality and the quantity of the DNA are insufficient for a successful autosomal STR typing, an mtDNA analysis can be performed. MtDNA has some characteristics that are especially advantageous in forensic DNA analysis. Due to the high copy number of mtDNA in each cell, there is a higher chance of detecting mtDNA in a severely degraded DNA sample compared to nuclear DNA, which only occurs in two copies in each cell. Another advantage is that it is maternally inherited and thus reference samples can be obtained from maternal relatives.

Approximately 610 bp of the hypervariable regions 1 and 2 (HV1 and HV2) located in the control region are routinely amplified and sequenced in forensic mtDNA analysis. The resulting sequences are compared to the rCRS and only nucleotide differences to the rCRS are reported. The DNA commission of the International Society of Forensic Genetics (ISFG) has developed guidelines for forensic mtDNA analysis, which include safety precautions, recommendations regarding nomenclature as well as guidance for interpretation of the results [93]. To exclude the possibility that the reference and the forensic evidence sample originate from the same source, at least two nucleotide differences is required. If only one nucleotide difference is ob-

served, the result is deemed inconclusive. If identical mtDNA sequences are observed, it cannot be excluded that the samples originate from the same source [93]. When this occurs, the frequency of the particular mtDNA sequence or haplotype in a population is estimated. However, since mtDNA is maternally inherited and lack recombination, it is not as discriminative as an analysis of nuclear STR markers and this is the major limitation of mtDNA analysis. Consequently, the product rule cannot be used to calculate the match probability and the frequency of an mtDNA profile in a population require a population database of mtDNA sequences. The EDNAP (European DNA profiling) mtDNA Population database (EMPOP) contains almost 11 000 mtDNA haplotypes (www.empop.org; version 2.1, August 2010) [94]. Thus, the statistics regarding an mtDNA analysis is not only limited by the maternal inheritance pattern but also by the number of haplotypes in the database. Nevertheless, an mtDNA analysis has the same exclusion value as nuclear STR typing. One way to increase the discrimination is to analyse variations in the coding region [95-97].

Heteroplasmy

Heteroplasmy is the presence of multiple mtDNA types within the same individual. This occurs when a mutation takes place in one of the thousands of mtDNA genomes that are present in the mitochondria. The result is a mixture of normal mtDNA and mutant mtDNA (heteroplasmy), which are transmitted differentially by cell division [98]. Identification of heteroplasmy can complicate the interpretation of the results in forensic genetics. Therefore, guidelines for mtDNA typing and interpretation of heteroplasmy are established by the forensic genetics community [11]. However, even though mtDNA has a high mutation rate, heteroplasmy is rarely observed in forensic mtDNA typing. Heteroplasmy can also increase the match probability of a forensic mtDNA analysis. This was demonstrated in the identification of the Romanov family, during which it was found that tsar Nicolas II and his brother shared a heteroplasmic position [89,99].

DNA damage and contamination in skeletal remains

Many of the difficulties associated with forensic genetic analysis also occur in the study of ancient DNA. Degradation of DNA begins immediately after death. The DNA is fragmented by intracellular nucleases, which is one of the reasons why long PCR amplicons are difficult to amplify in degraded materials that are commonly found at crime scenes. The surrounding environment also influences the extent of degradation of the DNA molecules. Two crime scenes are never identical and the biological materials collected at different crime scenes are never exposed to exactly the same conditions. Heat, humidity, microorganisms and humic acids (i.e. a major organic component of the

soil, produced by the degradation of dead organic substances) all contribute to degradation of DNA.

Aside from nucleases, the DNA in aged skeletal remains may be subjected to chemical modifications such as oxidative- and hydrolytic damage [5-7]. Hydrolytic deamination of cytosines to uracil results in miscoding lesions in which the cytosine is changed into a thymine (C-T) or a guanine is changed to adenine (G-A) during amplification. This is one of the major contributors to DNA damage in post-mortem samples such as old skeletal remains. However, C-T substitutions can be reduced by treatment with uracil N-glycosylase (UNG), which eliminates uracil from the DNA strand and thereby prevent introduction of the T nucleotide in these positions during PCR [7].

Another major challenge when analysing minute amounts of DNA from old skeletal remains is the risk of modern contamination that can question the authenticity of the results. Exogenous DNA can be introduced at several stages and it is impossible to know how many individuals that may have handled the samples over the years. DNA analyses of old human skeletal remains are often based on mtDNA. However, due to the maternal inheritance pattern and the limited variation seen in populations, the analyst may share an mtDNA profile with the analysed bone sample by chance. Thus, when analysing old human remains there is always the possibility of false positive results and it is therefore important to critically evaluate all aspects of the results of such analyses [100]. In addition to removing of the surface of the bone and taking the extensive safety precautions previously discussed, it has been shown that soaking the bones in commercial bleach (sodium hypochlorite, NaOCl) prior to DNA extraction is an effective decontamination method [101,102].

Technologies in forensic DNA analysis

Analysis by capillary electrophoresis

The routinely used technologies in forensic DNA typing (both for nuclear and mtDNA analysis) are based on fluorescence detection using CE [52-55,103]. As previously discussed, routine STR genotyping is based on fragment analysis with fluorescently dye labelled PCR primers. The markers are amplified in a multiplex reaction and thereafter size separated using CE. In CE, narrow sized capillaries are filled with a polymer solution that the samples can be run through. When a positive charge is applied, the negatively charged DNA molecules are transferred into the capillary. A laser beam at the end of the capillary causes the dye labelled fragments to fluoresce, which is detected by an optical device [104]. The dyes that are used for labelling the primers emit light at different wavelengths. Therefore, several fragments that

overlap in size can be analysed simultaneously as long as they are labelled with different dyes (*Figure 1*).

Sanger dideoxy sequencing

Dideoxy Sanger sequencing [105] has been the gold standard in DNA sequencing for almost three decades. The technology is based on hybridisation of a sequencing primer to a PCR product. This is followed by incorporation of deoxynucleotide triphosphates (dNTPs) and dideoxynucleotide triphosphates (ddNTPs), which lack a hydroxyl group at the 3'end. As a result, elongation of the synthesized strand is terminated when the polymerase incorporates a ddNTP. This results in a large amount of fragments that differ in length by one base. Today, fluorescently labelled ddNTPs are used and the fragments are size separated using CE [106,107]. The sequencing results are shown in a chromatogram (*Figure 3*). The sequencing of the human genome was accomplished by the fragmentation of DNA, which was then cloned into bacterial vectors and sequenced using the Sanger technology [22].

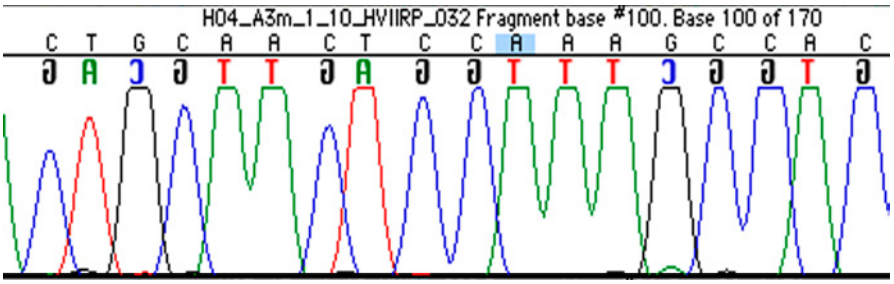

Figure 3. Chromatogram demonstrating the sequencing results of a part of the mitochondrial HV2 region of an ulna bone supposed to have been buried for approximately 70 years.

Pyrosequencing

Pyrosequencing is a non-electrophoretic sequencing technology based on real-time detection of Pyrophosphate (PPi) using an enzymatic cascade system that ultimately results in the generation of light [108]. Successful incorporation of a dNTP by DNA polymerase leads to the release of PPi. The PPi is converted to ATP by ATP sulfurylase, followed by generation of light when luciferase uses ATP to oxidize luciferin. The light is proportional to the number of incorporated nucleotides and unincorporated nucleotides as well as ATP are degraded by apyrase before the next nucleotide is added (*Figure 4*). The results are presented in a pyrogram, in which the ascending slope of the peaks reflects the activity of DNA polymerase and ATP sulfurylase. The activity of luciferase determines the height of the signal and the slope of the

descending curve is demonstrated by the efficiency of apyrase [109]. Pyrosequencing is suitable for SNP typing [110] and also for detection of mutations [111,112]. Deyde *et al.* used Pyrosequencing for the detection of drug resistance markers in the pandemic influenza A virus (H1N1) [113]. Other applications of Pyrosequencing have been in studies of DNA methylation [114,115] and typing of viruses [116,117]. In forensic genetics it has been applied in mtDNA sequencing [118], in quantification of mixtures [119], mtDNA coding region analysis [97], amelogenin-based sex determination [120] and in nuclear STR typing as demonstrated in Paper I and II in this thesis [121,122]. One advantage of Pyrosequencing when compared to Sanger sequencing is that it can read DNA in the region directly after the sequencing primer.

One of the limitations of the technology is the limited reading length of the nucleotide sequence. Insufficient polymerase activity is one factor limiting read length in Pyrosequencing [123]. However, the addition of single-stranded DNA binding protein (SSB) has been shown to significantly improve the read length and to reduce non-specific signals, especially in longer PCR products [123,124]. In addition, background signals observed in the later dispensations in the pyrogram, disturb the interpretation of correct signals and reduce the read length. The intensity of the signal peaks also decreases as sequencing proceeds because each nucleotide dispensation result in increases of the volume and a reduction of the efficiency of the enzymes [109]. Mashayekhi *et al.* show that the efficiency of apyrase decrease in later dispensations, due to the dilution effect caused by the increase in reaction volume of 0.07% after every nucleotide dispensation. This causes inefficient degradation of nucleotides and ATP and is therefore one of the main factors restricting the read length [125]. Mashayekhi *et al.* also present a simulation model to increase reading length by replacing the apyrase in the reaction mix with a washing step between every nucleotide dispensation and thereby remove inhibiting by-products [125].

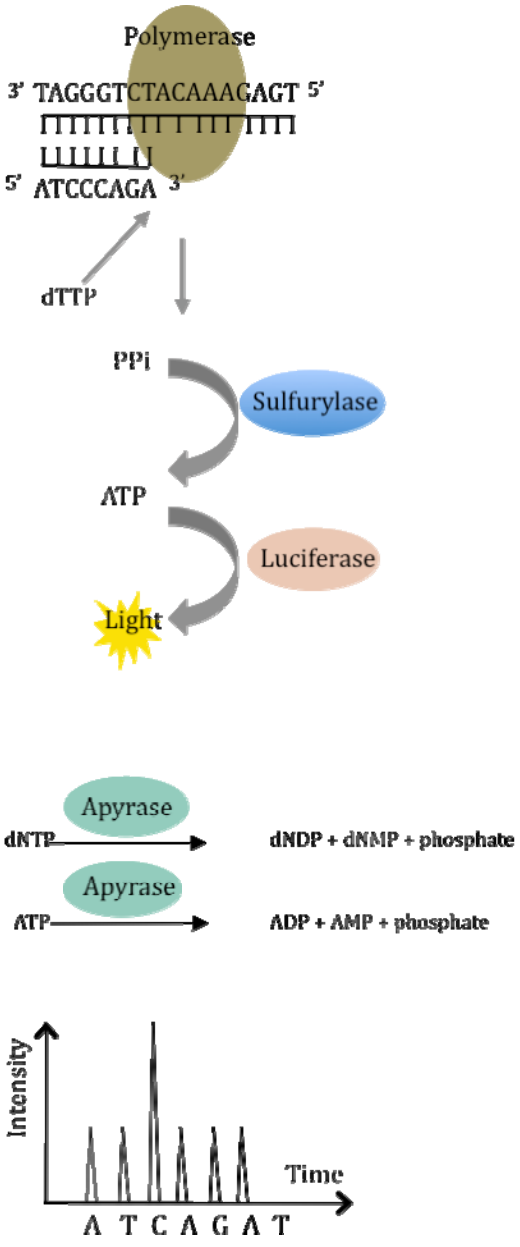

Figure 4. The principle of Pyrosequencing technology. The real-time detection of Pyrophosphate is based on incorporation of dNTPs by polymerase, which triggers a cascade system of enzyme reactions that ultimately generate light.

454

The Pyrosequencing principle is adapted in the 454 (Roche) whole genome sequencing technology [126]. 454 sequencing is one of the NGS technologies along with for instance the SOLID™ system (Applied Biosystems) and the Solexa Genome Analyzer II (Illumina). NGS technologies allow for genome wide sequencing with a massive increase in throughput and are highly suitable for *de novo* sequencing and resequencing [127]. In 454 sequencing, the DNA is fragmented followed by ligation to adapters before separation into single strands. The fragments are then bound to a bead, which contain an emulsion of water, oil and detergent, which forms a droplet around the bead. The DNA fragment in each droplet is thereafter amplified and the emulsion is broken with subsequent denaturation of the DNA strands. The beads, containing amplified single stranded DNA are added into picolitre-sized wells on a fibre-optic slide containing approximately 1.6 million wells and thereafter sequenced using the Pyrosequencing chemistry [126]. The Genome sequencer FLX system generates one million reads per run and each sequence read is on around 300-400 bp. The 454 sequencing method has successfully been used in studies of ancient DNA. A complete mitochondrial genome sequence of a 38 000 year old Neanderthal individual has been assembled as well as nuclear DNA sequencing of Neanderthals [128-130].

DNA quantification using real-time PCR

The commercially available STR typing kits recommends 0.5-1 ng of template DNA for successful amplification and detection of fragment lengths [131]. Too low quantity of DNA in the STR analysis can result in allele drop out, peak imbalances and low signals, especially in the longer fragments [12]. Too much input DNA can result in over expressed peak signals as well as background signals. Therefore, it is important that the amount of DNA in a forensic sample is accurately determined, both, to save precious evidence material and to select suitable markers for subsequent DNA analysis. Andréasson *et al.* have developed a system for quantification of nuclear- and mtDNA suitable for forensic samples using real-time PCR based on the TaqMan® assay [132]. In the TaqMan® assay, a probe is labelled with a reporter fluorophore that is attached to the 5' end and a quencher fluorophore in the 3' end. The 5'-3' exonuclease activity of the Taq polymerase cleaves the probe during the elongation phase of the PCR, releasing the reporter from the probe. This results in an increased reporter emission intensity that is detected by a charge coupled device (CCD) camera, since the proximity to the quencher is lost [133-135]. At a certain cycle the fluorescence reaches a threshold, also called threshold cycle (C_t). The C_t value is inversely proportional to the DNA concentration in the target, i.e. low levels of DNA results in a high C_t value. The DNA quantification of a sample is possible by the use of a standard curve with known concentrations of either nuclear DNA or mtDNA.

Present investigation

Aim

The aim of the papers presented in this thesis is focusing on evaluation of the Pyrosequencing technology for forensic DNA typing and efficient methods for mtDNA analysis of old skeletal remains.

Paper I

Y chromosomal STR analysis using Pyrosequencing technology

Background

DNA analysis of markers located on the Y chromosome is valuable in certain cases, particularly in sexual assault cases, in which evidence often contains a mixture of DNA from both a female victim and a male perpetrator. The forensic community has agreed on a core set of eight Y-STR markers, named the minimal haplotype (DYS19, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS393 and DYS385ab) to use in routine Y-STR analysis [69]. The routine Y-STR analysis is based on multiplex amplification of the minimal haplotype loci as well as the loci recommended by SWGDAM (DYS439 and DYS439). The amplified fragments are then size separated using CE. The technology is very robust and accurate, however, the commercially available kits require long fragments of up to 450 bp, to enable electrophoretic size separation. These longer fragments can be difficult to amplify if the evidence material is old and degraded. Therefore, the forensic community is in need of new efficient, robust and fast techniques that can be used on samples with very scarce amounts of DNA. Pyrosequencing is a non-electrophoretic DNA sequencing technology initially developed for analysis of short stretches of DNA [110,136]. In previous studies it has been used for sequencing mtDNA in forensic genetics [118] and for sex determination by analysis of short stretches of the amelogenin gene [120].

In paper I we demonstrate a novel Pyrosequencing-based assay for analysis of Y-STRs.

Results and discussion

The Pyrosequencing technology was used for analysis of seven minimal haplotype Y-STRs, DYS19, DYS389 I/II, DYS390, DYS391, DYS392, DYS393 and DYS438 (SWGDAM). PCR primers were designed creating amplicons between 72-233 bp. A total of 70 unrelated male individuals were typed and all loci displayed male specificity. The alleles were assigned by sequencing using a sequence directed dispensation order. Nucleotides were added to the reaction in the order in which they appear in the repeat unit. In this way, as soon as the growing strand reached the end of the STR, the dispensation order no longer matched the template sequence and reaction terminated. The sequence prior to the termination displays the number of repeat units and the length of the allele. Alleles were assigned for all of the loci examined and the allele frequencies among the 70 male individuals were determined. The gene diversity ranged from 0.386 to 0.734. According to the YHRD database (3.0) the haplotype most commonly observed in the study occurs with frequency of 0.93% (worldwide). The major advantage of this assay is its ability to detect sequence variants within or in proximity to the repeat, which may increase the discrimination capacity. This cannot be done with standard fragment analysis using capillary electrophoresis. For example, at the DYS393 locus an A/C SNP was detected in the first repeat unit in 14.3% of the analysed individuals (*Figure 5*). This resulted in one repeat unit less ($(AGAT)_{13}$ to $CGAT(AGAT)_{12}$). The system is sensitive and analysis of casework samples containing between 15 pg and 0.4 ng DNA were successfully amplified and sequenced.

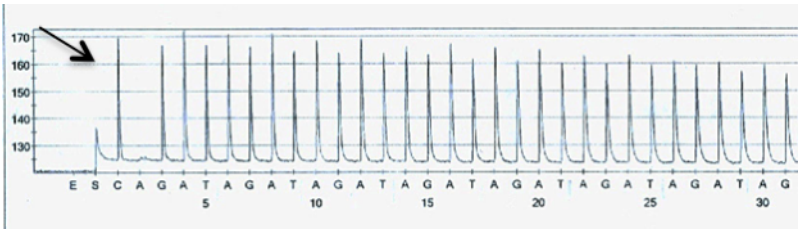

Figure 5. Pyrosequencing result of the DYS393 locus, demonstrating the A/C SNP in the first repeat unit (dispensation 1).

In conclusion, the Pyrosequencing assay for the analysis of Y-STRs can be useful in certain cases as well as for evaluation of new markers suitable for forensic DNA typing and fast database assembly.

Paper II

Forensic analysis of autosomal STR markers using Pyrosequencing

Background

STRs are routinely genotyped by length separation in forensic DNA analysis. Their abundance and selectively neutral nature makes them useful in many areas of genetic research. Due to their high degree of polymorphism at each locus they are very informative and are routinely used for individual identification in forensic genetics. A sequence-based assay could reveal additional information and an increased resolution due to sequence variants that can be detected within or in close proximity to the repeat. This cannot be observed by the routine fragment size analysis. Moreover, a rapid and robust sequence based system for genotyping STRs can be useful for rapid construction of databases and for the evaluation of new loci suitable in forensic STR typing.

In Paper II we present a novel strategy for genotyping autosomal STRs based on the Pyrosequencing technology.

Results and discussion

A total of ten markers routinely used in forensic DNA analysis were selected for use in testing the Pyrosequencing assay (CSF1PO, TH01, TPOX, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539 and Penta E). PCR primers were designed to generate amplicons in the range of 66-178 bp. In total, 114 Swedish individuals were genotyped together with a few forensic samples and allele frequencies, observed and expected heterozygosity were calculated. The results obtained were verified by fragment analysis. Alleles were assigned using a strategy similar to that described in Paper I. The termination recognition base (TRB) and the use of a sequence directed dispensation order give rise to a specific pattern in the downstream flanking sequence. Heterozygous alleles are possible to identify because the signal is decreased by 50% when the shorter allele terminates. The pyrograms arising from short simple repeats were more straightforward to interpret compared to ones arising from more complex repeats. The analysis of the compound repeats D3S1358 and D8S1179 commonly resulted in two possible genotypes that were impossible to differentiate. This was observed in 31% of the genotypes at the D8S1179 locus and in 36% of the genotypes at the D3S1358 locus. In all of these cases, one of the genotypes was concordant with fragment analysis. Moreover, sequence variants were identified at five loci. At the D13S317 locus a T/A SNP in the last repeat unit was detected in 92% of the typed individuals, resulting in the genotype 12/12[AATC] in the Pyrosequencing assay and 12/13 in the fragment analysis.

To conclude, this Pyrosequencing-based assay has the ability to detect sequence variants within or close to the repeat, resulting higher discrimination capability and can be useful for evaluation of less complex novel STRs and for rapid compilation of databases.

Paper III

Analysis of the putative remains of a European patron saint- St. Birgitta

Background

Saint Birgitta (Saint Bridget of Sweden) lived between 1303 and 1373 and was canonized as a Roman catholic saint in 1391. Following her death, her remains were transported from Rome to Vadstena and placed in a relic casket together with her daughter Katarina (1331-1381).

The approaches and methods used for analysis of DNA from ancient remains are very similar to those when analysing samples from crime scenes. Analysis of mtDNA is used in special criminal investigations when nuclear autosomal DNA typing is not possible to perform due to severe degradation of the DNA or lack of nuclear DNA. DNA in skeletal remains is often found in very scarce amounts and has moreover often been subjected to various environmental factors and chemical agents. This results in fragmentation of the DNA and base modifications that can interfere with the PCR reaction. In addition to degradation and fragmentation of the DNA, is the high risk of modern contamination. Analysis of mtDNA is commonly performed on aged skeletal remains due to the high copy number per cell, which increases the likelihood of successful amplification.

In Paper III, an anthropological and DNA based analysis of the putative remains of Saint Birgitta and her daughter Katarina is performed. The maternal relationship between the two sets of remains is investigated using mtDNA analysis and radiocarbon dating is also performed.

Results and discussion

The HV1 (221bp and 440 bp fragments) and HV2 (243 bp and 415 bp fragments) regions were amplified and sequenced. The bones were extracted with two different approaches of the initiating step of the extraction, including a soaking step in bleach in one approach. A sex determination analysis was performed based on sequencing of a short stretch of the amelogenin gene utilising the Pyrosequencing technology. The sequences obtained from both skulls demonstrated that they were of female origin. The results of mtDNA sequencing revealed differences between the skulls at six positions (five in HV1 and one in HV2) (Table 3). As a safety precaution, to avoid

contamination and sample mix up, two different analysts performed the DNA analysis at separate occasions. The mtDNA of the analysts were compared with the result obtained from the skulls and one of the analysts demonstrated a sequence result that was inconclusive. Therefore, in order to exclude that the sequence result was due to contamination from the analyst, a coding region analysis of position 3010G/A and 16519T/C was performed. An additional difference in 3010 between the analyst (3010G) and skull A (3010A) could be observed, excluding contamination.

During the analysis the skulls showed differences in the number of successful amplification reactions and skull B demonstrated a better quality of the sequences. In the mtDNA quantification based real-time PCR, skull B contained a higher mtDNA yield compared to skull A. To investigate if the skulls could be of different age due to the difference in sequence quality and quantity a radiocarbon dating was performed. Skull A was dated to the period 1215-1270 cal AD and skull B was dated to 1470-1670 cal AD, which do not correspond to the periods Saint Birgitta or Katarina were alive. However, it was necessary to consider the potential consequences of a reservoir effect on the precision of theses dates. Radiocarbon dating alone can cause a shift in age when studying the remains of individuals who consume large amounts of food from freshwater sources such as fish, resulting in an older radiocarbon age. Therefore, the natural mass fractionation of the stable isotopes $\delta^{13}C$ and $\delta^{15}N$ was measured. When viewing the results from this, it is not possible to completely exclude that skull A is from the 14th century. If skull A is from St. Birgitta, her diet would have been extensively dominated by fish, which is questionable according to the historical records concerning traditions in medieval Sweden.

In conclusion, the radiocarbon dating of the skulls result in a dating period that do not correspond to the time period St. Birgitta and her daughter lived. Moreover, the mtDNA analysis reveals a non-maternal relationship.

Table 3. *Sequencing results of the HV1 and HV2 regions from skull A (Saint Birgitta) and B (Katarina).*

	HV1					**HV2**
	16126	16189	16294	16296	16304	73
rCRS[1]	T	T	C	C	T	A
Skull A	T	C	C	C	T	A
Skull B	C	T	T	T	C	G

[1]rCRS, revised Cambridge Reference Sequence

Paper IV

DNA extraction and analysis of skeletal remains

Background

There are several efficient extraction methods for the recovery of DNA from skeletal remains. However, because the conditions under which the remains have been preserved vary in a large extent, the quality and quantity of the DNA can be very different.

Many extraction methods are based on silica-binding [137-141] or phenol-choloform [89,91,142]. Moreover, ethylene diamine tetra-acetic acid (EDTA) is often used to decalcify of the bones and proteinase K is added to digest proteins [143,144]. One major problem with old skeletal remains is the high risk of exogenous contamination. Different approaches to remove or destroy contamination have been investigated and soaking the bones in commercial bleach can be efficient for decontamination [101,102]. A perfect extraction method would be generally applicable in studies of skeletal remains, and would give high yields of DNA while removing PCR inhibitors as well as contaminants. However, many extraction protocols use different purification procedures and different concentrations of chemical agents, which indicates that no leading extraction method is preferred.

In paper IV, the evaluation of an efficient DNA extraction method for bone samples, supposed to have been buried for approximately 70 years, are investigated.

Results and discussion

The efficiency of DNA extraction based on a salting out procedure was investigated, in which three different pre-extraction protocols were evaluated (Table 4).

Table 4. *Three different initiation protocols based on addition of bleach, EDTA, proteinase K and SDS were evaluated.*

Extraction	**Bleach**	**EDTA**	**Proteinase K**	**SDS**
A	+	+	+	-
B	-	+	+	+
C	-	+	+	-

The number of mtDNA copies in extracts from a skull and an ulna were determined, and the success rate of the amplification and sequencing was evaluated. The protocols involved treating whole or pulverised bone samples

with two different concentrations of EDTA and proteinase K. One of the extraction protocols included a soaking step in commercial bleach (NaOCl). A sex determination based on sequencing of the part of the amleogenin gene was performed. The presence of a 6 bp deletion on the X-chromosome, observed in the sequencing results, established that both the skull and ulna were of female origin. MtDNA quantification was performed according to Andreasson *et al*. [132] and 0-14 110 mtDNA copies/100mg were obtained in the different extracts. The extracts were quantified without dilution and at dilutions in 1:10 and 1:20. The yield of mtDNA from the skull was considerably lower than that from the ulna, which probably was due to the more porous state of the skull and less compact bone material. The highest mtDNA yield, in extracts from the ulna, was observed in using protocol A (0-5200 mtDNA copies/100mg), whereas protocol C yielded the highest amount of mtDNA copies in extracts from the skull (0-1100 mtDNA copies/100mg). A total of 120 PCR reactions were performed on 20 ulna extracts (eight obtained using extraction protocol A and B respectively and four using extraction protocol C) (three dilutions and amplification of HV1 and HV2 region respectively) and a total of 96 PCR reactions were performed on 16 skull extracts. Using protocol C, the highest success rates in the amplification and sequencing were obtained with pulverised samples in both the skull and ulna. Furthermore, using extraction protocol C, 25% of the undiluted extracts resulted in detected amounts of mtDNA. By contrast, 92% of the undiluted extracts from extraction protocol B gave successful result in the amplification and sequencing, indicating that higher concentrations of EDTA (0.5M in protocol A and C compared to 6mM, in protocol B) result in inhibitory effects in the PCR. In total, 96% of the extracts demonstrated an identical mtDNA sequence profile (263G). Soaking pulverised skull samples in bleach resulted in reduced yields of mtDNA, and only 8% of these extracts were successfully amplified and sequenced. Moreover, extracts prepared using protocol B and C demonstrated signs of contamination, whereas no contamination was detected in extracts prepared using protocol A. One extract prepared using protocol B demonstrated a male profile in the sex determination analysis. In addition, the same extract contained a higher number mtDNA copies compared to the other extracts from extraction B as well as a sequence difference (150T). Thus, soaking the bones in bleach is a potential method for decontamination of old skeletal remains prior to DNA extraction, especially whole bone samples.

In this study we investigated DNA isolation, based on different treatments prior to an extraction that was based on a salting out method. The protocols can be utilised on whole as well as pulverised bone samples that in this case had been buried for approximately 70 years. Moreover, a soaking step in commercial bleach prior to the extraction can be useful for decontamination, but it will also reduce the DNA yield.

Concluding remarks and future perspectives

Forensic DNA analysis has become an indispensable tool in criminal investigations. A forensic DNA analysis can tie an individual to a crime scene by analysing forensic evidence, but just as important is the use of a forensic DNA analysis for exoneration of innocent individuals.

This thesis is focusing on evaluation of sensitive tools to implement in several areas in forensic genetics including efficient methods for analysis of old skeletal remains. Two Pyrosequencing based systems for the analysis of autosomal- as well as Y chromosomal STRs are presented. Markers on the Y chromosome are particularly useful in sexual assault cases, in which there often are mixtures of female and male DNA. Evaluation and addition of new polymorphic loci is furthermore important to improve forensic DNA analysis. In the most recent kits provided by the commercial companies, three miniSTR loci and two polymorphic loci have been included in addition to the European Standard Set (ESS) [145]. MiniSTRs can increase in the success rate when typing highly degraded samples, since the amplicons are shorter than 200 bp. In Paper I and II, the majority of the amplicons were designed to be shorter than 200 bp for optimal analysis of degraded samples. The sequence analysis based on the Pyrosequencing technology described in this thesis offer a rapid and robust strategy for evaluation of new loci as well as detection of sequence variants within and near the repeat.

Forensic samples are often present in minute amounts, highly degraded, can contain inhibitors and exogenous contamination. Old skeletal remains often present all of these challenges. Thus, old skeletal remains provide the ideal test for the assessment of novel method's sensitivity. In some cases the DNA is so degraded that analysis of mtDNA is the best alternative. The maternal inheritance of mtDNA makes it convenient for analysis of maternal relationships as well as for individual identification in mass disasters or missing person investigations. In this thesis a maternal relationship between the putative skull of St. Birgitta and her daughter Katarina was investigated and excluded. Skeletal remains are often exposed very diverse and harsh conditions and the extent of degradation and fragmentation of the DNA between samples can differ substantially. The variety of different conditions that remains are kept in also affects the quantity of DNA. It is therefore highly beneficial to have efficient DNA extraction methods that recover sufficient amounts of DNA for a subsequent molecular analysis. An optimal DNA extraction method also removes inhibitors as well as possible contami-

nation. In Paper IV, the efficiency in DNA recovery, using different protocols, from samples that have been buried for approximately 70 years was estimated. Our result indicates that high concentrations of EDTA are beneficial and that both whole as well as pulverised bone samples can be decontaminated using commercial bleach.

The field of forensic genetics is facing an interesting future. The rapid developments in massively parallel sequencing technologies for molecular analysis present intriguing possibilities in medical genetic research. The high throughput NGS technologies provide an enormous capacity in the generation of sequence data. In the future, resequencing of genomes will facilitate the study of for instance CNVs and rare genetic variants associated with complex human diseases. The implementation of NGS in forensic genetics will be interesting, but requires development of software as well as suitable bioinformatics tools to correctly handle the amount of data.

One interesting aspect in forensic research is the prediction of physical traits. The ability to predict a suspect's eye- or hair colour can provide important information in criminal investigations, especially in cases where the suspect is completely unknown. Strong association between blue/brown eye colour and certain SNP loci in the *HERC2* gene have been demonstrated [146-148]. However, the genetics behind physical appearance constitute a complex pattern of multiple genes [149]. Hence, more work is needed to identify markers suitable for the prediction of physical traits useful in forensic genetics [150].

Another interesting field in forensic genetics is the identification of body fluids. Determination of a body fluid present in a crime scene stain can provide valuable information to a criminal investigation. Juusola and Ballantyne describe a promising system for body fluid identification that is based on mRNA profiling [151-153].

Biological evidence found at crime scenes constitutes some of the most important forensic evidence in criminal investigations of the 21st century. New technologies are important for the progression of the field. The systems presented in this thesis provide rapid and accurate evaluation of novel less complex STR loci suitable for forensic DNA typing as well as fast database assembly. Moreover, the benefits of mtDNA analysis and improved DNA extraction has been presented based on analysis of skeletal remains aged between 100 and 700 years exposed to a variety of degrading conditions.

Acknowledgements

The work presented in this thesis was performed at the Department of Genetics and Pathology, Rudbeck Laboratory at Uppsala University.

There are many people I would like to thank for making this thesis possible:

My supervisor, *Marie Allen*, for believing in me and giving me the chance to start my degree project in the forensic group and continue as a PhD student. My co-supervisor, *Ulf Gyllensten,* for interesting discussions at meetings and seminars.

All present and former members of the forensic group: *Martina*, my "sawing/drilling in the basement partner" ☺, for all talks regarding everything between forensics and life. It is definitely to few discussions about last night's CSI these days. Also for all the valuable comments when reading this thesis. *Anna-Maria,* for introducing me to Rudbeck and forensics. *Hanna,* for all the fun and for making the office look so much better. I still have the knobs on the desk where you used to sit. *Charlotte*, for fun conversations. *Maria M,* for organising and taking care of everything at work so nicely when I was away on maternity leave. The present girls: *Maria L*, for your enthusiasm and interest in the forensic field. *Anki,* for always being at work before me. It is nice knowing that I am not the only one getting up before six in the morning. *Mia,* for our endless talks about children ☺. *Viktoria*, for all the help in the lab during the summer, and all other students.

The Uppsala Genome Center, especially *Inger J* and *Joanna* for all the help with GeneMapper. All other colleagues at Rudbeck, for creating a nice atmosphere and for pleasant talks in the corridor, lab or in the lunchroom, making it a fun place to work at, *Emma, Ann-Sophie, Anna W*, *Inger G, Jessica, Lucia, Ludmila E, Marie S, Anna-Karin* and *Malin E* and all others that have worked at floor 2 or 3 between 2003 and 2010.

All members of the PhD student council at IGP for fun meetings and all administrative personnel at Rudbeck.

The Kjell and Märta Beijer Foundation for the financial support.

Finally, there are some important people I would like to thank outside of Rudbeck who have contributed to this thesis in a different way:

My best friend *Pernilla*, for all crazy and wonderful things that have happened during our friendship.

My friends in Uppsala, *Marja*, for all the fun at Norrland's and all other times we have shared during the past eleven years, both happy and sad. *Ida*, for all the fun and for being such a good friend throughout the years.

My mother, *Britt-Inger*, for always believing in me and for all the help with Wille and Vidar during the time writing this thesis. My father, *Lennart*, for always encouraging me, without you I would not have been where I am today. *Jonas*, the best brother in the world. For our (almost) every day talks, over the telephone.

Johan, the love of my life. For taking care of the boys when I have been at work (by far to many evenings). Just for making me laugh every day. My two sunshines: *Wille*, my own little proof that life can never be controlled and be absolutely perfect anyway and *Vidar*, the other joy of my life. Every day with you two is an adventure. But most importantly, you remind me what is important in life. I am so lucky to have the three of you in my life.

References

1. Styrman, L.P.G.L. Personal Communication. (Gävle, 2010).
2. Mardis, E.R. The impact of next-generation sequencing technology on genetics. *Trends Genet* **24**, 133-141 (2008).
3. Holland, M.M., Cave, C.A., Holland, C.A. & Bille, T.W. Development of a quality, high throughput DNA analysis procedure for skeletal samples to assist with the identification of victims from the World Trade Center attacks. *Croat Med J* **44**, 264-272 (2003).
4. Tsokos, M., Lessig, R., Grundmann, C., Benthaus, S. & Peschel, O. Experiences in tsunami victim identification. *Int J Legal Med* **120**, 185-187 (2006).
5. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715 (1993).
6. Hoss, M., Jaruga, P., Zastawny, T.H., Dizdaroglu, M. & Paabo, S. DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Res* **24**, 1304-1307 (1996).
7. Hofreiter, M., Jaenicke, V., Serre, D., Haeseler Av, A. & Paabo, S. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res* **29**, 4793-4799 (2001).
8. Butler, J.M., Shen, Y. & McCord, B.R. The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J Forensic Sci* **48**, 1054-1064 (2003).
9. Wilson, I.G. Inhibition and facilitation of nucleic acid amplification. *Appl Environ Microbiol* **63**, 3741-3751 (1997).
10. Paabo, S. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Natl Acad Sci U S A* **86**, 1939-1943 (1989).
11. Bar, W., Brinkmann, B., Budowle, B., Carracedo, A., Gill, P.*, et al.* DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. *Int J Legal Med* **113**, 193-196 (2000).
12. Gill, P., Whitaker, J., Flaxman, C., Brown, N. & Buckleton, J. An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forensic Sci Int* **112**, 17-40 (2000).
13. Budowle, B., Eisenberg, A.J. & van Daal, A. Validity of low copy number typing and applications to forensic science. *Croat Med J* **50**, 207-217 (2009).
14. Gill, P. Application of low copy number DNA profiling. *Croat Med J* **42**, 229-232 (2001).
15. Irwin, J.A., Leney, M.D., Loreille, O., Barritt, S.M., Christensen, A.F., Holland, T.D., Smith, B.C. & Parsons, T.J. Application of low copy number STR typing to the identification of aged, degraded skeletal remains. *J Forensic Sci* **52**, 1322-1327 (2007).
16. Jeffreys, A.J., Wilson, V. & Thein, S.L. Individual-specific 'fingerprints' of human DNA. *Nature* **316**, 76-79 (1985).

17. Gill, P., Jeffreys, A.J. & Werrett, D.J. Forensic application of DNA 'fingerprints'. *Nature* **318**, 577-579 (1985).
18. Reynolds, R., Sensabaugh, G. & Blake, E. Analysis of genetic markers in forensic DNA samples using the polymerase chain reaction. *Anal Chem* **63**, 2-15 (1991).
19. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* **51 Pt 1**, 263-273 (1986).
20. Watson, J.D. & Crick, F.H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738 (1953).
21. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C*., et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
22. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J*., et al.* The sequence of the human genome. *Science* **291**, 1304-1351 (2001).
23. Kruglyak, L. & Nickerson, D.A. Variation is the spice of life. *Nat Genet* **27**, 234-236 (2001).
24. Feuk, L., Carson, A.R. & Scherer, S.W. Structural variation in the human genome. *Nat Rev Genet* **7**, 85-97 (2006).
25. Weber, J.L. Informativeness of human (dC-dA)n.(dG-dT)n polymorphisms. *Genomics* **7**, 524-530 (1990).
26. Ellegren, H. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* **5**, 435-445 (2004).
27. Miesfeld, R., Krystal, M. & Arnheim, N. A member of a new repeated sequence family which is conserved throughout eucaryotic evolution is found between the human delta and beta globin genes. *Nucleic Acids Res* **9**, 5931-5947 (1981).
28. Tautz, D. & Renz, M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res* **12**, 4127-4138 (1984).
29. Toth, G., Gaspari, Z. & Jurka, J. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res* **10**, 967-981 (2000).
30. Nakamura, Y. DNA variations in human and medical genetics: 25 years of my experience. *J Hum Genet* **54**, 1-8 (2009).
31. Schlotterer, C. Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109**, 365-371 (2000).
32. Bhargava, A. & Fuentes, F.F. Mutational dynamics of microsatellites. *Mol Biotechnol* **44**, 250-266 (2010).
33. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell* **72**, 971-983 (1993).
34. Wooster, R., Cleton-Jansen, A.M., Collins, N., Mangion, J., Cornelis, R.S*., et al.* Instability of short tandem repeats (microsatellites) in human cancers. *Nat Genet* **6**, 152-156 (1994).
35. Sinden, R.R., Potaman, V.N., Oussatcheva, E.A., Pearson, C.E., Lyubchenko, Y.L. & Shlyakhtenko, L.S. Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. *J Biosci* **27**, 53-65 (2002).
36. Gill, P., Kimpton, C.P., Urquhart, A., Oldroyd, N., Millican, E.S., Watson, S.K. & Downes, T.J. Automated short tandem repeat (STR) analysis in forensic casework--a strategy for the future. *Electrophoresis* **16**, 1543-1552 (1995).
37. Walsh, P.S., Fildes, N.J. & Reynolds, R. Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Res* **24**, 2807-2812 (1996).

38. Edwards, A., Civitello, A., Hammond, H.A. & Caskey, C.T. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am J Hum Genet* **49**, 746-756 (1991).
39. Kimpton, C.P., Gill, P., Walton, A., Urquhart, A., Millican, E.S. & Adams, M. Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *PCR Methods Appl* **3**, 13-22 (1993).
40. Budowle, B., Moretti TR, Niezgoda SJ, Brown, BL. CODIS and PCR-Based Short Tandem Repeat Loci: Law Enforcement Tools. in *Proceedings of the second European Symposium on Human identification* 73-88 (Promega corporation, Innsbruck, Austria, 1998).
41. Butler, J.M. Genetics and genomics of core short tandem repeat loci used in human identity testing. *J Forensic Sci* **51**, 253-265 (2006).
42. Meloni, R., Laurent, C., Campion, D., Ben Hadjali, B., Thibaut, F.*, et al.* A rare allele of a microsatellite located in the tyrosine hydroxylase gene found in schizophrenic patients. *C R Acad Sci III* **318**, 803-809 (1995).
43. Burgert, E., Crocq, M.A., Bausch, E., Macher, J.P. & Morris-Rosendahl, D.J. No association between the tyrosine hydroxylase microsatellite marker HUMTH01 and schizophrenia or bipolar I disorder. *Psychiatr Genet* **8**, 45-48 (1998).
44. Schneider, P.M., Bender, K., Mayr, W.R., Parson, W., Hoste, B.*, et al.* STR analysis of artificially degraded DNA-results of a collaborative European exercise. *Forensic Sci Int* **139**, 123-134 (2004).
45. Wiegand, P. & Kleiber, M. Less is more--length reduction of STR amplicons using redesigned primers. *Int J Legal Med* **114**, 285-287 (2001).
46. Coble, M.D. & Butler, J.M. Characterization of new miniSTR loci to aid analysis of degraded DNA. *J Forensic Sci* **50**, 43-53 (2005).
47. Hill, C.R., Kline, M.C., Coble, M.D. & Butler, J.M. Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples. *J Forensic Sci* **53**, 73-80 (2008).
48. Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P.*, et al.* Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077-1082 (1998).
49. Chakraborty, R., Stivers, D.N., Su, B., Zhong, Y. & Budowle, B. The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. *Electrophoresis* **20**, 1682-1696 (1999).
50. Gill, P. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int J Legal Med* **114**, 204-210 (2001).
51. Sanchez, J.J., Phillips, C., Borsting, C., Balogh, K., Bogus, M.*, et al.* A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* **27**, 1713-1724 (2006).
52. Gill, P., Koumi, P. & Allen, H. Sizing short tandem repeat alleles in capillary array gel electrophoresis instruments. *Electrophoresis* **22**, 2670-2678 (2001).
53. Krenke, B.E., Viculis, L., Richard, M.L., Prinz, M., Milne, S.C.*, et al.* Validation of a male-specific, 12-locus fluorescent short tandem repeat (STR) multiplex. *Forensic Sci Int* **148**, 1-14 (2005).
54. Mulero, J.J., Chang, C.W., Calandro, L.M., Green, R.L., Li, Y., Johnson, C.L. & Hennessy, L.K. Development and validation of the AmpFlSTR Yfiler PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system. *J Forensic Sci* **51**, 64-75 (2006).

55. Budowle, B., Adamowicz, M., Aranda, X.G., Barna, C., Chakraborty, R.*, et al.* Twelve short tandem repeat loci Y chromosome haplotypes: genetic analysis on populations residing in North America. *Forensic Sci Int* **150**, 1-15 (2005).
56. Krenke, B.E., Tereba, A., Anderson, S.J., Buel, E., Culhane, S.*, et al.* Validation of a 16-locus fluorescent multiplex system. *J Forensic Sci* **47**, 773-785 (2002).
57. Ensenberger, M.G., Thompson, J., Hill, B., Homick, K., Kearney, V.*, et al.* Developmental validation of the PowerPlex((R)) 16 HS System: An improved 16-locus fluorescent STR multiplex. *Forensic Sci Int Genet* (2009).
58. Johnson, P. & Williams, R. DNA and Crime Investigation: Scotland and the 'UK National DNA Database'. *Scott J Crim Justice Stud* **10**, nihms6806 (2004).
59. Butler, J.M. *Forensic DNA Typing*, (2005).
60. Graves, J.A. & Schmidt, M.M. Mammalian sex chromosomes: design or accident? *Curr Opin Genet Dev* **2**, 890-901 (1992).
61. Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L.*, et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825-837 (2003).
62. Ross, M.T., Grafham, D.V., Coffey, A.J., Scherer, S., McLay, K.*, et al.* The DNA sequence of the human X chromosome. *Nature* **434**, 325-337 (2005).
63. Charlesworth, B. The evolution of sex chromosomes. *Science* **251**, 1030-1033 (1991).
64. Ali, S. & Hasnain, S.E. Genomics of the human Y-chromosome. 1. Association with male infertility. *Gene* **321**, 25-37 (2003).
65. Cooke, H.J., Brown, W.R. & Rappold, G.A. Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. *Nature* **317**, 687-692 (1985).
66. Jobling, M.A. & Tyler-Smith, C. The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* **4**, 598-612 (2003).
67. Jobling, M.A. & Tyler-Smith, C. Fathers and sons: the Y chromosome and human evolution. *Trends Genet* **11**, 449-456 (1995).
68. de Knijff, P., Kayser, M., Caglia, A., Corach, D., Fretwell, N.*, et al.* Chromosome Y microsatellites: population genetic and evolutionary aspects. *Int J Legal Med* **110**, 134-149 (1997).
69. Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C.*, et al.* Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med* **110**, 125-133, 141-129 (1997).
70. Prinz, M., Boll, K., Baum, H. & Shaler, B. Multiplexing of Y chromosome specific STRs and performance for mixed samples. *Forensic Sci Int* **85**, 209-218 (1997).
71. Prinz, M., Ishii, A., Coleman, A., Baum, H.J. & Shaler, R.C. Validation and casework application of a Y chromosome specific STR multiplex. *Forensic Sci Int* **120**, 177-188 (2001).
72. Shewale, J.G., Sikka, S.C., Schneida, E. & Sinha, S.K. DNA profiling of azoospermic semen samples from vasectomized males by using Y-PLEX 6 amplification kit. *J Forensic Sci* **48**, 127-129 (2003).
73. Roewer, L., Krawczak, M., Willuweit, S., Nagy, M., Alves, C.*, et al.* Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *Forensic Sci Int* **118**, 106-113 (2001).
74. Willuweit, S. & Roewer, L. Y chromosome haplotype reference database (YHRD): update. *Forensic Sci Int Genet* **1**, 83-87 (2007).
75. Ayub, Q., Mohyuddin, A., Qamar, R., Mazhar, K., Zerjal, T., Mehdi, S.Q. & Tyler-Smith, C. Identification and characterisation of novel human

Y-chromosomal microsatellites from sequence database information. *Nucleic Acids Res* **28**, e8 (2000).
76. Shewale, J.G., Nasir, H., Schneida, E., Gross, A.M., Budowle, B. & Sinha, S.K. Y-chromosome STR system, Y-PLEX 12, for forensic casework: development and validation. *J Forensic Sci* **49**, 1278-1290 (2004).
77. Mayntz-Press, K.A. & Ballantyne, J. Performance characteristics of commercial Y-STR multiplex systems. *J Forensic Sci* **52**, 1025-1034 (2007).
78. Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R.*, et al.* Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457-465 (1981).
79. Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. & Howell, N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* **23**, 147 (1999).
80. Wallace, D.C., Brown, M.D. & Lott, M.T. Mitochondrial DNA variation in human evolution and disease. *Gene* **238**, 211-230 (1999).
81. Giles, R.E., Blanc, H., Cann, H.M. & Wallace, D.C. Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A* **77**, 6715-6719 (1980).
82. Chen, X., Prosser, R., Simonetti, S., Sadlock, J., Jagiello, G. & Schon, E.A. Rearranged mitochondrial genomes are present in human oocytes. *Am J Hum Genet* **57**, 239-247 (1995).
83. Sutovsky, P., Moreno, R.D., Ramalho-Santos, J., Dominko, T., Simerly, C. & Schatten, G. Ubiquitin tag for sperm mitochondria. *Nature* **402**, 371-372 (1999).
84. Sutovsky, P., Moreno, R.D., Ramalho-Santos, J., Dominko, T., Simerly, C. & Schatten, G. Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. *Biol Reprod* **63**, 582-590 (2000).
85. Brown, W.M., George, M., Jr. & Wilson, A.C. Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A* **76**, 1967-1971 (1979).
86. Bogenhagen, D. & Clayton, D.A. The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. Quantitative isolation of mitochondrial deoxyribonucleic acid. *J Biol Chem* **249**, 7991-7995 (1974).
87. Holland, M.M., Fisher, D.L., Mitchell, L.G., Rodriquez, W.C., Canik, J.J., Merril, C.R. & Weedn, V.W. Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War. *J Forensic Sci* **38**, 542-553 (1993).
88. Allen, M., Engstrom, A.S., Meyers, S., Handt, O., Saldeen, T., von Haeseler, A., Paabo, S. & Gyllensten, U. Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. *J Forensic Sci* **43**, 453-464 (1998).
89. Gill, P., Ivanov, P.L., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E. & Sullivan, K. Identification of the remains of the Romanov family by DNA analysis. *Nat Genet* **6**, 130-135 (1994).
90. Wilson, M.R., DiZinno, J.A., Polanskey, D., Replogle, J. & Budowle, B. Validation of mitochondrial DNA sequencing for forensic casework analysis. *Int J Legal Med* **108**, 68-74 (1995).
91. Bender, K., Schneider, P.M. & Rittner, C. Application of mtDNA sequence analysis in forensic casework for the identification of human remains. *Forensic Sci Int* **113**, 103-107 (2000).
92. Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M.L. & Wallace, D.C. Classification of European

mtDNAs from an analysis of three European populations. *Genetics* **144**, 1835-1850 (1996).
93. Bar, W., Brinkmann, B., Budowle, B., Carracedo, A., Gill, P.*, et al.* Guidelines for mitochondrial DNA typing. DNA Commission of the International Society for Forensic Genetics. *Vox Sang* **79**, 121-125 (2000).
94. Parson, W. & Dur, A. EMPOP--a forensic mtDNA database. *Forensic Sci Int Genet* **1**, 88-92 (2007).
95. Parsons, T.J. & Coble, M.D. Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome. *Croat Med J* **42**, 304-309 (2001).
96. Coble, M.D., Just, R.S., O'Callaghan, J.E., Letmanyi, I.H., Peterson, C.T., Irwin, J.A. & Parsons, T.J. Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. *Int J Legal Med* **118**, 137-146 (2004).
97. Andreasson, H., Nilsson, M., Styrman, H., Pettersson, U. & Allen, M. Forensic mitochondrial coding region analysis for increased discrimination using pyrosequencing technology. *Forensic Sci Int Genet* **1**, 35-43 (2007).
98. Wallace, D.C. Mitochondrial DNA sequence variation in human evolution and disease. *Proc Natl Acad Sci U S A* **91**, 8739-8746 (1994).
99. Ivanov, P.L., Wadhams, M.J., Roby, R.K., Holland, M.M., Weedn, V.W. & Parsons, T.J. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nat Genet* **12**, 417-420 (1996).
100. Gilbert, M.T., Bandelt, H.J., Hofreiter, M. & Barnes, I. Assessing ancient DNA studies. *Trends Ecol Evol* **20**, 541-544 (2005).
101. Kemp, B.M. & Smith, D.G. Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. *Forensic Sci Int* **154**, 53-61 (2005).
102. Salamon, M., Tuross, N., Arensburg, B. & Weiner, S. Relatively well preserved DNA is present in the crystal aggregates of fossil bones. *Proc Natl Acad Sci U S A* **102**, 13783-13788 (2005).
103. Stewart, J.E., Aagaard, P.J., Pokorak, E.G., Polanskey, D. & Budowle, B. Evaluation of a multicapillary electrophoresis instrument for mitochondrial DNA typing. *J Forensic Sci* **48**, 571-580 (2003).
104. Butler, J.M., Buel, E., Crivellente, F. & McCord, B.R. Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis* **25**, 1397-1412 (2004).
105. Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-5467 (1977).
106. Prober, J.M., Trainor, G.L., Dam, R.J., Hobbs, F.W., Robertson, C.W., Zagursky, R.J., Cocuzza, A.J., Jensen, M.A. & Baumeister, K. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* **238**, 336-341 (1987).
107. Pettersson, E., Lundeberg, J. & Ahmadian, A. Generations of sequencing technologies. *Genomics* **93**, 105-111 (2009).
108. Ronaghi, M., Uhlen, M. & Nyren, P. A sequencing method based on real-time pyrophosphate. *Science* **281**, 363, 365 (1998).
109. Ronaghi, M. Pyrosequencing sheds light on DNA sequencing. *Genome Res* **11**, 3-11 (2001).
110. Ahmadian, A., Gharizadeh, B., Gustafsson, A.C., Sterky, F., Nyren, P., Uhlen, M. & Lundeberg, J. Single-nucleotide polymorphism analysis by pyrosequencing. *Anal Biochem* **280**, 103-110 (2000).

111. Ahmadian, A., Lundeberg, J., Nyren, P., Uhlen, M. & Ronaghi, M. Analysis of the p53 tumor suppressor gene by pyrosequencing. *Biotechniques* **28**, 140-144, 146-147 (2000).
112. Bravo, L.T., Tuohy, M.J., Ang, C., Destura, R.V., Mendoza, M., Procop, G.W., Gordon, S.M., Hall, G.S. & Shrestha, N.K. Pyrosequencing for rapid detection of Mycobacterium tuberculosis resistance to rifampin, isoniazid, and fluoroquinolones. *J Clin Microbiol* **47**, 3985-3990 (2009).
113. Deyde, V.M., Sheu, T.G., Trujillo, A.A., Okomo-Adhiambo, M., Garten, R., Klimov, A.I. & Gubareva, L.V. Detection of molecular markers of drug resistance in 2009 pandemic influenza A (H1N1) viruses by pyrosequencing. *Antimicrob Agents Chemother* **54**, 1102-1110 (2010).
114. Tost, J., Dunker, J. & Gut, I.G. Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing. *Biotechniques* **35**, 152-156 (2003).
115. Colella, S., Shen, L., Baggerly, K.A., Issa, J.P. & Krahe, R. Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites. *Biotechniques* **35**, 146-150 (2003).
116. Gharizadeh, B., Kalantari, M., Garcia, C.A., Johansson, B. & Nyren, P. Typing of human papillomavirus by pyrosequencing. *Lab Invest* **81**, 673-679 (2001).
117. Adelson, M.E., Feola, M., Trama, J., Tilton, R.C. & Mordechai, E. Simultaneous detection of herpes simplex virus types 1 and 2 by real-time PCR and Pyrosequencing. *J Clin Virol* **33**, 25-34 (2005).
118. Andreasson, H., Asp, A., Alderborn, A., Gyllensten, U. & Allen, M. Mitochondrial sequence analysis for forensic identification using pyrosequencing technology. *Biotechniques* **32**, 124-126, 128, 130-123 (2002).
119. Andreasson, H., Nilsson, M., Budowle, B., Frisk, S. & Allen, M. Quantification of mtDNA mixtures in forensic evidence material using pyrosequencing. *Int J Legal Med*, 1-8 (2006).
120. Tschentscher, F., Frey, U.H. & Bajanowski, T. Amelogenin sex determination by pyrosequencing of short PCR products. *Int J Legal Med* **122**, 333-335 (2008).
121. Edlund, H. & Allen, M. Y chromosomal STR analysis using Pyrosequencing technology. *Forensic Sci Int Genet* **3**, 119-124 (2009).
122. Divne, A.M., Edlund, H. & Allen, M. Forensic analysis of autosomal STR markers using Pyrosequencing. *Forensic Sci Int Genet* **4**, 122-129 (2010).
123. Ehn, M., Ahmadian, A., Nilsson, P., Lundeberg, J. & Hober, S. Escherichia coli single-stranded DNA-binding protein, a molecular tool for improved sequence quality in pyrosequencing. *Electrophoresis* **23**, 3289-3299 (2002).
124. Ronaghi, M. Improved performance of pyrosequencing using single-stranded DNA-binding protein. *Anal Biochem* **286**, 282-288 (2000).
125. Mashayekhi, F. & Ronaghi, M. Analysis of read length limiting factors in Pyrosequencing chemistry. *Anal Biochem* **363**, 275-287 (2007).
126. Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S.*, et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380 (2005).
127. Metzker, M.L. Sequencing technologies - the next generation. *Nat Rev Genet* **11**, 31-46 (2010).
128. Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U.*, et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722 (2010).
129. Green, R.E., Malaspinas, A.S., Krause, J., Briggs, A.W., Johnson, P.L.*, et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416-426 (2008).

130. Green, R.E., Krause, J., Ptak, S.E., Briggs, A.W., Ronan, M.T., *et al.* Analysis of one million base pairs of Neanderthal DNA. *Nature* **444**, 330-336 (2006).
131. PromegaCorporation. Technical Manual #TMD022-PowerPlex1 16 HS System. (Promega Corporation, Madison, WI, 2009).
132. Andreasson, H., Gyllensten, U. & Allen, M. Real-time DNA quantification of nuclear and mitochondrial DNA in forensic analysis. *Biotechniques* **33**, 402-404, 407-411 (2002).
133. Holland, P.M., Abramson, R.D., Watson, R. & Gelfand, D.H. Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proc Natl Acad Sci U S A* **88**, 7276-7280 (1991).
134. Livak, K.J., Flood, S.J., Marmaro, J., Giusti, W. & Deetz, K. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl* **4**, 357-362 (1995).
135. Heid, C.A., Stevens, J., Livak, K.J. & Williams, P.M. Real time quantitative PCR. *Genome Res* **6**, 986-994 (1996).
136. Nordstrom, T., Ronaghi, M., Forsberg, L., de Faire, U., Morgenstern, R. & Nyren, P. Direct analysis of single-nucleotide polymorphism on double-stranded DNA by pyrosequencing. *Biotechnol Appl Biochem* **31 (Pt 2)**, 107-112 (2000).
137. Boom, R., Sol, C.J., Salimans, M.M., Jansen, C.L., Wertheim-van Dillen, P.M. & van der Noordaa, J. Rapid and simple method for purification of nucleic acids. *J Clin Microbiol* **28**, 495-503 (1990).
138. Hoss, M. & Paabo, S. DNA extraction from Pleistocene bones by a silica-based purification method. *Nucleic Acids Res* **21**, 3913-3914 (1993).
139. Yang, D.Y., Eng, B., Waye, J.S., Dudar, J.C. & Saunders, S.R. Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *Am J Phys Anthropol* **105**, 539-543 (1998).
140. Rohland, N. & Hofreiter, M. Ancient DNA extraction from bones and teeth. *Nat Protoc* **2**, 1756-1762 (2007).
141. Lee, H.Y., Park, M.J., Kim, N.Y., Sim, J.E., Yang, W.I. & Shin, K.J. Simple and highly effective DNA extraction methods from old skeletal remains using silica columns. *Forensic Sci Int Genet* (2010).
142. Hanni, C., Brousseau, T., Laudet, V. & Stehelin, D. Isopropanol precipitation removes PCR inhibitors from ancient bone extracts. *Nucleic Acids Res* **23**, 881-882 (1995).
143. Hagelberg, E. & Clegg, J.B. Isolation and characterization of DNA from archaeological bone. *Proc Biol Sci* **244**, 45-50 (1991).
144. Loreille, O.M., Diegoli, T.M., Irwin, J.A., Coble, M.D. & Parsons, T.J. High efficiency DNA extraction from bone by total demineralization. *Forensic Sci Int Genet* **1**, 191-195 (2007).
145. Hill, C.R., Duewer, D.L., Kline, M.C., Sprecher, C.J., McLaren, R.S., *et al.* Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex((R)) ESX 17 and ESI 17 Systems. *Forensic Sci Int Genet* (2010).
146. Kayser, M., Liu, F., Janssens, A.C., Rivadeneira, F., Lao, O., *et al.* Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *Am J Hum Genet* **82**, 411-423 (2008).
147. Sturm, R.A., Duffy, D.L., Zhao, Z.Z., Leite, F.P., Stark, M.S., Hayward, N.K., Martin, N.G. & Montgomery, G.W. A single SNP in an evolutionary conserved

region within intron 86 of the HERC2 gene determines human blue-brown eye color. *Am J Hum Genet* **82**, 424-431 (2008).
148. Mengel-From, J., Borsting, C., Sanchez, J.J., Eiberg, H. & Morling, N. Human eye colour and HERC2, OCA2 and MATP. *Forensic Sci Int Genet* **4**, 323-328 (2010).
149. Pulker, H., Lareu, M.V., Phillips, C. & Carracedo, A. Finding genes that underlie physical traits of forensic interest using genetic tools. *Forensic Sci Int Genet* **1**, 100-104 (2007).
150. Kayser, M. & Schneider, P.M. DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations. *Forensic Sci Int Genet* **3**, 154-161 (2009).
151. Juusola, J. & Ballantyne, J. Messenger RNA profiling: a prototype method to supplant conventional methods for body fluid identification. *Forensic Sci Int* **135**, 85-96 (2003).
152. Juusola, J. & Ballantyne, J. Multiplex mRNA profiling for the identification of body fluids. *Forensic Sci Int* **152**, 1-12 (2005).
153. Juusola, J. & Ballantyne, J. mRNA profiling for body fluid identification by multiplex quantitative RT-PCR. *J Forensic Sci* **52**, 1252-1262 (2007).

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 606

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. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-131904

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2010