Single Nucleotide Polymorphisms: Characterisation and Application to Profiling of Degraded DNA

By

Shaikha Hassan Sanqoor M.Sc.

A thesis Submitted to the University of Central Lancashire
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

October 2009

DECLARATION

I declare that the work contained in this thesis has not been previously submitted for any other award from an academic institution. To the best of my knowledge and belief, the thesis contains no materials previously published or written by another person except where due reference is made.

Signed ---------- --------Date----------

Shaikha H Sanqoor

ABSTRACT

Single nucleotide polymorphisms (SNPs) are one of the forensic markers used to resolve the problem of DNA typing from degraded samples. It has been found in previous studies that when profiling heavily degraded forensic samples the small amplicon required for SNP analysis has an advantage over the larger STR loci, which are routinely used in forensic case work.

A total of 66 SNPs from the non-coding region of the 22 pairs of autosomal chromosomes were identified and SNP assays developed. Instead of selecting the SNPs from the available GenBank® sites, SNPs were typed from Arab individuals from Kuwait and United Arab Emirates (UAE) to identify polymorphic SNPs.

In order to obtain SNP data from Arab populations, a total of 10 unrelated Arab individuals from Kuwait and UAE were typed. The Affymetrix GeneChip® Mapping 250 K Array Sty I was employed to generate profiles for approximately 238,000 SNPs. Only autosomal SNPs were selected from the data.

Following selection, allele frequencies were estimated using the SNaPshot™ technique (Applied Biosystems) with 25 UAE individuals. For this technique, PCR forward and reverse primers were designed to generate PCR products less than 150 bp. The single base extension primers were designed to hybridise 1 bp upstream from the target SNP.

SNP characterization, including HardyWeinberg equilibrium and pair wise linkage disequilibrium, was carried out using the software package Arlequin v 3.1. Allele frequencies were calculated using Excel spreadsheets. PowerStats v.12 software used for discrimination power and match probability estimation.

All the 66 SNPs were polymorphic with average heterozygosity levels of 47%. A high heterozygosity level is very valuable for forensic application improving the individualization of forensic samples (Vallone et al. 2005). The probability that two individuals having identical genotype profile was found to be very low, 3.058×10^{-25}. The combined power of discrimination was found to be 0.999999999. This indicated that the selected SNPs met the parameters needed for forensic application.

The SNPs genotype sensitivity gave profiles from minute amounts of DNA template as little as 100 pico grams (pg) and optimal and reproducible results at 300 pg of DNA template.

The profiling of DNA from forensic samples is not always possible. This can be due to insufficient amount of samples being recovered and in many cases, DNA degradation. Biological materials that are recovered from the scene of the crime have often been exposed to sub-optimal environmental conditions such as high temperature and humidity.

SNPs performance on degraded samples was tested on artificially degraded saliva and semen samples. Controlled temperature and humidity experiments were performed to study the effect of these environmental factors on the samples. Also uncontrolled experiments on samples being subjected to different weather conditions (UK summer and UAE winter and summer) was performed in order to study and compare both weather effects on saliva samples. The triplex sets of SNPs that were developed for such study showed full allele profiles when compared to STRs, the current method used in forensic labs. In addition, SNPs produced a higher success rate than STRs when tested with samples obtained from human teeth remains and on samples subjected to DNase 1 digestion. The small size of SNPs, between 90 and 147 base pair (bp), showed more resistance to degradation than the STRs size ranging between 100 and 360 bp.

This study demonstrated that the 66 SNPs selected are useful markers when the typing of degraded samples by STRs fails to produce complete or partial profiles.

I dedicate this thesis with love to my late father and family

CONTENTS

List of Figures

List of Tables

ACKNOWLEDGMENTS

All thanks are due to Allah, the creator, who has power over all things.

There are a number of people who supported me during my research project. I would like to thank Dubai Police Head Quarters for their financial support to conduct this project, especially to General Khamis Al Muzainah, Brigadier Mohammad Saad Al Sharif and to Lieutenant Ahmed Al Mansoori.

I would like to thank my supervisor Dr William Goodwin who has provided me with guidance and advice throughout the course of my PhD project and Dr Sibte Hadi for his advice. Also I would like to thank Dr Arati Iyengar and Dr Judith Smith for their help and support. Many thanks go to Professor Jaipaul Singh and Dr Amal Shervington for their advice and help. I am particularly grateful to Dr Fred Harris for his suggestions to me during the writing of this thesis.

I would like to express my appreciation to National Centre of Meteorology & Seismology (Abu Dhabi, UAE), UAE Air Force & Air Defence Meteorology Centre and UK Meteorology Centre for providing me with the weather conditions data.

Thanks to Dubai Police Crime Laboratory for providing 100 blood samples. I would also like to thank Latheqia Sallam from Abu Dhabi Forensic Science Laboratory and Dr Mohammed Al-enizi from Kuwait General Department of Criminal Evidence who have provided Arab blood samples used for screening.

Many thanks to all my friends and colleagues in the Research Office who have supported and encouraged me throughout my project especially Nathalie, Shahid, Adnan, Glenda, Ash, Shanthi Helen, Cat and Alicia. I would like to extend my thanks to all people in the ITAV unit especially Barbara and Mohammad Asif for their help. Also

I would like to thank my friends Dr Aisha Khalifa for her support and Dr Ahmed Abdullah Ahmad for his help.

Finally, a very big thank you goes to my family. I am forever indebted to my parents and my sister Moza for their love, support and encouragement. Thanks are also due to my brothers, Mohammad, Obaid, Saeed, Abdul Aziz and Adil for their inspiration.

CHAPTER 1

INTRODUCTION

1.1. Overview

The majority of forensic analyses are concerned with the identification, characterisation and matching of forensic evidence. Frequently, the forensic scientist is asked to characterise biological samples from the scene of a crime for comparison with a potential suspect. Biological samples may include blood, semen, and saliva stains (Patzelt, 2004). Another category of forensic genetics is based around the testing of biological relationships and the identification of human remains, which may have been subjected to environmental insult.

1.2. Classic Genetic Markers

The suggestion that genetic markers may be applied to identify forensic samples is not a new concept (Altukhov and Salmenkova, 2002). The discovery of immunological and biochemical markers such as haemoglobin, blood grouping (ABO) and acid phosphatase, have been developed and applied to forensic analysis since 1915 (Patzelt, 2004; Jobling and Gill, 2004). These classic markers provide valuable evidence. However, these genetic markers show only small levels of individual variation and it is therefore difficult in many cases to produce a profile with a very high match probability. For example, the ABO blood group system can be used to classify people into only four different types: blood groups A, B, AB and O. The matching of an ABO type between a forensic blood stain and suspect therefore provides only weak statistical evidence for true association. Furthermore, these markers are unstable and frequently deteriorate in forensic specimens due to environmental effects such as heat, humidity and time (Budimlija et al., 2003).

1.3. Human Genome

1.3.1. Genomic Deoxyribonucleic Acid

Deoxyribonucleic acid (DNA) is the genetic material found in the cell nucleus. The human body is composed of trillions of cells, each cell, with the exception of red blood cells contains 46 chromosomes. The human genome is composed of 3.2 giga base pairs (Gb) of DNA (in a haploid cell). Individuals share approximately 99.9% homology through their genetic code; their genetic differences are determined by the remaining 0.1% of DNA (Baltimore, 2001; Li et al., 2006).

DNA contains length and sequence polymorphisms (Figure 1.1). The polymorphisms that have received most attention are related to disease, which lead directly to an individual developing an illness. Analysing regions of the genome that are not subject to selection pressure has also allowed DNA to be used to study human evolution. In addition, DNA analysis offers valuable information in forensic science with polymorphisms allowing the typing and identification of biological materials (Budowele et al., 2005).

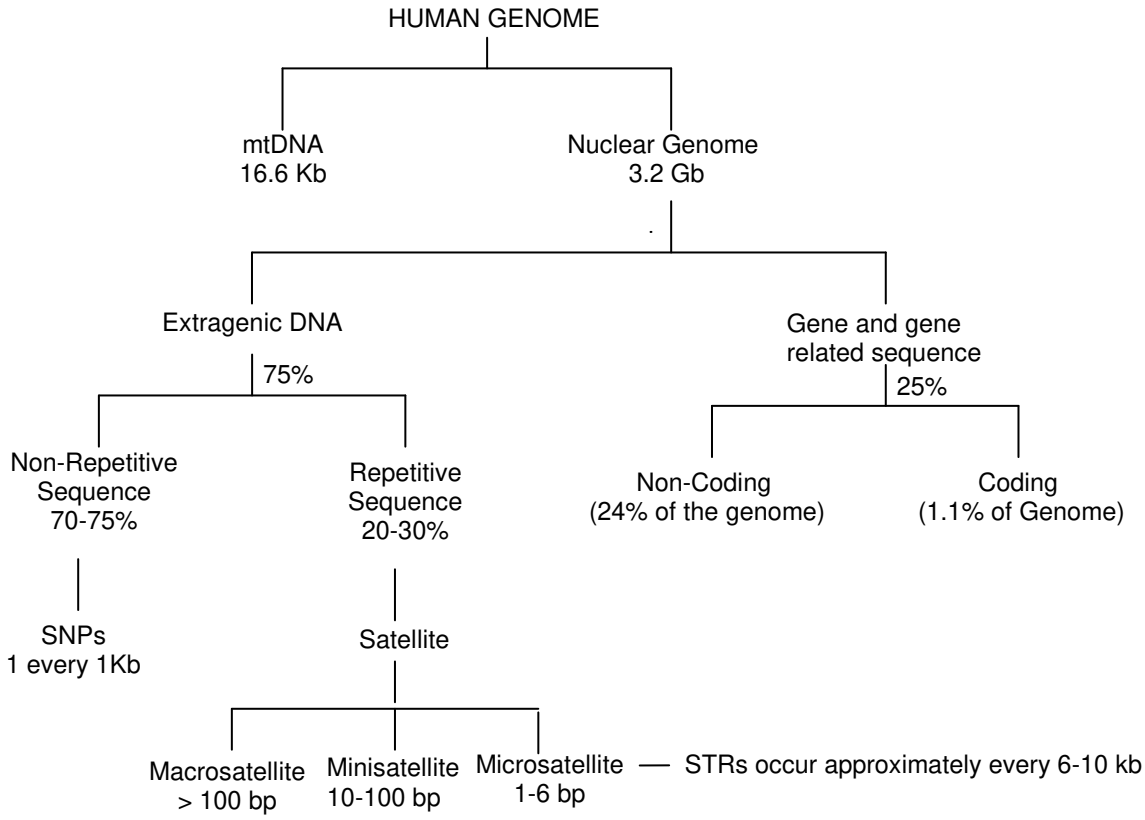


Figure 1.1. Shown above is a schematic diagram, which was adapted from Kashayab et al. (2004) and shows DNA polymorphisms in the human genome.

1.3.1.1. Coding Region

The portion of gene sequence in the human genome that is translated to protein is located in the coding regions, which are called exons, and represent only 1.1% of the genome (Baltimore, 2001). This region is responsible for an individual's phenotype such as skin colour and hair type, as well as all the underlying biochemical processes.

1.3.1.2. Noncoding Region

As reported by Venter et al. (2001) and Collins et al. (2004) in the analysis of the human genome sequence, noncoding DNA accounts for 99% of the genome. Most of

the genetic variation between humans is found within these noncoding regions (Sachidanandam et al., 2001).

1.3.2. DNA Polymorphisms

The a lleles are alternative forms of a gene that represent variation at specific position on chromosome and when the allele of a particular marker is present at 1% or greater in a given population, then that particular marker is considered to be polymorphic (Brookes, 1999).

Forensic DNA analysis began in 1985 after the discovery by Jeffreys et al. (1985) of variable number tandem repeats (VNTRs) or minisatellites. Minisatellites consist of a core region of DNA, which is typically 10 bp to 100 bp and is repeated tandemly. The variation of VNTRs between individuals exists due to different numbers of the core unit (Jeffreys et al., 1985).

VNTR technology was limited because it required a relatively large amount of high molecular weight DNA, which was not available from many forensic samples (Patzelt, 2004).

1.3.3. Polymerase Chain Reaction Mediated Analysis

Advances in molecular biology have made it possible to explore DNA variation directly. This, in turn, has led to the development of powerful DNA typing systems and the majority of these systems are based on the polymerase chain reaction (PCR), which is an enzymatic process by which a specific region of DNA is replicated many times to yield several million copies of a particular sequence (Saiki et al., 1985; Mullis et al., 1986). DNA amplification technology based on PCR is ideally suited for the analysis of

forensic samples, due to its sensitivity, its speed and its ability to provide sufficient copies of target sequences of DNA required for forensic comparison (Schneider et al., 2004; Kline et al., 2005).

1.3.3.1. Short Tandem Repeats

Short tandem repeats (STRs), also known as microsatellites, consist of tandem repeat sequences (Figure 1.2), with repeats consisting of 1-6 bp (Krawczak and Schmidtke, 1994). STRs are abundant throughout the human genome and occur on average every 6,000-10,000 bp (Beckmann and Weber, 1992).

Commercially available kits generate products that range between 100 bp and 450 bp. PCR-based systems, unlike VNTRs, require only one nanogram (ng) of DNA (Butler, 2007), and by typing several loci (typically at least 9 loci) simultaneously, high levels of discrimination can be achieved. The probability of two unrelated individuals having the same AmpFℓSTR® SGM plus® (which profiles 10 STR loci) profile is approximately 1 in 10^{-13} (Butler et al., 2003; Gill, 2002; Tsukada et al., 2002).

Using STRs to analyse highly degraded DNA in samples collected from crime scenes, including burnt and highly decomposed remains, is not always possible (Gill, 2002). In such samples, the DNA length is subjected to a reduction and ultimately larger STRs such as FGA (in the SGM plus®) loci are affected and allelic drop-out may be observed (Butler, 2006).

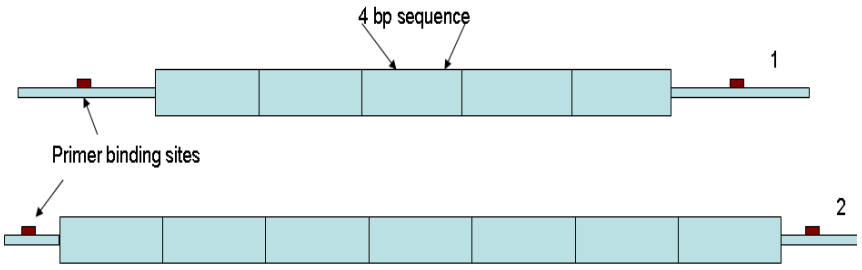


Figure 1.2. Shown above are two STR alleles containing 5 and 7 repeats of the core repeat. Also shown are the PCR primer binding sites that flank the repeat region.

1.3.3.2. Mini Short Tandem Repeats

Since PCR product sizes are governed by the primer binding site (Butler et al., 2004). In many cases it is possible to reduce the size of most PCR products by moving the primer binding site closer to the core repeat of the STRs (Figure 1.3) (Tsukada et al., 2002; Butler et al., 2003).

However, some STRs loci are not suitable for forensic analysis due to unsuitable primer sites or larger allele sizes, such as D13S317 and FGA (Butler et al., 2003). Moreover, the discriminatory power of commercial mini STR kits is lower than standard STR kits markers; only 8 loci are currently available in a commercial multiplex kit.

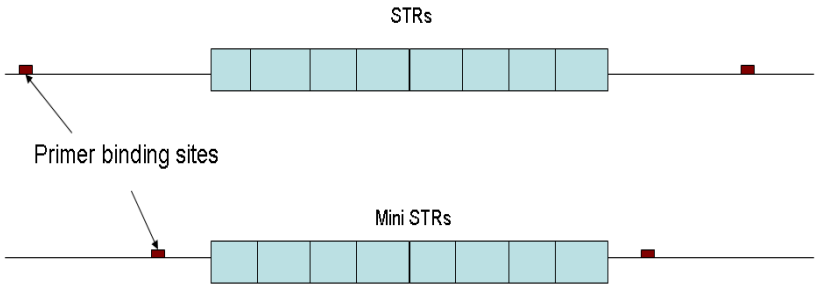


Figure 1.3. Shown above is a schematic diagram illustrating the difference of PCR primer binding sites between a STR and mini STRs. In the case of the mini STR, the primers bind nearer to the repeats.

The mini STR kits are designed for use along with one of the standards STR kit, for example the AmpFℓSTR® MiniFiler™ is designed to be used with the AmpFℓSTR® Identifiler® kit.

1.3.3.3. Y- Chromosome STRs

STRs markers on the Y chromosome can be considered as a fundamental tool in a number of forensic identification applications (Jobling, 2001; Gill et al., 2001; Sanchez et al., 2003). These STRs contain male genetic information (Butler, 2006) and thus may be applied to sexual assault cases where a mixture of male and female DNA is likely to be found (Jobling, 2001; Kayser, 2007). These STRs can also be useful in cases where the male genetic information is crucial, such as paternity cases, especially in the absence of the father, necessitating the testing of more distant relatives (Gill et al., 2001; Sanchez et al., 2003).

Despite their utility in forensic application, STRs on the Y chromosome encounter limitations as markers due to their haplotype nature, and lack of meiotic recombination. Consequently, their impact in forensic cases is reduced in terms of discrimination: the

genetic features of these STRs are inherited and passed from one generation to another among related males without change. However Y STRs can be applied for exclusion purposes (Palo et al. 2007).

1.3.3.4. Mitochondrial DNA

Human mitochondrial DNA (mtDNA) consists of approximately 16.5 kb (16,569 bp) of closed, double stranded, circular DNA (Holland and Parsons, 1999). Most of the sequence variation in this DNA is found in 2 hypervariable segments: hypervariable segment I (HVS-I) and hypervariable segment II (HVSII (Holland and Parsons, 1999).

In the context of forensic DNA typing, mtDNA is a powerful tool for typing damaged forensic samples. This is due to the fact that cells contain a high mitochondrial copy number, which is greater than 1000 per cell (Salas et al., 2007). The relative abundance of mtDNA makes it suitable to recover genetic information for forensic identification where the amount of nuclear DNA present is insufficient for analysis or the DNA is in a highly degraded state (Vallone et al., 2004; Niederstätter et al., 2006).

Due to the maternal mode of inheritance of mtDNA, the match probability of two individuals sharing the same profile is relatively high.

1.3.3.5. Low Copy Number

Full STR profiles can be routinely obtained from 250 picograms (pg) of DNA (Gill, 2001). The amount of template DNA recovered from many forensic samples is adequate (Clayton et al., 1995). However, in many cases, such as with touch DNA, insufficient DNA for standard profiling is recovered (Wolff and Gemmell, 2008).

To generate DNA profiles from samples with low copy number (LCN) different strategies have been employed to overcome the loss of genetic information (Mulero et al., 2008). These include: increasing the number of cycles from standard PCR protocol from 28-30 to 34 cycles, which was found to favour of number of detected alleles (Gill, 2001; Kloosterman and Kersbergen, 2003); reducing the PCR volume; filtration of the amplicon to remove ions that compete with DNA when being injected into the capillary; and adding more amplified product to the denature formamide; increasing injection time (Budowle et al, 2001; Forster et al, 2008). However, although these modifications to PCR and detection methodology led to improvements in some cases, ambiguous results that often interfere with the analysis of profiles led many forensic laboratories to stop using the method. Because of the sensitivity of the new method to contamination, exogenous DNA can be amplified along side the evidential DNA, introducing unrelated alleles. In addition, unbalanced alleles in heterozygote samples are often observed (Budowle et al., 2001; Gill et al., 2001).

1.4. Single Nucleotide Polymorphisms

Single nucleotide polymorphism (SNPs) in the human genome are the change of single nucleotides at a particular loci (Figure 1.4). On the basis of the number of alleles in each locus, SNPs are counted as biallelic polymorphisms, however, triallelic SNPs are also known to occur at a very low frequency within the human genome (Brookes, 1999).

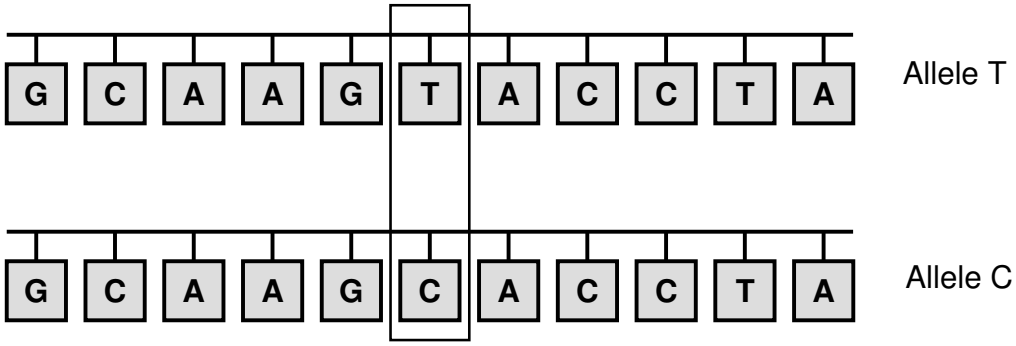


Figure 1.4. Shown above is a schematic diagram illustrating two DNA strands carrying a SNP: T and C.

SNPs occur, on average, every 1000 bp in the human genome, which leads to a high quantity of SNPs, most of which lie outside the coding region of the genome (Collins et al., 2001; Cooper et al., 1985; Metzker, 2005; Venter et al., 2001). These SNPs constitute more than 80% of genome variation with the remaining 20% of variation due to length polymorphisms, insertions, deletions and duplications (Haff and Smirnov, 1997).

The announcement of sequence mapping of the human genome in 2001 by the international human genome sequencing consortium, a worldwide collaboration of different groups, has increased the scientific communities's knowledge of SNPs greatly. The collaborating groups included: the haplotype map consortium (HapMap) (Sobrino et al., 2005), the SNP consortium (TSC) (Thorisson and Stein, 2003), and a number of other private groups and foundations such as academic centres and pharmaceutical companies (Halim and Altsbuler, 2001). Sequencing the human genome has provided researchers with tools and strategies to understand genetic variations, and the relation of phenotypes and the genes associated with particular diseases in humans (Gray et al., 2000).

1.4.1. Methods for the detection of SNPs

Large numbers of SNP sequences have been discovered over the past few years, which has led to a large amount of data becoming available for forensic applications (Thorisson and Stein, 2003). However, with the completion of the Human Genome Project, the discovery of SNPs has put great pressure on DNA technologists to design techniques and methods to meet the demand of researchers and scientists (Jenkins and Gibson, 2002).

In choosing a particular technique for SNP detection, it is important to consider the three main principles that govern the process:

- allelic discrimination reactions;

- detection techniques; and

- assay formats (Landegren et al., 1998; Sobrino et al., 2005).

1.4.1.1. Allelic Discrimination Reactions

Allelic discrimination reactions are methods to determine the type of variants of sequence on target DNA. On the basic alleles, variants can be classified as either homozygous; that is where two of the same kinds of variants are present, or heterozygous, where two different variants are present (Vallone et al., 2004).

Based on the mechanisms of the allelic discrimination reactions, different basic principles can be applied, including: allele specific hybridization (Wallace et al., 1979), primer extension (Syvanen, 1999), oligonucleotide ligation (Chen et al., 1998) and invasive cleavage (Olivier et al., 2002).

In the following outline, each discrimination reaction method is illustrated with examples for both its detection and assay format methods.

1.4.1.2. Allele Specific Hybridisation (ASH)

This method, also known as allele specific oligonucleotide (ASO), is based on the difference of thermal stability between two probes that hybridise with the target DNA (Wallace et al., 1979). The probe that is complementary to the variant SNP has a relatively high melting temperature. Conversely, the probe that has a mismatched sequence has a relatively low melting temperature. The product of allelic discrimination can be detected by many techniques, for example, fluorescence resonance energy transfer (FRET), which is the basis of the TaqMan assay, as shown in Figure 1.5 (Oliver et al., 2000; McGuigan and Ralston, 2002).

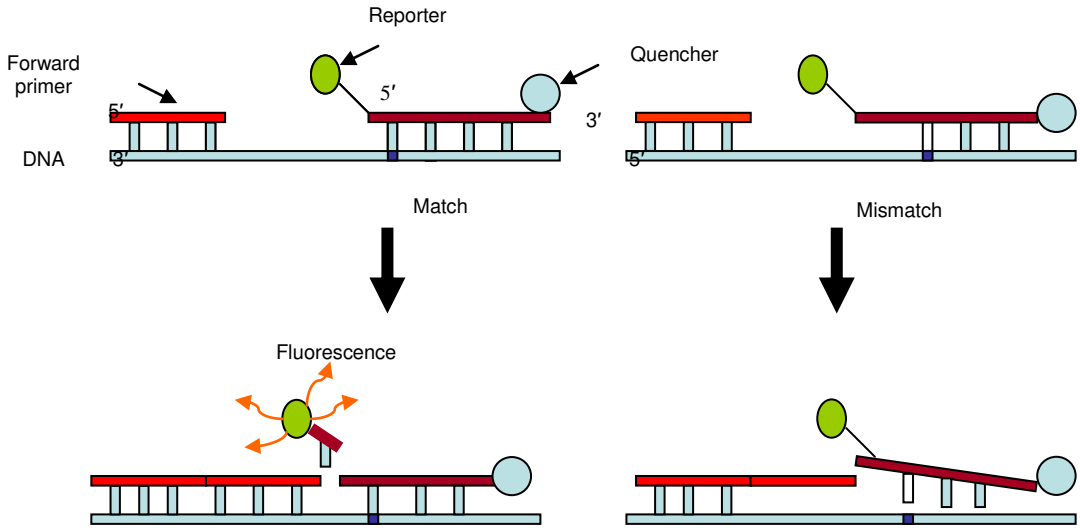


Figure 1.5. Shown above is a schematic representation of ASH using a TaqMan® probe. Illustrated is primer binding and allelic discrimination, which is achieved by the selective annealing of match probe and template sequence. The assay is based in the 5′ exonuclease activity of *Taq* polymerase. When the probe is intact the quencher interacts with the fluorophore (reporter) by fluorescence resonance energy transfer (FRET), quenching its fluorescence. In the extension step, the 5′ nucleotide, that has the fluorescent dye attached, is cleaved by the 5′ exonuclease activity of the *Taq* polymerase, leading to an increase in fluorescence of the reporter dye. A mismatched probe is displaced without fragmentation and no fluorescence is detected. Adapted from Livak (1999).

1.4.1.3. Primer Extension (PE)

This is one of the most frequently used detection methods currently used for SNP genotyping and is also known as minisequencing (Syvanen, 1999; Sanchez et al., 2003) and single base primer extension (SBE) (Inagaki et al., 2002). The mechanism of this method is based on the activity of DNA polymerase. However, PE methods can be divided into two types based on the principle of the extension mechanisms of the primer. In the first type, the primer binds upstream to the variant sequence on the target DNA. The dideoxynucleotide (ddNTP) that is complementary to polymorphic position is incorporated at the 3′ end of the primer by DNA polymerase (Syvanen, 1999). The product can then be detected by microarrays as used by the Affymetrix method (Divne and Allen, 2005) or electrophoresis as in the SNaPshot™ technique (Figure 1.6)

(Budowel, 2004). The second type involves the primer annealing to the polymorphic sequence and being extended by DNA polymerase only if it is a perfect match, with the product being determined using a technique such as pyrosequencing (Figure 1.7) (Ronaghi, 2001).

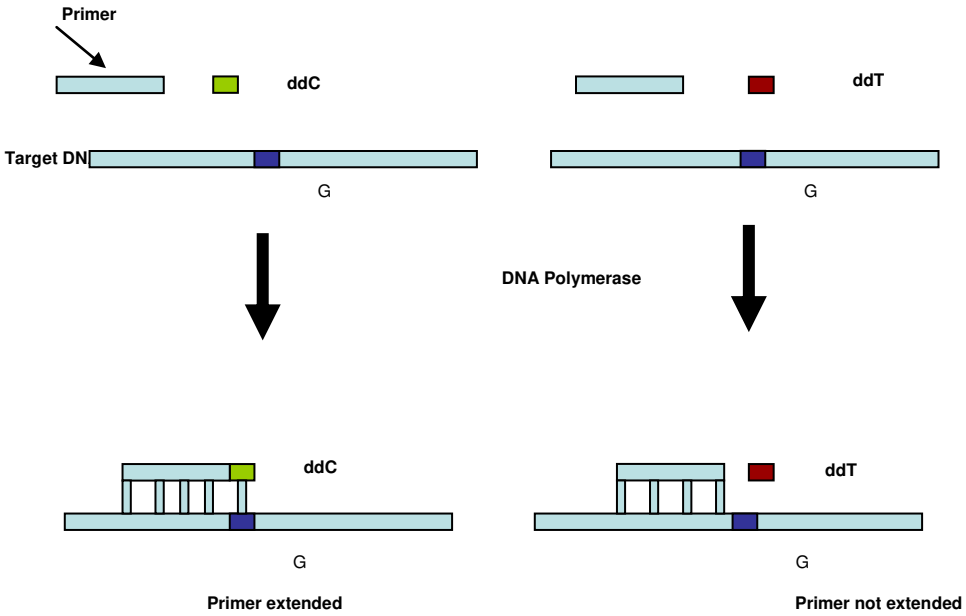


Figure 1.6. Shown above is a schematic diagram of PE using a single nucleotide primer extension assay. Under optimised conditions, a primer anneals to its target DNA immediately upstream to the SNP and is extended with single ddNTP complementary to the polymorphic base. The SNP patterns can be determined by the electrophoretic peaks as in SNaPshot™. This figure was adapted from Sobrino et al. (2005).

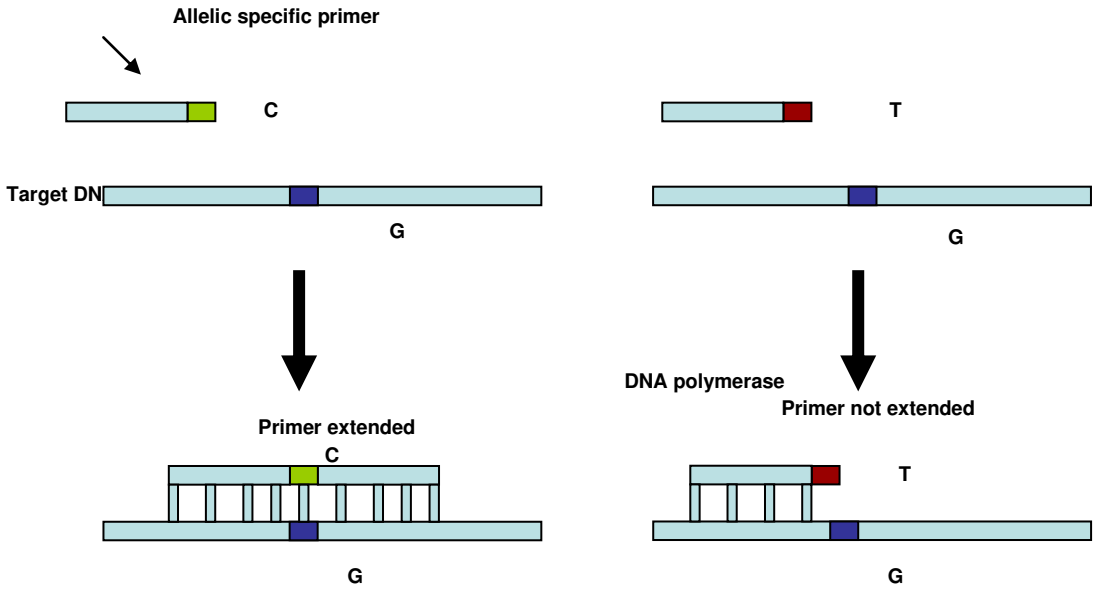


Figure 1.7. Shown above is a schematic representation of PE using an allelic specific extension. When there is a perfect match, the primer is extended by DNA polymerase Sobrino et al. (2005).

1.4.1.4. Allele Specific Oligonucleotide Ligation (ASOL)

The ASOL method requires three probes one of which is a generic probe that is designed to anneal to just one sequence on the polymorphic site (downstream) and two others which are allele specific probes. The generic probe and allele specific probes hybridise to the target DNA in tandem; the 5′ end of the generic probe joins to the 3′ end of the allele specific probe. However, the heterozygous sample will have both allele specific probes matched to the polymorphic sites on both strands (Figure 1.8) (Landegren et al., 1988).

The principle of this method depends on two factors: the first of these factors is hybridisation of the generic probe to the sequence adjacent to the SNP and the match between the sequences on allele specific probe to the SNP on the target DNA. The second of these factors is the ability of the ligase enzyme to join the two probes together by covalent bonding (Landegren et al., 1988).

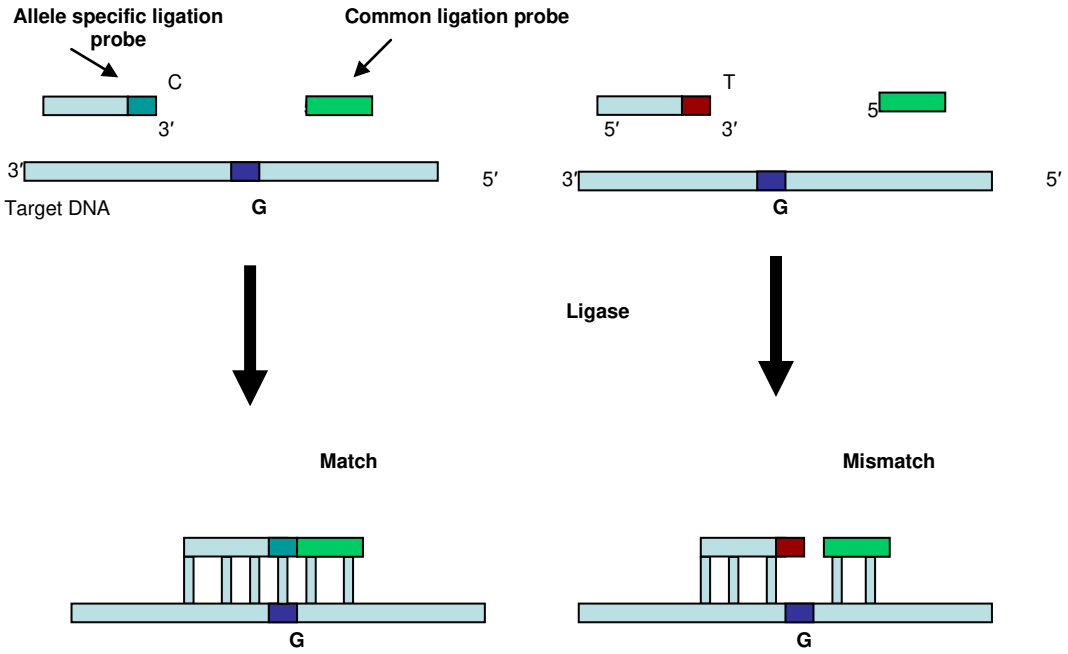


Figure 1.8. Shown above is a schematic diagram of ASOL. The common probe is hybridised adjacent to the allelic-specific probe. When there is a perfect match of the allelic-specific probe, DNA ligase joins both allelic-specific and common probes Adapted from Sobrino et al. (2005).

1.4.1.5. Invasive Cleavage

The reaction of this method is performed directly on genomic DNA, without prior amplification and is carried out in two stages (Figure 1.9) (Rao et al., 2003; Olivier et al., 2002; Lu et al., 2004).

The concept of the TaqMan assay (FRET) can be utilised in this method to monitor the alleles. The quencher is placed at the 3′ end of the allele specific probe and the labelled dye at the 5′ arm. The signal is only released when the invasive structure is formed on the target DNA (perfect match) (Olivier et al., 2002; Lu et al., 2004).

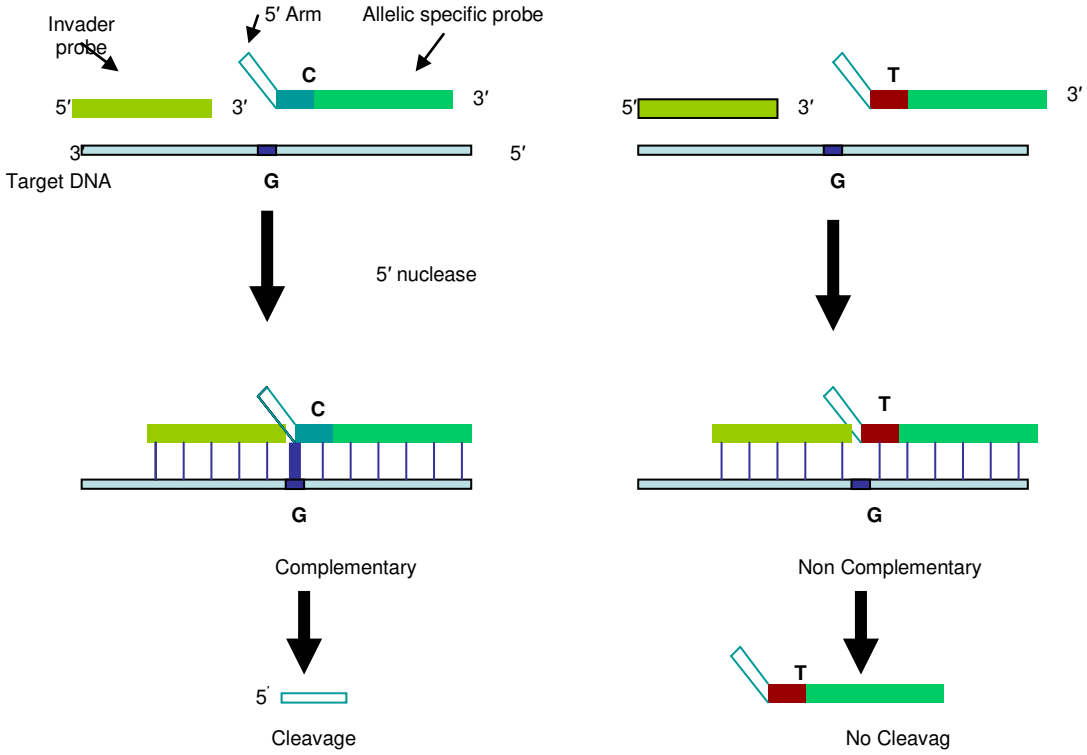


Figure 1.9. Shown above is a schematic illustration of the invasive cleavage allelic discrimination reaction. The invader probe and allele- specific probe anneal to the target DNA with an overlap of one nucleotide forming a structure that is recognised by 5′ exonuclease, releasing the 5′ arm of the allele specific probe. If the allele specific probe is not match the nucleotide at the SNP position, cleavage will not occur. Adapted from Sobrino et al. (2005).

1.4.2. Detection Methods

As was described above, the detection of SNPs at specific loci is dependent up on the mechanism of the allelic discriminatory reactions. Some discrimination reactions can be measured using different platforms.

1.4.3. Assay Format of SNP

There are two different categories which are related to SNP assay format. The first category of assay involves homogenous reaction in which the assay is performed in solution in a closed tube, as in the SNaPshot technique. The second category of assay,

which is normally referred to as heterogenous reaction, involves a solid support like microarray chip such as used in the Affymetrix technique (Gibson, 2006).

1.5. Forensic Biological Evidence

The purpose of forensic science is to identify and match biological samples. The recovery and analysis of DNA from such samples is the challenge for forensic scientists. Most of the biological samples such as blood, semen, saliva and tissue, which are found at the scene of crime, are exposed to environmental insult before collection. This can lead to degradation, especially in hot climates such as those found in the Arabian Gulf region. A large amount of forensic evidence can be lost using conventional STR technology (Bender et al., 2004).

1.6. DNA Degradation

It is well established that DNA can easily fragment in biological samples. Within cells, segments of double helix DNA are protected to some degree through association with the histones (Lewin, 2004). However, the linker DNA that connects the nucleosomes is more vulnerable and is often the point at which DNA degradation starts to occur (Coble and Butler, 2005).

Microorganisms can accelerate the breakdown of DNA. Deposited cellular material is a good source of nutrients for microorganisms, such as bacteria and fungi. Such microorganisms will secrete nucleases and, if the environmental conditions allow their growth, they can rapidly destroy the entire DNA (Bender et al., 2004; Vacca et al., 2005). Even without microorganisms, the breakdown of the cellular structure of deposited material will leave the DNA exposed to the cells' own nucleases (Pääbo et al., 2004; Neaves et al., 2009).

In addition to enzymatic effects, some chemical substances can also affect the DNA strands. For example, the hydrogen bonds that are present at the carbon atoms number 1, 2, 3, 4 and 5 of deoxyribose sugar of the DNA strand can react with compounds, such as hydrogen peroxide through oxidation (Pogozelski and Tullius, 1998). Also, chemical compound like nitric oxide (N_2O_2), can cause damage to DNA through deamination (removal of amino group) from both pyrimidines and purines bases (Nguyen et al., 1992). These oxidation and deamination processes lead to modification of primary structure of the DNA strand.

If the cellular material is exposed to direct sunlight the nitrogenous bases of DNA have the ability to absorb energy emitted by UV radiation (Hall and Ballatyne, 2004). This can lead to a photochemical reaction which alters the primary structure of the DNA strand leading to the formation of pyrimidine dimers (Mitchell et al., 1992). This does not destroy the DNA, but the cross-linking renders the DNA inert in a PCR.

1.7. Aims of the Project

Within the forensic field, there is a need for new markers that can overcome the problems encountered in typing degraded DNA (Budowele et al., 2005). SNPs represent the smallest available polymorphic markers.

In the present study, the focus will be on the identification of SNPs that may be informative in a forensic context within the Arab Population. To achieve this aim, individuals from the United Arab Emirates (UAE) and Kuwait have been employed for the first time as candidates to develop the use of SNP identification in forensic applications.

It was decided to generate the data from unrelated Arab individuals from Kuwait and UAE, instead of selecting available SNPs from the GenBank®. To obtain such data, the Affymetrix® technique was used.

The resulting SNP candidates from the autosomal chromosomes were then evaluated using the SNaPshot™ technique. Rigorous strategies and criteria were used to select SNPs. A series of statistical calculations were also used to determine the informative value of the SNP markers for the use in forensic analysis.

Finally based on the statistical calculations such as heterozygosity and discrimination power, at the completion of this research 66 of the best SNPs were selected as potential forensic markers. Their utility for the analysis of degraded DNA was assessed using both simulated and real forensic cases.

1.8. Population Overview

1.8.1. United Arab Emirates

The United Arab Emirates (UAE) comprises seven Emirates that were united in December 2, 1971 to form the State of UAE. Abu Dhabi is its capital and the political Emirate, whilst Dubai is the second Emirate and is famous for business and as a tourist attraction. Other Emirates of the UAE include: Sharjah, Ajman, Umm Al Qaiwain, Ras Al Khaimah and Fujairah.

UAE is a part of the Gulf Cooperation Council (GCC), which consists of six Gulf Countries; Bahrain, Kuwait, Oman, Saudi Arabia, Qatar, and UAE.

According to the 2006 census, the population of the UAE stood at 4.43 million. The indigenous inhabitants are called Emirati and constitute 20% of the total population.

The rest of the population are migrants and include South Asian (Indians, Pakistanis and Bangladeshis), Afghanis, Iranians, along with people from other Arab countries such as Palestine, the Yemen and Oman (www.vesitabudhabi.ae). Geographically, the UAE is situated along the coast of southern Arabian Gulf Sea, sharing borders with Oman, and Saudi Arabia (Figure 1.12).

Figure 1.10. Shown above is a map of the UAE indicating its borders with neighbouring GCC Countries. Saudi Arabia is located to the west, south and southeast whilst Oman lies to the southeast and northeast. Figure 1.12 was obtained from the UAE Ministry of Information and Culture, (1992).

1.8.2. Kuwait

The State of Kuwait is a part of the GCC, with a population of 963,571 Kuwaiti nationals according to the 2005 census (Al-Ghunaim, 2007). In addition to Kuwaitis, other people living and working in Kuwait include Iranians, Asians, and members of other Arab nations such as Palestine and Egypt. The state of Kuwait is situated on the northern tip of the Arabian Gulf Sea, sharing borders with Saudi Arabia and Iraq (Figure1.13).

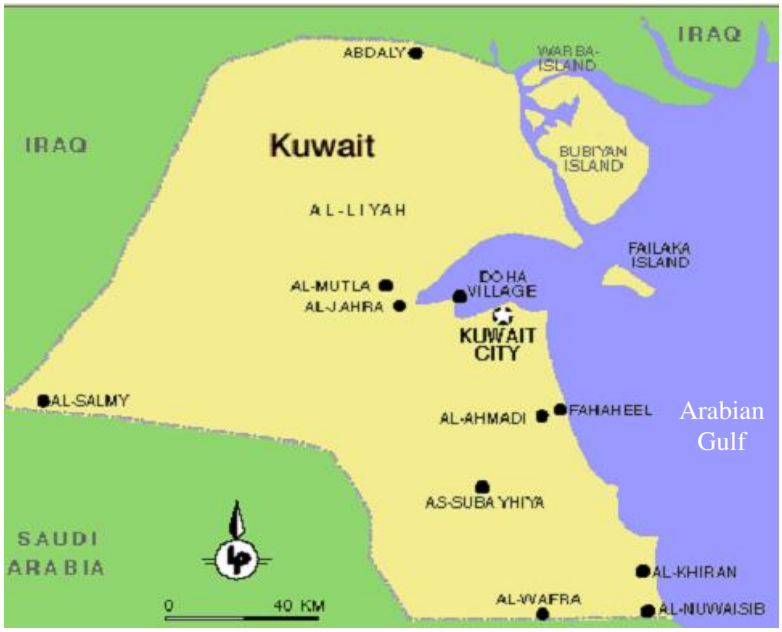


Figure 1.11. Shown above is a map of Kuwait indicating its borders with Saudi Arabia, which is located to the south west, and Iraq, which lies to the west and north.

CHAPTER 2

MATERIALS and METHODS

2.1. Sample Collection

In the following work, all samples were given with informed consent and were anonymised upon receipt. Samples of dried blood from 5 unrelated Kuwaiti individuals were collected and stored on FTA® paper by the Kuwait General Department of Criminal Evidence. Samples of dried blood from 5 unrelated UAE Arab individuals were collected by the Abu Dhabi Forensic Science Laboratory and placed on cotton swatches. To carry out the population study samples of dried blood from 100 unrelated United Arab Emirates (UAE) individuals were collected by the Dubai Police Crime Laboratory. The UAE samples were collected and stored on FTA® cards (Whatman® Bioscience, UK).

2.2. Affymetrix SNP Screening

2.2.1. Extraction and Purification of DNA

2.2.1.1. DNA Extraction

An area (1 cm^2) of cotton or FTA® card (from 5 Kuwaiti) was cut using sterile scissors and placed into a 1.5 ml tube (ELKay, UK). Using the modified method of Foran (2006), 500 µl of extraction buffer (0.01 M Tris, 0.01 M EDTA, 0.1 M NaCl and 2% SDS), 10 µl of 1 M DTT (Promega, US), and 20 µl of Proteinase K (20 mg/ml) (Qiagen Ltd, UK) was added to the tube. Samples were pulse vortexed and incubated on a Techne DB-2A heating block (Techne, USA) at 37 °C overnight (more than 10 h). Samples were removed from the heating block, briefly centrifuged at 13,000 rpm (Eppendorf 5415D, radius 6.4 cm) to remove condensation from the sides of the tube and purified as described in Section 2.2.1.2

2.2.1.2. Organic Solvent Purification

The following protocol was carried out in a flow hood. After the overnight incubation the samples, which were observed to be reddish coloured solutions, were individually transferred to a 1.5 ml tube, leaving behind the cotton/FTA® Card residue. As a first step, to each tube, 500 µl of phenol/chloroform/isoamyl alcohol in the ratio 25:24:1 (v/v) and at pH 8.0 (Fisher Bio Reagents, UK) was added. Each tube was then inverted several times until the solution appeared milky, vortexed and centrifuged at 13,000 rpm for 5 min. The pale yellow supernatant was removed so as not to disturb the lower organic phase, and retained. The retained supernatant was transferred into a new 1.5 ml tube. To each tube, a further 400 µl of phenol/chloroform/isoamyl alcohol was added and the previous step repeated. The resulting semi-clear supernatant was transferred into a Centricon® filter MY-100 membrane (Millipore, UK) and 1X TE buffer (1.0 M Tris HCl, 0.1 M EDTA, pH 8.0; Sigma, UK) was added to make the volume up to 2 ml. Each tube was then centrifuged (Falcon 6/300 Sanyo, radius 11.7 cm) at 3,500 rpm for 15 minutes (mins). The DNA sample in the filter was washed with TE buffer and centrifuged at 3,500 rpm for 15 mins. The filter was then inverted into a storage tube and centrifuged at 3,500 rpm for 5 mins. The resulting DNA samples were collected (approximately 35 µl) and stored at 4 °C for future use.

2.2.2. DNA Quantification

DNA samples that were extracted as described in Section. 2.2.1.1, were quantified using real-time PCR.

2.2.2.1. Application of the Quantifiler™ Human DNA Quantification Kit

DNA concentrations in samples were determined using the Quantifiler™ Human DNA Quantification Kit (Applied Biosystems, USA) with the ABI 7500 real-time PCR machine (Applied Biosystems). The procedure was carried out according to the manufacturer's protocol with the exception that the final volume of the reaction was reduced by half. Using 0.2 ml tubes, serial dilutions of the DNA standard, which was provided by the manufacturer, were prepared with TE buffer (Section 2.2.1.2) to give final DNA concentrations of 50, 16.5, 5.56, 1.85, 0.62, 0.21, 0.07 and 0.02 ng/µl. These DNA dilutions were stored in -20 °C for further use.

The total volume for the reaction was 12.5 µl, which comprised 5.25 µl of Quantifiler PCR Reaction Mix, 6.25 µl of Quantifiler Human Primer Mix and 1 µl of the DNA sample, including the non-template control (NTC) and the DNA standard. The reaction was prepared in a master mix. A MicroAmp™ optical 96-well reaction plate (Applied Biosystems) was placed on its base (MicroAmp™ splash free 96-well base) and 11.5 µl from the master mix was loaded into each well. Then, 1 µl of diluted DNA standard was loaded into the corresponding wells: each standard was set up in duplicate. Next to the standards, two wells were set for NTC into which 1 µl of TE buffer was loaded, then 1 µl of each sample was added into its corresponding well. When samples and standards were loaded, care was taken to avoid the formation of air bubbles.

The plate was sealed with an optical adhesive cover (Applied Biosystems) and placed into the ABI 7500, which was switched on prior to the reaction preparation. The thermal cycler protocol was performed in two stages: stage 1, hold at 95.0 °C for 10 minute (min); stage 2 consisted of 40 cycles at 95 °C for 15 seconds (s) followed by

60.0 °C for 1 min. After completion of the amplification the DNA concentration for each sample was estimated in ng /µl.

2.2.3. Whole Genome Amplification

2.2.4. Overview

Whole genome amplification is a well established technique to help overcome situations where there is insufficient DNA for analysis (Schneider et al., 2004). In the present study, whole genome amplification was used to increase the amount of DNA in samples from UAE and Kuwaiti individuals that were < 50 ng /µl to the levels that were required to conduct analysis using the Affymetrix Genechip®.

2.2.5. REPLI-g® Midi Kit

Whole genome amplification was performed with the QIAGEN REPLI-g® Midi kit. The method was based on the use of enzyme phi 29 (Φ 29) DNA polymerase.

The procedure for whole genome amplification was carried out according to the manufacturer's instructions. To a series of 1.5 ml microcentrifuge tubes was added 5 µl of reaction buffer, D1, and 5 µl of DNA sample containing < 50 ng of genomic DNA. A positive control sample was also prepared, containing 10 ng of DNA. All the tubes were briefly centrifuged at 13,000 rpm and then incubated at room temperature (23 °C) for 3 mins. After incubation, 10 µl of buffer N1 was added to each tube, the solution was mixed and briefly centrifuged at 13,000 rpm. To each tube was then added, 29 µl of REPLI-g Midi reaction buffer and 1 µl of REPLI-g Midi DNA polymerase, which was prepared as a master mix. Each tube was then incubated on a heating block overnight (30 °C) for 16 h and the reactions terminated by heating the block to 65 °C for 3 mins.

After cooling, the samples were removed and retained for further use. To assess the results of whole genome amplification, incubated samples were analysed using agarose gel electrophoresis (AGE) as described in Section 2.2.5.1.

2.2.5.1. Agarose Gel Electrophoresis (AGE)

AGE was conducted using a 0.5% (w/v) SeaKem® LE agarose gel in a tray tank (6 cm × 6 cm), which was submerged under TAE buffer (per 1000 ml: 4.84 g Tris Base, 1.14 ml glacial acetic acid, 2 ml 0.5 EDTA (pH 8.0)). Samples for AGE were prepared as follows: 2 µl of DNA were separately placed in test tubes and to each was added 2 µl of distilled water (dH_2O), 1 µl of gel loading buffer, and 6 × bromophenol blue (ABgene). As a size marker, a similar sample was also prepared except that amplified DNA was replaced with 2 µl of a *Lamda Hin*d III 23 kilo base pair (kb) ladder (ABgene™, UK). Immediately prior to use, the *Lamda Hin*d III ladder solution was heated at 56 °C for 15 mins. The gel was run at 100 V for 30 mins, stained in 0.5 µg/ml ethidium bromide (EtBr) and visualised using a UV transilluminator (Bio Doc- It™ Imaging System, US).

2.3. SNPs Screening

2.3.1. Affymetrix® GeneChip® Human Mapping 250K Array Sty 1

SNP analysis was conducted on samples (Section 2.1) obtained from 10 unrelated individuals from Kuwait and the UAE. The samples were analysed using GeneChip® Human Mapping 250K Array Sty 1. Due to specialist instrumentation requirements and the unavailability of essential equipment at the University of Central Lancashire, the samples were sent for analysis to Geneservice Ltd, UK.

2.3.2. Selection of Candidate SNPs

2.3.2.1. Software

Microsoft Office Access

Microsoft® Office Access 2003 was used to accommodate the high volume of SNP data obtained in this study. For further data analysis, Microsoft® Office Excel 2003 was employed

2.3.3. Identification of SNPs

For the initial identification of SNP markers, two different strategies were followed. First, a total of 238,304 SNPs from each Kuwaiti and UAE individual were linked together by Microsoft® Office Access 2003. The link was set to allow the combination of data from each individual into one group. This link was made by accessing the national centre for biotechnology information (NCBI) reference identifiers (dbSNP rs). Second, the data were rearranged according to autosomal chromosomes to reflect the number of SNPs in each chromosome that would be selected. For initial screening, SNPs with confidence values less than 0.09 were selected. This value was part of the Affymetrix® 250K chip analysis properties that were determined during SNP genotyping. This confidence value (< 0.09) permitted pooling of the data whose probability value indicated that more than 91% of SNPs were correctly genotyped. This, in turn, allowed a further reduction of the data size to a few thousand candidate SNPs. Ultimately, the reduction in size of the SNPs became appropriate for transference to an Excel sheet for further assessments.

The data was sorted according to the frequencies in ascending order using Excel. The SNPs with frequencies of 0.45 – 0.55 for each allele were selected.

2.3.4. Strategies and Criteria

In order to confirm the status of SNPs, and to determine conclusive screening results, several databases were interrogated. These included: Ensembl (http://www.ensembl.org), the Haplotype Map (HapMap) database (http://hapmap.org), the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov). As the above sites became publicly available during the course of the present research, they were incorporated into the data analysis strategy.

Also, a review of the existing literature identified a number of other properties to consider when selecting SNPs.

On the basis of position of the SNPs on the chromosomes, the following selection criteria were used:

1- The position of currently used STR markers in forensic analysis were identified and SNP candidates were selected at least 1 Mb from these regions.

2- SNPs that occurred at a distance of at least 100 kb from each other were targeted, as this distance was found to reduce the association between SNPs (Sanchez et al., 2006, Phillips et al., 2004).

3- To ensure the availability of specific regions for primer design and to prevent any complication during this process SNPs were selected so as to be 100 bp from any other characterised polymorphism (Sanchez et al., 2006).

5- Only SNPs that were located in the intergenic region were selected.

2.3.5. Design of PCR Primers

The PCR primer pairs (forward and reverse) used in this study was designed using the publicly available software: Primer3 (http://www.fro.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and Oligonucleotide Properties Calculator software (http://www.basic.nothwestern.edu/biotools/oligocalc.html). The design properties were based on singleplex primer conditions.

Template sequences 150 bp from both sides of the SNP marker were selected as primer binding sites and 20-30 bases upstream and downstream from the SNP sites were excluded as candidate PCR primer binding sites. The amplicon size was kept at less than 150 bp, to maximise amplification efficiency when typing degraded samples. The G-C contents of each primer was in the range of 35-60%, and in order to avoid hairpin formation, the 3′ end of each primer was checked for any complementary sequence to other parts of the primer as well as primer – primer interaction for each primer pair (Sanchez and Endicott, 2006).

To ensure the specificity of each primer for the target sequence, the test for non-specific target sites within the genome was determined using NCBI basic local alignment search tool (BLAST) program (www.ncbi.nlm.nih.gov/BLAST).

2.3.6. Primer Synthesis and Purity

The primers were synthesized by Invitrogen™ and were delivered desalted and lyophilised. Stock solutions of 100 µM primers were prepared by appropriate dilution with TE buffer. For example, primers supplied as 24.0 nanomoles were diluted with 240 µl of 1 × TE buffer. Stocks were kept at -20 °C, while an aliquot of 10 µM working solution for each primer was kept at 4 °C.

2.3.7. PCR Primer Optimisations

Each primer pair was optimised using single locus amplification. The PCR optimisation were carried out using thermal cyclers GeneAmp® 2700 , GeneAmp ® 9700 and Veriti ™ (Applied Biosystem) with the following PCR conditions: samples contained 0.5 ng of DNA template and primer 0.32 µM in a total reaction volume of 12.5 µl containing 1.1 X ReadyMix™ PCR master mix (ABgene™ UK). The $MgCl_2$ concentration in the reaction was adjusted to 2.5 mM by adding 1.0 mM from 25 mM stock (Applied Biosystems).

Each primer pair was tested using the following singleplex PCR conditions and cycle programme (Table 2.1).

Table 2.1. Indicated below are the cycling conditions and PCR Programmes for PCR primer optimization.

	Steps	Program A	Program B	Program C	Program D	Program E
Stage 1	Denature	95 °C 3 min	95 °C 3 min	95 °C 3 min	95 °C 3 min	95 °C 3 min
Stage 2	Denature	94 °C 1 min	94 °C 1 min	94 °C 1 min	94 °C 1 min	94 °C 1 min
	Annealing	56 °C 1 min	58 °C 1 min	60 °C 1 min	62 °C 1 min	64 °C 1 min
	Extended1	72 °C 1 min	72 °C 1 min	72 °C 1 min	72 °C 1 min	72 °C 1 min
	Extended 2	65 °C 7 min	65 °C 7 min	65 °C 7 min	65 °C 7 min	65 °C 7 min
Stage 3	Hold [a]	12 °C	12 °C	12 °C	12 °C	12 °C

[a] Hold is the final step for PCR till samples are removed from the PCR cycler.

All programmes were run for 30 cycles.

2.3.7.1. Gel Analysis of PCR Products

The PCR products of singleplex amplification were checked using AGE. Electrophoresis was conducted as described in Section 2.2.5.1 except that: a 2.5% (w/v) SeaKem® LE agarose gel and a tray tank (12 cm x 6 cm), which was loaded with 1 × TBE buffer (per 1000 ml: 10.8 gm Tris base, 5.5 gm Boric Acid, 4 ml 0.5 M EDTA at pH 8.0 at room temperature) were used. In addition, a 20 bp ladder (ABgene™) was used as a size marker. Samples for AGE were prepared as follows: 2 µl of amplified PCR products and the size marker were separately placed in test tubes and to each was added 2 µl of distilled water, and 1 µl of gel loading buffer (ABgene™).

2.3.8. Singleplex PCR Reaction

One PCR programme to amplify all primers individually was set up according to the conditions for PCR optimisation as described in Section 2.3.7 except that the following conditions were employed, based on the modified methodology of Sanchez and Endicott (2006): stage 1 was conducted at 95 °C for 3 mins; stage 2 at 94 °C for 1 min, 60 °C for 1 min, 72 °C for 1 min; this was repeated for 30 cycles, and then the reaction was incubated at 65 °C for 7 mins followed by 12 °C until samples were removed from the thermocycler. Three independent replicates were performed for each primer pair.

2.3.9. Gel analysis of Singleplex PCR Product

The amplified products of the PCR reaction were assayed as described in 2.2.5.1 except that a 2.5 % agarose gel was used.

2.3.10.PCR Reaction Clean Up

The remaining PCR products were purified to remove any excess of primers and dNTPs that were not incorporated during the amplification. The purification was carried out with the MinElute™ PCR purification spin column (Qiagen) following the manufacturer's protocol. The PCR product was eluted in 10 µl of elution buffer (EB). Alternatively, 0.5 µl ExoSAP-IT kit® (USB®, Germany) was added to 1 µl of PCR product and incubated at 37 °C for 15 mins, and inactivated at 80 °C for 15 mins, as indicated by the manufacturer's protocol.

2.3.11. Design of Single Base Extension Primers

Single base extension (SBE) primers were designed to hybridise to the target DNA one base from the 3′ end of polymorphic SNPs. Unless stated otherwise, the programmes, conditions and properties described in Section 2.3.5 were used to design SBE primers. Essentially, sequences, which were approximately 30 bp upstream and downstream of the SNP site, were selected as primer binding sites. The annealing temperature was kept between 60 °C ± 2 °C (Lindblad-Toh et al., 2000). During the initial stages of primer design, a number of the primers were made of different sizes (extended) by adding multiples of four poly-thymidine tail (poly T) to the 5′ end of the primers, as suggested by the Applied Biosystems SNaPshot® User's Manual (Biosystems, 2000).

2.3.12.Synthesis and Purities of SBE Primers

The SBE primers were synthesised by Invitrogen™ and delivered in a lyophilised form. Primers that were less than 30 bases were delivered as desalted and primers more than 30 bases in length purified using reverse phase chromatography. A stock solution of primers (100 µM) was prepared by adding the appropriate volume of 1X TE buffer

(Section 2.2.1.2), which was then kept at -20 °C. However, for more immediate use, 10 µM aliquots were prepared for each primer and kept at 4 °C.

2.3.13. Screening of SBE Primers

SBE primers were screened against non-template PCR amplicon to check as to whether any possible self extension or any unrelated peaks would be produced. The screening was carried out according to the manufacturer's protocol with the exception that the final volume of the reaction was reduced by half. The reaction components were 2 µl of SNaPshot™ mix, 0.5 µM (0.5 µl) of SBE primer and 2.5 µl dH_2O. Thermal cycling conditions were applied as described in the SNaPshot™ protocol: 96 °C for 10 s, 50 °C for 5 s, and 60 °C for 30 s, for 25 cycles. The product of the SNaPshot was purified and analysed as described below Sections 2.3.15 and 2.3.16.

2.3.14. Primer Extension Reaction

The primer extension reactions were carried out in a total volume of 5 µl, which comprised: 2 µl of SNaPshot™ mix, 0.5 µl of SBE primer (0.5 µM), 1.5 µl of dH_2O and 1 µl of PCR singleplex amplicons. Each reaction was performed with positive and negative controls as described in the manufacturer's protocol. Thermal cycling conditions for the reaction were as described in Section 2.3.13.

2.3.15. Removal of Unincorporated ddNTPs

The excess of fluorescently labelled ddNTPs in the primer extension reaction were removed by the addition of shrimp alkaline phosphatase (SAP). 1 µl of SAP (1 unit/µl; USB®, Germany) was added to the reaction tube, the reaction contents mixed briefly

and incubated at 37 °C for 40 mins, then at 90 °C for 5 mins to inactivate the enzyme (Vallone et al., 2005). The purified samples were kept at 4 °C.

2.3.16. ABI 310 PRISM® Genetic Analyzer

In a 200 µl PCR tube, 1 µl of SAP- treated primer extension products was diluted in 10 µl of Hi-Di™ formamide and 0.3 µl GeneScan™ 120-LIZ internal size standard (Applied Biosystems). The samples were mixed, briefly centrifuged at 13,000 rpm and then incubated at 95 °C for 5 mins. The samples were placed on ice prior to capillary electrophoresis (CE) ABI 310 PRISM® Genetic Analyzer as in Section 2.3.17.

2.3.17. ABI 310 PRISM® Genetic Analyzer Set Up

The separation of the SBE products was performed in a 47 cm long capillary (36 cm well-to-read) (Web Scientific Ltd, UK) using POP™4 polymer (Applied Biosystems). Electrophoresis running buffer (Applied Biosystems) was used in 1X concentration. The GS POP 4 (1 ml) E5 run module with dye set DS- 02 (filter set E5): dR110 (blue), dR6G (green), dTAMRA™ (yellow), dROX™ (red) and LIZ® (orange) was used with the following parameters: run temperature 60 °C, syringe pump time 150 s, pre-run voltage 15 kV, pre run time 120 s, injection time 5 s, and injection voltage 15 kV, run voltage 15 kV, run time 24 mins. Data analyses were performed using the software: GeneScan™ version 3.7 and GeneMapper ™ ID version 3.1. Three independent replicates were performed for each SNP reaction.

2.4. Sampling of UAE Individuals

2.4.1. Extraction Procedure

Blood from 100 UAE individuals were collected as described in Section 2.1 and DNA extracted as indicated in Section 2.2.2.1. These samples were then purified as described in Section 2.4.2.

2.4.2. Purifications

DNA extracted from the blood of 100 UAE individuals (Section 2.4.1) was purified using phenol/chloroform/isoayml alcohol as described in Section 2.2.1.2, except that a Microcon® YM-30 membrane (Millpore, UK) was used to concentrate the sample, which retained 15-20 µl. The supernatant from the second step of this protocol, which was a phenol/chloroform wash, was transferred to the microcon filter and the volume was brought up to the edge of the tube by adding 1X TE buffer. The microcon was centrifuged at 13,000 rpm for 12 mins (MSE-micro Centaur, SANYO) at room temperature (23 °C) and the filtrate was discarded. Approximately 400 µl of 1X TE was then added as a washing step to the microcon filter and the whole centrifuged at 13,000 rpm for 10 mins. The microcon filter was inverted into a new microcon collection tube and centrifuged at 1000 rpm for 3 mins. Approximately 20 µl of sample was collected and the stock tubes were stored at -20 °C. DNA in these samples was then quantified as described in Section 2.4.3.

2.4.3. Quantification

An estimation of DNA concentration from the 100 UAE samples was determined using the Quantifiler™ Human DNA Quantification Kit (Applied Biosystems) with ABI 7500 real time PCR (Applied Biosystems), as described Section 2.2.2.1

2.4.4. SNP Genotyping

In order to obtain quantitative information of allele frequencies of the candidate SNPs, each SNP was tested with 25 UAE samples. The reactions were carried out in singleplex. The analysis was performed as described in Sections 2.3.8 to 2.3.10 and 2.3.14, and 2.3.15 to 2.3.17.

2.4.5. Sensitivity Study

In order to determine the threshold amount of DNA to be correctly genotyped using SNPs, two DNA samples from two volunteer individuals were studied. Buccal samples on sterile cotton swabs were collected and allowed to air dry at room temperature (22 °C) for approximately 1 h.

DNA was extracted using Qiagen® QIAamp® DNA Mini Kit. The extraction was performed according to the manufacturer's protocol instruction for spin extraction as described in Section 2.4.6.

2.4.6. Qiagen® QIAamp® DNA Mini Kit Spin Extraction

The swab head containing the buccal sample was cut and placed in 1.5 ml tube. To this tube, 400 µl of 1X phosphate buffered saline (PBS: 137 mM $NaCl_2$, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 mM KH_2PO_4 at pH 7.4), 20 µl of proteinase K (Qiagen®) and 400

µl of buffer AL (provided by the manufacturer) were added. The tube was briefly vortexed and incubated at 56 °C for 2 h. The tube was then centrifuged at 13,000 rpm to remove any condensation left on the cap, 400 µl of 100% ethanol was added and the tube vortexed. Approximately 700 µl of the extracted sample was transferred into a spin column, which had previously been placed in a 2 ml tube (both provided by the manufacturer), and centrifuged at 8000 rpm for 1 min. The solution in the bottom tube was discarded and the last step was repeated until all remaining extracted sample was transferred into the column. 500 µl of AW1 solution (provided by the manufacturer) was added to the spin column, which was placed into a new 1.5 ml tube and the column was then centrifuged at 8000 rpm for 1 min. The solution from the lower tube was discarded and 500 µl of AW2 (provided by the manufacturer) was added to the column, centrifuged at 13,000 rpm for 1 min and solution from the bottom tube was discarded. The spin column was centrifuged once more at 13,000 rpm for 1 min to remove any residual ethanol. The 1.5 ml tube was removed and discarded, and the spin column placed in a fresh 1.5 ml tube with its cap cut and 150 µl of elution buffer (AE) was added. The spin column was let to stand for 1 min at room temp (23 °C) to allow the DNA sample to be eluted from the spin column filter into the solution. The column was then centrifuged at 8000 rpm for 1 min and DNA that had collected in the bottom tube was transferred into a fresh capped tube, and store at 4 °C for further analysis.

The extracted DNA was quantified using Quantifiler™ Human DNA Quantification Kit (Applied Biosystems) as described in Section 2.2.2.1.

2.4.7. Sequential Dilution of DNA

DNA from the two different buccal swabs extracted in Section 2.4.6 was diluted with 1 X TE buffer to give solutions with final DNA concentrations of: 100 pg/µl, 200 pg/µl,

300 pg/µl, 400 pg/µl, 500 pg/µl, 1000 pg/µl, 2000 pg/µl, 4000 pg/µl, and 8000 pg/µl. These dilution factors were based on the DNA concentration values obtained in Section 2.4.6.

2.4.8. SNP Amplification and Genotyping

The loci of four SNPs from four different chromosomes (Table 2.2) were included in this study. PCR was performed thrice at all the dilutions described in Section 2.3.8 and 2.3.10. The triplicate singleplex genotyping method was performed using a ABI 310 Prism® Genetic Analyser following the SBE reaction as described in Sections 2.3.14 to 2.3.15 using the conditions described in Sections 2.3.16 and 2.3.17.

SNP genotypes and relative fluorescence units (RFU) values for each homozygote and heterozygote peaks in each dilution were observed and assessed.

Table 2.2. Indicated below are the position on chromosome, the SNP type and PCR length for each of the 4 SNP loci used in the sensitivity study in Chapter 5.

SNP code	SNP ref	Position	SNP genotype	PCR Length (bp)
4-2	rs7684079	4	A/C	130
12-1	rs6487665	12	C/T	119
17-3	rs1872236	17	A/C	147
19-2	rs17304618	19	A/G	110

2.4.9. Multiplexing of SNP

To study the effect of degradation on the SNPs assay (Chapter 6), two sets of triplex PCR mixtures were used. The length of PCR products were categorised by size: small (<100 bp), medium (100-120 bp) and large (130-147 bp) as shown in Table 2.3. In order

to distinguish each SNP locus from others carrying the same fluorescent ddNTP dyes SBE primers were selected to be of different lengths (Schoske et al., 2003). Also, SNPs in the triplex sets were selected to contain the 4 possible labelled nucleotides (C, G, A and T).

Table 2.3. Indicated below are the PCR and SBE primers in the triplex sets with their SNP reference and position.

SNP code	SNP ref	SNP genotype	Position	PCR size (bp)	SBE size (bp)
		Triplex 1			
4-4	rs9995245	A/G	4	90	28
19-2	rs17304618	A/G	19	110	58
13-4	rs2892545	C/T	13	142	37
		Triplex 2			
21	rs8130475	A/G*	21	92	28
18-3	rs9950394	C/T*	18	119	54
17-3	rs1872236	A/C	17	147	42

* Genotypes are for reverse sequence.

2.4.10. Triplex Optimisation

PCR Conditions

Each set of the triplex (Section 2.4.9) was screened for primer dimer formation using the AutoDimer program (Vallone and Butler, 2004). The optimisation was carried out in a 12.5 µl reaction volume containing: 1.1 X ReadyMix™ PCR master mix (ABgene™) and 0.5 ng/µl DNA. The procedure was performed in 4 PCR reactions, each containing different concentration of primers, which were: 0.2 µM, 0.32 µM, and 0.4 µM, while the concentration of $MgCl_2$ was kept constant in each tube at 2.5 mM as in the

singleplex reaction (Table 2.4). PCR products for each reaction were checked using AGE as described in Section 2.2.5.1, except that a 2.5% (w/v) agarose gel was used. Based on the number and intensity of bands present, the relevant concentration of primers was determined. Then the $MgCl_2$ concentration was optimised while PCR primer concentration was kept constant: assays of each triplex set was performed in two PCR tubes each with 2.5 and 3.0 mM of $MgCl_2$ present (Table 2.5). The result was accepted when all three bands in the triplex were sharply defined. Based on this analysis, the final optimal primer concentrations were found to range between 0.2 and 0.4 µM whilst that of $MgCl_2$ was found to be 3 mM. All other conditions for triplex optimisation were as described for the singleplex reaction in Section 2.3.7. The thermal cycling programme was carried out as described in Section 2.3.8. After the cycling program, the reactions were then left at 12 °C until samples were removed from the thermocycler.

Table 2.4. Indicated below are the PCR primer optimizations for triplex 1 and 2. PCR tubes 1 to 3 contained equal concentration of primers, while for tube 4, the primers were mixed in different concentrations. For all the reactions, the $MgCl_2$ concentration was kept constant at 2.5 mM.

	Set 1(4-4, 13-4 and 19-2)				Set 2 (21-17-3 and 18-3)		
PCR Tube	Primers Code	Primer Conc. µM	$MgCl_2$ Conc. mM	PCR Tube	Primers Code	Primer Conc. µM	$MgCl_2$ Conc. mM
1	all primers	0.2	2.5	1	all primers	0.2	2.5
2	all primers	0.32	2.5	2	all primers	0.32	2.5
3	all primers	0.4	2.5	3	all primers	0.4	2.5
4	4-4 & 13-4 19-2	0.2 0.4	2.5	4	21 &17-3 18-3	0.2 0.4	2.5

Table 2.5. Indicated below are the optimal $MgCl_2$ concentrations for analysis of triplex set 1 and 2 when the concentration of primers are kept constant.

Set 1(4-4, 13-4 and 19-2)				Set 2 (21-17-3 and 18-3)			
PCR Tube	Primers Code	Primer Conc. µM	$MgCl_2$ Conc. mM	PCR Tube	Primers Code	Primer Conc. µM	$MgCl_2$ Conc. mM
1	4-4 & 13-4 19-2	0.4 0.2	2.5	1	21 &17-3 18-3	0.2 0.4	2.5
2	4-4 & 13-4 19-2	0.4 0.2	3.0	2	21 &17-3 18-3	0.2 0.4	3.0

SBE Reaction Conditions

As PCR primers, the SBE primers were also checked for primer dimmer formation using the AutoDimer program. The triplex reaction was then carried out as described in Section 2.3.14, except that the SBE primers concentration used in all cases was 0.2 µM.

Purifications

The end products of PCR and SBE reactions were purified to remove excess primers and unused ddNTPs by using 0.5 µl of ExoSAP-IT kit® (USB®) and 1 µl of SAP (USB®) as described in Sections 2.3.10 and 2.3.15.

2.4.11. Triplex Genotyping

Analysis of the optimised triplex set was carried out using the ABI 310 Prism® Genetic Analyser as described in Sections 2.3.16 and 2.3.17.

2.5. Degradation Assessments

2.5.1. Controlled Environmental Conditions

In this method, in order to generate degraded DNA samples were exposedto environmental insult with the humidity and temperature controlled in the laboratory.

A. Humidity and Temperature

A 50 µl of sample (saliva/semen) was pipetted onto a sterile cotton swab (COPAN) and kept in an incubator at 37 °C with a humidity of 98% ± 2% for a period of 18 days.

The humid environment was prepared as follows: layers of tissue paper were saturated with distilled water (dH_2O) and folded to fit a solid plastic container. The swabs were placed into a rack inside the container so that they were not touching. An EL-USB2-RH/ temperature Data logger (LASCAR electronics, UK) inside the container was used to monitor the humidity and temperature during the experiment (Figure 2.1). The USB data logger was set up in accordance with the manufacturer's instructions. In order to prevent the loss of water vapours, the container was tightly sealed and incubated at 37 °C in a hybridisation oven (HYBAID™, UK). Samples were removed at 3 days intervals and stored at -20 °C until processed further.

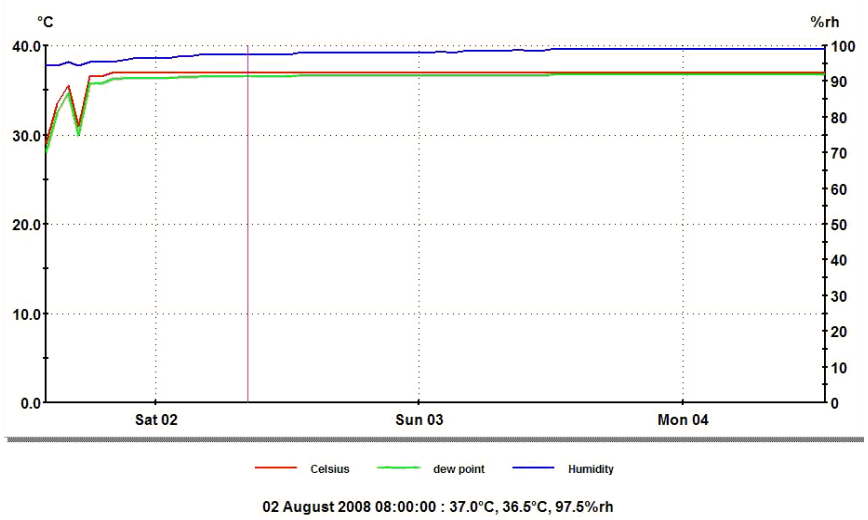


Figure 2.1. Shown above is the data over a 3 day incubation period were recorded on the USB data logger. The relative humidity percentage (% rh) was 98% and the temperature was 37 °C.

B. Room Temperature

The samples were prepared as described above in humidity and temperature and kept at room temperature in a laminar flow hood cabinet. Temperature was recorded using a thermometer; every 3 days sample was removed and stored at -20 °C. Temperature ranged from 21-24 °C.

2.5.2. Environmental Conditions

UAE Weather in December/ January and September

Two uncontrolled experiments were conducted in two different UAE climates: December 2007 to January 2008 (Table 2.6), with temperatures ranging between 21 °C and 24 °C; and September 2008, with temperatures ranging between 35°C and 39 °C (Table 2.7).

50 µl of saliva from a female donor was added onto a microscopic glass slide. The samples were placed outside exposed to environmental conditions. The samples were removed after a set of period of 3, 6, 12 days and the temperature was taken from the recorded using UAE weather forecasting service (Table 2.8) and (Table 2.9).

Table 2.6. Indicated below are the UAE weather conditions in December/January for degraded saliva samples.

Duration (days)	3	6	12
Start date at 1 pm	20/12/07	20/12/07	20/12/07
Start temp (°C)	24	24	24
End date	23/12/07 1pm	26/12/07 1. 05 pm	01/01/08 1 pm
Weather condition	partially cloudy	sunny - partially cloudy	sunny-partially cloudy
End temp (°C)	21	22	24

Table 2.7. Indicated below are UAE weather conditions in September/October for degraded saliva samples.

Duration (days)	3	6	12	18
Start date at 10 am	18/09/08	18/09/08	18/09/08	18/09/08
Start temp (°C)	35	35	35	35
End date	21/09/08 10.30	24/09/08 10 am	30/09/08 10 am	06/10/08 10 am
Weather condition	sunny	sunny	sunny	sunny
End temp (°C)	39	37	35	34

Table 2.8. Shown below are the December 2007 hourly data obtained from Met Office UAE.

Date	Time (24 hr clock)	Relative Humidity %	Temperature (°C)
20/12/2007	00:00	75	20.8
20/12/2007	01:00	78	20.2
20/12/2007	02:00	78	19.4
20/12/2007	03:00	79	18.7
20/12/2007	04:00	63	20.4
20/12/2007	05:00	57	21.5
20/12/2007	06:00	47	24.2
20/12/2007	07:00	46	24.7
20/12/2007	08:00	46	24.8
20/12/2007	09:00	45	24.9
20/12/2007	10:00	48	24.7
20/12/2007	11:00	52	24.3
20/12/2007	12:00	53	24.0
20/12/2007	13:00	55	23.5
20/12/2007	14:00	59	23.2
20/12/2007	15:00	62	22.9
20/12/2007	16:00	65	22.5
20/12/2007	17:00	65	22.1
20/12/2007	18:00	68	21.8
20/12/2007	19:00	70	21.5
20/12/2007	20:00	72	20.6
20/12/2007	21:00	73	20.1
20/12/2007	22:00	72	19.8
20/12/2007	23:00	69	19.5
Average		62.4	22.2

Table 2.9 Shown below are the September hourly data obtained from Met Office UAE.

Date	Time (24 hr clock)	Relative Humidity %	Temperature (°C)
18/09/2008	00:00	61	29.0
18/09/2008	01:00	61	28.4
18/09/2008	02:00	61	28.2
18/09/2008	03:00	55	30.3
18/09/2008	04:00	53	30.8
18/09/2008	05:00	49	33.1
18/09/2008	06:00	43	35.5
18/09/2008	07:00	44	35.6
18/09/2008	08:00	43	36.0
18/09/2008	09:00	41	35.8
18/09/2008	10:00	43	35.1
18/09/2008	11:00	48	34.5
18/09/2008	12:00	47	34.5
18/09/2008	13:00	52	34.0
18/09/2008	14:00	54	33.6
18/09/2008	15:00	52	33.2
18/09/2008	16:00	56	32.8
18/09/2008	17:00	59	32.0
18/09/2008	18:00	62	31.3
18/09/2008	19:00	62	30.7
18/09/2008	20:00	62	30.3
18/09/2008	21:00	64	29.6
18/09/2008	22:00	86	28.9
18/09/2008	23:00	70	28.4
Average		55.3	32.2

UK Weather Conditions

The experiment was conducted in August 2008 (Table 2.10). A 50 µl of saliva from a female volunteer was added onto each microscopic glass slide. Samples were placed outside and exposed to weather conditions such as light, UV and humidity. However, the experiment was conducted in a covered, outside environment, to prevent the sample from being washed away by the rain. The temperature was taken from the recorded using UK weather forecasting service (Table 2.11) Samples were removed at 3 day intervals and stored at - 20 °C until the experiment was completed.

Table 2.10. UK weather conditions in August for degraded saliva samples.

Duration (days)	3	6	9	12	15	18
Start date at 12 pm	01/08/08	01/08/08	01/08/08	01/08/08	01/08/08	01/08/08
Start temp (°C)	19	19	19	19	19	19
End date	04/08/08 12.45 pm	07/08/08 13.05 pm	10/08/08 12 pm	13/08/08 12 pm	16/08/08 12.05pm	19/08/08 12 pm
Weather conditions	Cloudy-raining	Raining	Cloudy	Raining	Raining	Raining
End temp (°C)	18	18	19	19	19	17

Table 2.11. An example of the hourly data obtained from Met Office UK.

Date	Time (24 hr clock)	Relative Humidity %	Temperature (°C)
04/08/2008	00:00	92.4	10.9
04/08/2008	01:00	96.1	10.5
04/08/2008	02:00	96.2	11.2
04/08/2008	03:00	95.1	11.7
04/08/2008	04:00	95.1	12.0
04/08/2008	05:00	92.7	12.3
04/08/2008	06:00	89.5	13.0
04/08/2008	07:00	88.6	13.9
04/08/2008	08:00	91.1	14.6
04/08/2008	09:00	95.5	14.4
04/08/2008	10:00	90.3	15.5
04/08/2008	11:00	85.6	16.9
04/08/2008	12:00	78.0	17.3
04/08/2008	13:00	72.1	18.4
04/08/2008	14:00	62.9	19.0
04/08/2008	15:00	64.6	19.0
04/08/2008	16:00	57.0	19.7
04/08/2008	17:00	64.6	18.9
04/08/2008	18:00	64.1	18.5
04/08/2008	19:00	67.7	17.5
04/08/2008	20:00	74.7	15.6
04/08/2008	21:00	83.6	13.0
04/08/2008	22:00	87.5	13.8
04/08/2008	23:00	87.7	14.3
Average		82.1	15.1

2.5.3. Reference Samples

Reference samples were taken at the start of each experiment to represent time zero. The samples were prepared as follows: 50 µl of the sample was placed onto a sterile cotton swab and kept for approximately 1 h at room temperature (22 °C) to air dry. Samples were then stored at -20 °C until all experiments were completed and ready for extraction.

2.5.4. Extraction and Quantification

2.5.5. DNA Extraction from Semen Stains

The extraction procedure was carried out following the protocol in the QIAamp® DNA Investigator Handbook (Qiagen 2007) as described below in Section 2.5.6. The concentration of extracted DNA was estimated using the Quantifiler® Human DNA Kit as described in Section 2.2.2.

2.5.6. QIAamp® DNA Investigator

DNA was extracted according to the QIAamp® DNA Investigator Handbook protocol for isolation of DNA from sexual assaults. The swab heads, containing the semen, were cut off and the samples were placed into a 1.5 ml microcentrifuge tube, with 400 µl of ATL (Qiagen), 20 µl (2 mg/ml) of Proteinase K (Qiagen) and 10 µl of 1 M DTT (0.13 g/ml) added. The sample was pulse vortexed, incubated in a dry block at 56 °C for 2 h with vortexing approximately every 10 mins to ensure maximal lysis. After incubation, the tube was centrifuged at 13,000 rpm, 400 µl of AL (Qiagen) added, after which the sample was vortexed again and incubated in a dry block at 70 °C for 10 mins. Following incubation, the sample was briefly centrifuged at 13,000 rpm and 300 µl of 96% ethanol

was added. The sample was again briefly centrifuged at 13,000 rpm. A spin column (Qiagen) was placed into a 2 ml collection tube (Qiagen) and approximately 700 µl of the extracted sample was transferred into the column. The column was centrifuged at 8,000 rpm for 1 min, and the solution in the collection tube was discarded. The above step was repeated until all the extracted solution was transferred into the column. 500 µl of AW1 (Qiagen) was added and centrifuged at 8,000 rpm for 1 min. The solution from the collection tube was discarded, 500 µl of AW2 (Qiagen) was then added and the column centrifuged at 13,000 rpm for 1 min. To remove any trace of AW2, the the sample was centrifuged for a further 3 mins. The spin column was placed into clean microcentrifuge with its cap removed, the column was uncapped and kept at room temperature for 1 min. 150 µl of AE buffer was added into the spin column incubated at room temperature for 1 min and was centrifuged at 13,000 rpm for 1 min. DNA was recovered and transferred into new capped 1.5 ml microcentrifuge tube and stored at 4 °C until quantification.

The concentration of extracted DNA was estimated using the Quantifiler® Human DNA Kit (Applied Biosystems) as described earlier in Section 2.2.2.1.

2.5.7. DNA Extraction from Saliva Stains

The saliva sample on the microscopic glass slide was transferred onto sterile cotton swab as follows: a dry swab was moistened with 1X TE buffer, and used to lift up the sample from the glass slide. The extraction procedures for all saliva samples were carried out using Qiagene™ QIAamp® DNA Mini Kit as described in Section 2.4.6. (Chapter 2). The DNA was quantified using Quantifiler® Human DNA Kit as described above.

2.5.8. Amplification and Genotyping

To evaluate the efficiency of the degradation study, both sets of samples (saliva and semen) generated under all above conditions were examined using two different methods; SNP and STR analysis.

2.5.9. SNP Typing

Amplification was carried out based on the results obtained from Quantifiler® Human DNA Kit. SNP amplification was carried out in 2 separate triplexes (Section 2.4.9) using 0.5 ng of template, except for samples with low concentrations (<0.1 ng/µl), where the DNA template ranged from 0.06-0.24 ng. The thermal cycling was carried out in a GeneAmp® 9700 (Applied Biosystems) as described in Section 2.4.10. PCR products were purified using 0.5 µl of ExoSAP-IT Kit® (USB®, Germany) with 1.0 µl of PCR product as described in Section 2.3.10.

ABI SNaPshot™ Multiplex Kit was used to genotype SNP with SBE primer triplex method in two reactions. The reactions were performed according to the manufacturer's protocol as described earlier (2.3.14) with 0.2 µM of SBE primer triplex, these six loci for each DNA sample were profiled. Unincorporated ddNTPs were removed by using 1 µl of SAP (USB®).

Genotypes for the SNPs were detected on ABI 310 PRISM® Genetic Analyzer using the E5 run module.

2.5.10. STR Typing

STR typing was performed using the commercial AmpFℓSTR® SGM Plus® Kit (Applied Biosystems, Foster City, USA) according to the manufacturer's instructions,

except that the reaction volume was reduced by 1/4. For SNP analysis, DNA templates ranging from 0.06 ng to 0.5 ng were amplified in an STR reaction buffer consisting of 4.83 µl of GeneAmpF*I*STR® PCR Reaction Mix, 2.53 µl of AmpliF*I*STR® SGM plus® Primer set, 0.23 µl of AmpliTaq Gold® DNA polymerase at 1.25 unit/µl and 4.91 µl dH_2O. A thermal cycling GeneAmp® 9700 (Applied Biosystems) was used for the amplification with the following conditions: stage 1, 95 °C for 11 mins; stage 2, 94 °C for 1 min, 59 °C for 1 min, 72 °C for 1 min, for 28 cycles, and incubation at 60 °C for 45 mins followed by 12 °C until the samples were analysed.

1 µl of PCR product obtained above and 1 µl of AmpFℓSTR® SGM Plus® Allelic Ladder were separately diluted with 10 µl of Hi-Di™ formamide and 0.5 µl GeneScan™ 500 ROX™ size standards, in a 200 µl PCR tube. The allelic ladder and the PCR samples were then immediately placed into the genetic analyzer without a denaturation step (Butler et al., 2003). STR alleles were separated electrophoretically using ABI Prism® 310 Genetic Analyzer (Applied Biosystems) and run module filter GS STR POP 4 (1 ml) F for dye set DS- 32 (filter set F): 5-FAM (blue), JOE (green), NED (yellow) and ROX (red). The capillary electrophoresis was performed using a 47 cm capillary (Web Scientific Ltd, UK) using POP™4 polymer, and 1X electrophoresis running buffer (Applied Biosystems). Data analyses were performed using software GeneScan™ version 3.7 and GeneMapper™ ID version 3.1.

2.5.11. Extraction and Purification of Teeth samples

2.5.11.1. Cleaning

The surfaces of the teeth were cleaned from dirt and any debris. Each tooth was placed in a sterile 50 ml plastic tube, approximately 15 ml of dH_2O was added, and the tube was manually agitated approximately 10 times. The dH_2O was removed from the tube

and 15 ml of 10% bleach was added and the tube was agitated 10 times. The bleach was then removed and the teeth were rinsed in 15 ml of dH_2O to remove any trace of bleach that could interfere with the later analysis. Following the removal of dH_2O, the teeth were submerged in 95% ethanol and the tube was agitated again before the ethanol was removed.

Following cleaning, the teeth were air dried under a flow hood cabinet overnight.

2.5.11.2. Grinding

Each cleaned tooth was ground separately in a Freezer Mill (SPEX CertiPrep^{INC} 6750) following the manufacturer's instructions. The bone powder was then placed in a sterile 15 ml tube and stored at -20 °C.

2.5.11.3. Extraction

Decalcification

The removal of calcium from tooth powder before extraction can help during the extraction of DNA (Loreille et al., 2007). Approximately 100 mg of powdered tooth was placed in a 5 ml tube (SARSTEDT AG & Co. Nümbrecht, Germany). Following the protocol in Loreille et al 2007, 1 ml of 0.5 M EDTA at pH 8.0 (Sigma, UK) was added and the tube was gently shaken a few times to mix the powder and EDTA. The mixture was then placed in the fridge at 4 °C overnight (for more than 16 h). After incubation the tube was centrifuged at 2000g (spectrafuge 24 D- Labnet) for 2 mins and the supernatant solution removed leaving behind the powder.

Qiagen DNeasy® Blood and Tissue Kit

DNA was extracted from the decalcificated bone powder using the DNeasy® Blood and Tissue kit (Qiagen) with a modification. 1 ml of ATL buffer (Qiagen), 100 µl

(20mg/ml) of Protinase K (Qiagen) and 10 µl of 1 M DTT (0.13gm/ml) was added into the tube containing the bone powder. The tube was placed in a rotator (HYBAID-Micro-4) at 55 °C for approximately 72 h (until most of bone powder had dissolved). Following incubation, the sample was centrifuged at 8000 rpm for 1 min to remove any residue on the inner side of tube as a result of overnight incubation and the supernatant solution was transferred into a new 5 ml tube. 1ml of AL buffer (Qiagen) was added into the tube, the sample was mixed and incubated at 70 °C in the rotator for 30 mins. The sample was the briefly centrifuged at 8000 rpm and 1 ml of absolute ethanol (Qiagen) was added before the tube was mixed. A spin column (Qiagen) was placed into a 2 ml collection tube (Qiagen) and the extracted sample was transferred onto the column. The column was centrifuged at 8000 rpm for 1 min and the solution in the collection tube was discarded. The above step was repeated until all the extracted solution was transferred into the column. The spin column was placed onto new collection tube and 500 µl of AW1 buffer (Qiagen) was added and centrifuged at 8000 rpm for 1 min, the solution from the collection tube was discarded, 500 µl of AW2 buffer (Qiagen) was then added and the column centrifuged at 13,000 rpm for 1 min. To remove any trace of AW2, the sample was centrifuged for a further 1 min at 13,000 rpm, the collection tube was removed and the column was placed into 1.5 ml microcentrifuge tube. The DNA was then eluted using two steps. 25 µl of AE buffer (Qiagen) was added onto the spin column, incubated at room temperature for 5 mins and centrifuged at 8000 rpm for 1 min. The eluted DNA was then removed into a new 1.5 ml tube. The elution step was repeated using 25 µl of AE buffer onto the same column and the above steps repeated. The samples were stored at 4 °C until quantification.

2.5.11.4. Quantification

The concentration of extracted DNA was estimated using the Quantifiler® Human DNA Kit (Applied Biosystems) with the ABI 7500 real time PCR (Applied Biosystems). The procedure was carried out as described earlier in Section 2.2.2.1. Each bone sample was quantified in duplicate.

CHAPTER 3

IDENTIFICATION of POLYMORPHIC SNPs

3.1. Overview

The potential of SNPs as a forensic tool has been widely acknowledged over the last few years. The most attractive feature of SNPs is their short amplicon size and therefore their suitability for analysis of degraded DNA (Butler, 2007; Inagaki et al., 2004). Also, because of their low mutation rates from one generation to the next, SNPs can be used to test kinship (Sachidanandam et al., 2001). SNP mutation rates are found to be 10^{-8} compared to 10^{-3} for STRs, which are the current forensic method used for DNA profiling (Butler et al., 2007).

3.1.1. SNP Classification

The biallelic nature of SNPs provide three different genotype variations (Butler et al., 2007). If the alleles at an SNP locus are G and A, then the possible genotypes for both alleles can be GG, AA, and GA. However, classification of any SNP is based on six categories dependent on the variation of the four nitrogenous bases (A, C, G, and T) at each locus on the DNA strand. These classifications are A↔G, C↔T, A↔C, A↔T, C↔G, and T↔G, but since DNA occurs in double complementary strands (Figure 3.1), then typical basic classification of SNPs can be explained as A↔G (T↔C), C↔T (G↔A), A↔C (T↔G), A↔T (T↔A), C↔G (G↔C) and T↔G (A↔C), where the bases in the brackets represent the complementary strand (Brookes, 1999).

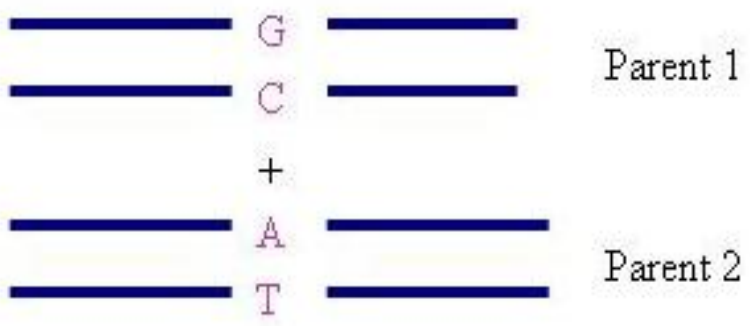


Figure 3.1. Shown above is a schematic diagram representing variation at a locus with SNP G/A on the two complementary strands. The complementary strands contain the bases C/T.

3.2. Aims of this Chapter

The main objectives of this chapter are:

- To analyse the SNPs (approximately 250,000) in 10 of Arab individuals from the United Arab Emirates and Kuwait (5 individuals from each country).

- To select 100 SNPs from all autosomal chromosomes with balanced minor and major allele frequencies. These SNPs should be distributed proportionally on the 22 autosomal chromosomes.

3.3. Methods

3.3.1. Samples

The samples used in this study were blood stains on FTA® cards, obtained from five Arab individuals from Kuwait, and blood stains on five cotton swatches, obtained from five Arab individuals from the UAE.

The purpose of including these two Arab populations was to generate in-house SNP data that could be used to identify informative SNPs for forensic purposes. Also, when this study was conducted, it was the first time that samples from UAE and Kuwaiti individuals had been used in this type of investigation.

3.3.1.1. DNA Extraction and Quantification

Extraction of DNA from the 10 samples was performed using a standard phenol/chloroform procedure following digestion with Proteinase K as described in Section 2.2.1.1. This method was selected in order to achieve a high yield of DNA template (Dixon et al., 2005a). Following extraction, the concentration of DNA was estimated using the Quantifiler® Human DNA Quantification kit (Applied Biosystems) with the ABI 7500 real-time PCR. Samples with insufficient concentrations (< 50 ng/µl) were amplified using phi 29 DNA polymerase, as described in Section 2.2.5. The extraction and quantification of samples was carried out as described in Sections 2.2.1.1 to 2.2.2.

3.3.2. Genotyping Methods and Techniques

3.3.2.1. Affymetrix® GeneChip® Technique

Allele Specific Hybridisation Method

Allele specific hybridisation is the basis of the Affymetrix GeneChip® system (Figure 3.2). This method is based on the annealing of a labelled amplicon containing the polymorphic site to a probe that is attached to an array (Goto et al., 2002). Annealing occurs as the amplicon contains the complementary sequence to the probe (Wallace et al., 1979). The hybridisation reaction is washed to remove any mismatch strands, enabling the complementary strands to be detected.

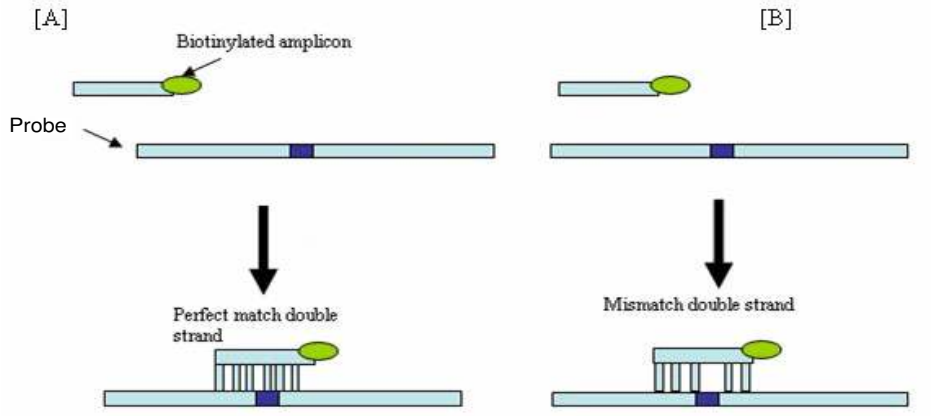


Figure 3.2. Shown above is an illustration of the allele specific hybridisation method. [A] represents a biotinylated single strand amplicon which hybridises perfectly with the complementary probe sequence to form a stable double strand; [B] represents a mismatch double strand which is removed during the post-hybridisation wash.

GeneChip® Method

The main feature of GeneChip® is the capability to detect thousands of SNPs in a single reaction. Each microarray contains sets of DNA probes with the SNP sequences that were selected from GenBank®. These probes are designed to be sensitive and specifically to hybridise only to the target sequence (Liu et al., 2003). In this project GeneChip® Mapping 250K Arrays Sty kit was used (Figure 3.3).

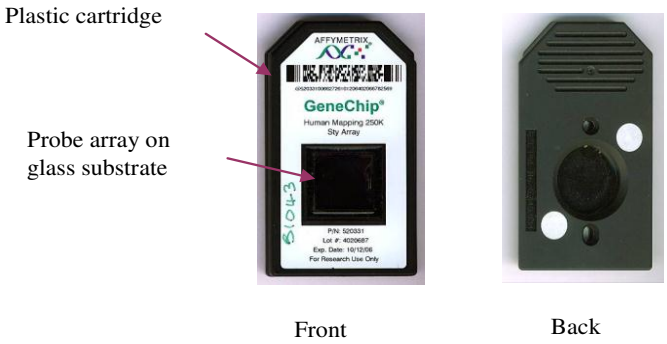


Figure 3.3. Shown above is the Affymetrix® GeneChip® Probe Array consisting of a square glass substrate mounted in a plastic cartridge. The glass contains an array of oligonucleotides mounted on its inner surface.

In this method there were three main steps: (1) PCR amplification of the DNA sequence containing the target SNP; (2) fragmentation of PCR products using endonuclease DNase I; (3) labelling of PCR products and hybridisation to the probes in the arrays (Figure 3.4).

Genomic DNA (250 ng) was digested using the restriction enzyme Sty which cut the target DNA into segments that were, on average, between 250 bp and 1,000 bp. The digested fragments become the substrate for the adapter ligation enzyme which attached an adapter. A single common primer, complementary to the adapter, was used to amplify the fragments (Matsuzaki et al., 2004). The PCR products were then fragmented by the enzyme DNase I. Finally, the fragments were biotinylated before hybridisation to the array probes by allele specific hybridisation. Subsequently, only the complementary sequences attached to array probes would be detected after purification and staining with Streptavidin Phycoerythrin. Genotyping Analysis Software (GTYPE) and GeneChip® Operating software (GCOS) were used for SNP detection.

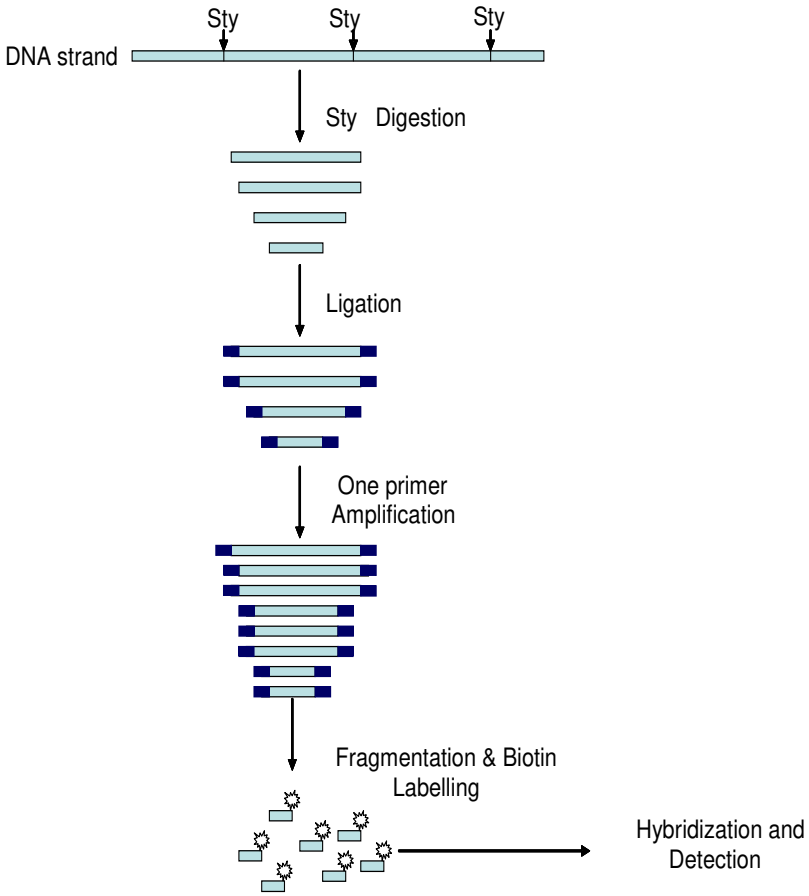


Figure 3.4. Shown above is the digestion of human genomic DNA with Sty and then the ligation of an adapter which contains a PCR primer site. The DNA is amplified, using the common primer, and the fragments are then digested by *DNAse* I to an average size of less than 180 bp, labelled with biotin, and then hybridised to the GeneChip® Mapping 250K Array. Figure 3.4 was adapted from Matsuzaki et al. (2004).

3.3.2.2. Strategies and Criteria for SNPs Selection

In order to obtain informative SNP markers, strategies and criteria were formulated. Based on the previous strategies that were described in Section 2.3.4, the selection of 100 SNPs as an initial target was carried out. The number of SNPs selected on each chromosome was in proportion to the length of the individual chromosomes (Table 3.1).

Table 3.1 Shown below are the different number of SNPs that were selected on each autosomal chromosome in the genome. The target number of SNPs selected was based on the size of each chromosome. Chromosome length was obtained from Ensembl Genome Browser (www.ensembl.org).

Chromosome	Chromosome size (Mb)	Percentage (Mb%)	Target number of SNPs
1	247	8.6	9
2	243	8.5	9
3	200	7.0	7
4	191	6.7	7
5	181	6.3	6
6	171	6.0	6
7	159	5.5	5
8	146	5.1	5
9	140	4.9	5
10	135	4.7	5
11	134	4.7	5
12	132	4.6	5
13	114	4.0	4
14	106	3.7	4
15	100	3.5	3
16	89	3.1	3
17	79	2.8	3
18	76	2.7	3
19	64	2.2	2
20	62	2.2	2
21	46.9	1.6	1
22	50	1.7	1
Total	2866	100	100

SNPs from Y and X chromosomes where eliminated from the selection, profiles of autosomal SNP exhibit high variability due to chromosomal assortment recombination and mutation leading to low match probability (Jobling and Gill, 2004). Y-chromosome is male specific and less diverse than autosomal SNPs as mutation is the only function to diversity for the Y haplotypes, therefore Y profiles show relatively high match probability (Jobling and Gill, 2004). Profiles from X chromosome showed less variation from the autosomal profiles, this due to low heterozygosity level on X chromosome; possibly due to strong selection on the X chromosome which is owing to the hemizgosity in male (Sachidanandam, et al., 2001).

Table 3.2 Quantification results for DNA in UAE and Kuwait samples used for Affymetrix® Genotyping.

Quantification values (ng/µl)		
No	UAE Samples	Kuwait Samples
1	47.8	5.8
2	90.2	5.2
3	100	4.3
4	65.4	7.2
5	126.3	5.0

3.4. Results

3.4.1. DNA Extraction

During the quantification of DNA, which was extracted from Kuwait and UAE specimens, some of the samples were found to be less than 50 ng/µl (Table 3.2).

3.4.2. Whole Genome Amplification

3.4.2.1. Phi 29(Φ29) DNA Polymerase

The DNA concentration required for the Affymetrix® genotyping method is 50 ng/µl. Therefore the samples with a concentration less than this were amplified using Φ29 DNA polymerase using the Qiagen REPLI-g® Midi kit (Figure 3.5).

The strand displacement amplification mechanisms of Φ29 DNA polymerase overcame the need for the re-extraction of the samples with, the DNA amplified directly from the original extracts.

Figure 3.5. Shown above are the results of 1% agarose gel eletrophoresis of DNA samples following whole genome amplification using REPLI-g Midi Kit. Lane I is a 23 Kb *Hin*d III ladder; lane 2 is the positive control, lanes 3-8 are Kuwait and UAE samples (samples with quantification results < 50 ng/µl) respectively.

3.4.2.2. SNP Genotyping

As specialised instruments and software were required for SNP screening using the Affymetrix technique, the DNA from the 10 samples (Section 3.4.1) were sent to an external supplier (Geneservice Ltd, UK). The SNP data were returned in the form of notepad file: for each sample a separate notepad file was supplied.

3.1.1. Analysis of SNP Data

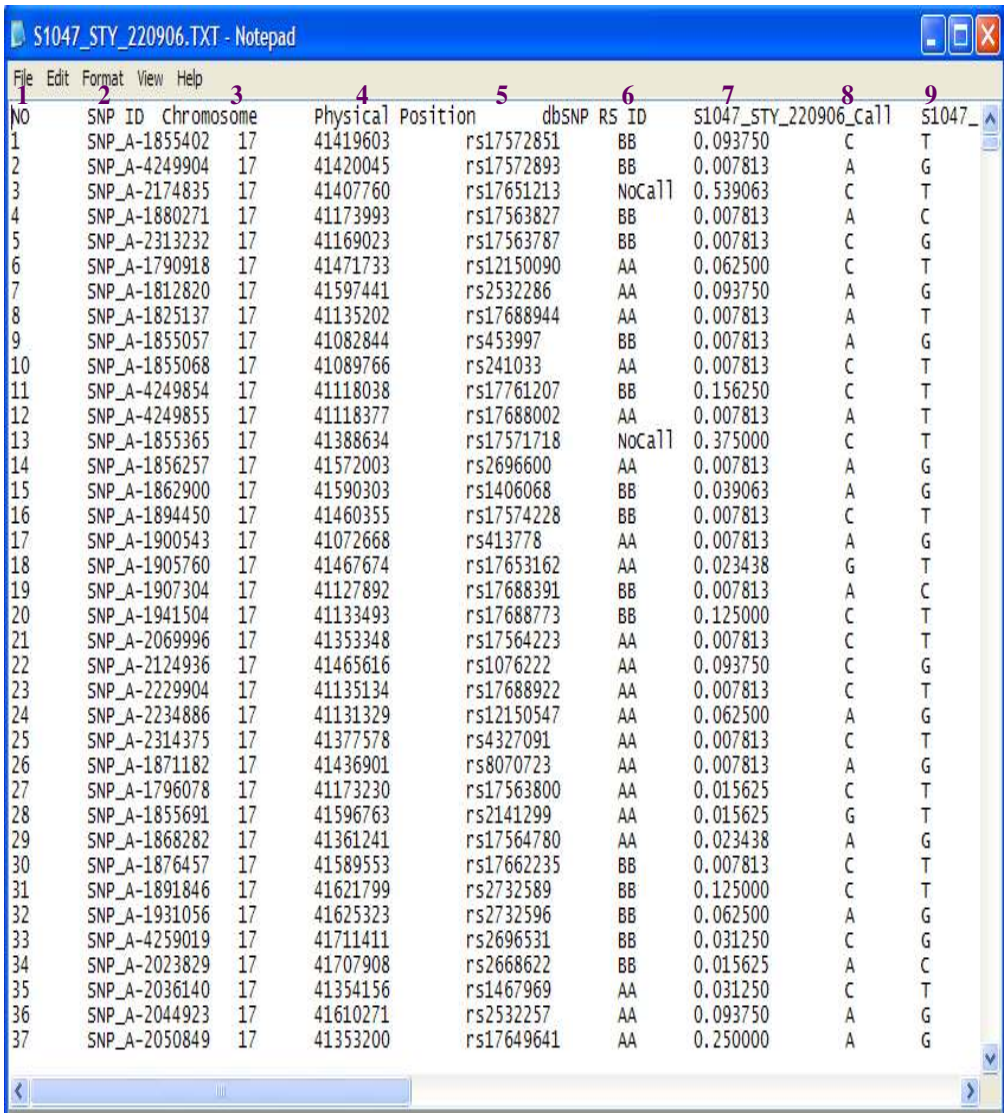
For the initial selection of SNPs and in order to process the large amount of data generated by Affymetrix (approximately 238,000 SNPs for each of the samples from Kuwait and UAE) in the form of a notepad document (Figure 3.6) the data were analysed using Microsoft® Office Access and Microsoft® Office Excel.

3.1.1.1. Microsoft® Office Access

The process consisted of two steps.

1. Copying the SNP data from the notepad documents into Microsoft® Office Access software.

The first stage was to create separate tables. Since there were 10 separate notepad documents obtained from the Affymetrix® genotyping 10 tables were designed. (Figure 3.7) and then the appropriate data from the notepad were imported into each of the tables (Figure 3.8).



S1047_STY_220906.TXT - Notepad

File Edit Format View Help

1 NO	2 SNP ID	3 Chromosome	4 Physical Position	5 dbSNP RS ID	6 S1047_STY_220906_Call	7	8	9 S1047_
1	SNP_A-1855402	17	41419603	rs17572851	BB	0.093750	C	T
2	SNP_A-4249904	17	41420045	rs17572893	BB	0.007813	A	G
3	SNP_A-2174835	17	41407760	rs17651213	NoCall	0.539063	C	T
4	SNP_A-1880271	17	41173993	rs17563827	BB	0.007813	A	C
5	SNP_A-2313232	17	41169023	rs17563787	BB	0.007813	C	G
6	SNP_A-1790918	17	41471733	rs12150090	AA	0.062500	C	T
7	SNP_A-1812820	17	41597441	rs2532286	AA	0.093750	A	G
8	SNP_A-1825137	17	41135202	rs17688944	AA	0.007813	A	T
9	SNP_A-1855057	17	41082844	rs453997	BB	0.007813	A	G
10	SNP_A-1855068	17	41089766	rs241033	AA	0.007813	C	T
11	SNP_A-4249854	17	41118038	rs17761207	BB	0.156250	C	T
12	SNP_A-4249855	17	41118377	rs17688002	AA	0.007813	A	T
13	SNP_A-1855365	17	41388634	rs17571718	NoCall	0.375000	C	T
14	SNP_A-1856257	17	41572003	rs2696600	AA	0.007813	A	G
15	SNP_A-1862900	17	41590303	rs1406068	BB	0.039063	A	G
16	SNP_A-1894450	17	41460355	rs17574228	BB	0.007813	C	T
17	SNP_A-1900543	17	41072668	rs413778	AA	0.007813	A	G
18	SNP_A-1905760	17	41467674	rs17653162	AA	0.023438	G	T
19	SNP_A-1907304	17	41127892	rs17688391	BB	0.007813	A	C
20	SNP_A-1941504	17	41133493	rs17688773	BB	0.125000	C	T
21	SNP_A-2069996	17	41353348	rs17564223	AA	0.007813	C	T
22	SNP_A-2124936	17	41465616	rs1076222	AA	0.093750	C	G
23	SNP_A-2229904	17	41135134	rs17688922	AA	0.007813	C	T
24	SNP_A-2234886	17	41131329	rs12150547	AA	0.062500	A	G
25	SNP_A-2314375	17	41377578	rs4327091	AA	0.007813	C	T
26	SNP_A-1871182	17	41436901	rs8070723	AA	0.007813	A	G
27	SNP_A-1796078	17	41173230	rs17563800	AA	0.015625	C	T
28	SNP_A-1855691	17	41596763	rs2141299	AA	0.015625	G	T
29	SNP_A-1868282	17	41361241	rs17564780	AA	0.023438	A	G
30	SNP_A-1876457	17	41589553	rs17662235	BB	0.007813	C	T
31	SNP_A-1891846	17	41621799	rs2732589	BB	0.125000	C	T
32	SNP_A-1931056	17	41625323	rs2732596	BB	0.062500	A	G
33	SNP_A-4259019	17	41711411	rs2696531	BB	0.031250	C	G
34	SNP_A-2023829	17	41707908	rs2668622	BB	0.015625	A	C
35	SNP_A-2036140	17	41354156	rs1467969	AA	0.031250	C	T
36	SNP_A-2044923	17	41610271	rs2532257	AA	0.093750	A	G
37	SNP_A-2050849	17	41353200	rs17649641	AA	0.250000	A	G

Figure 3.6. Shown above is an example of how data for approximately 238,000 SNPs was stored after Affymetrix® genotyping. The information in this example was for sample identification during analysis as S1047. Numbers: [1] represents serial number, [2] represents Affymetrix SNP ID, [3] represents the chromosome number, [4] represents the position of SNPs on the chromosome, [5] represents NCBI Database reference SNP ID (dbSNP rs ID), [6] represents the allele call type (S104-STY_220906), for example rs7572851 is BB (nucleotide TT), [7] represents confidence values, [8-9] represent allele types (A/B), for example the SNP rs7572851 is CT.

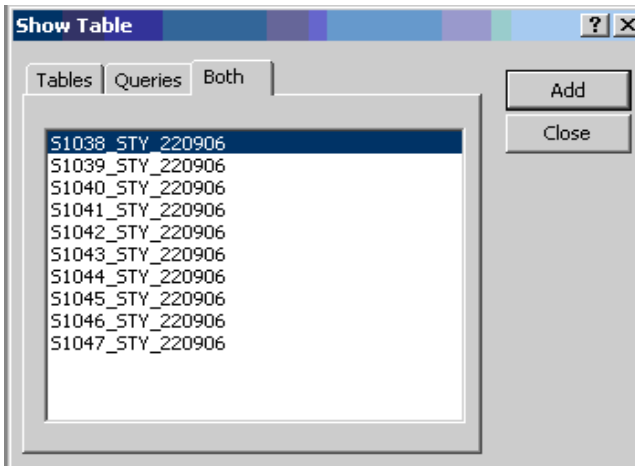


Figure 3.7. Shown above are 10 tables representing 10 different samples copied from the Affymetrix® to Microsoft® Office Access for further processing.

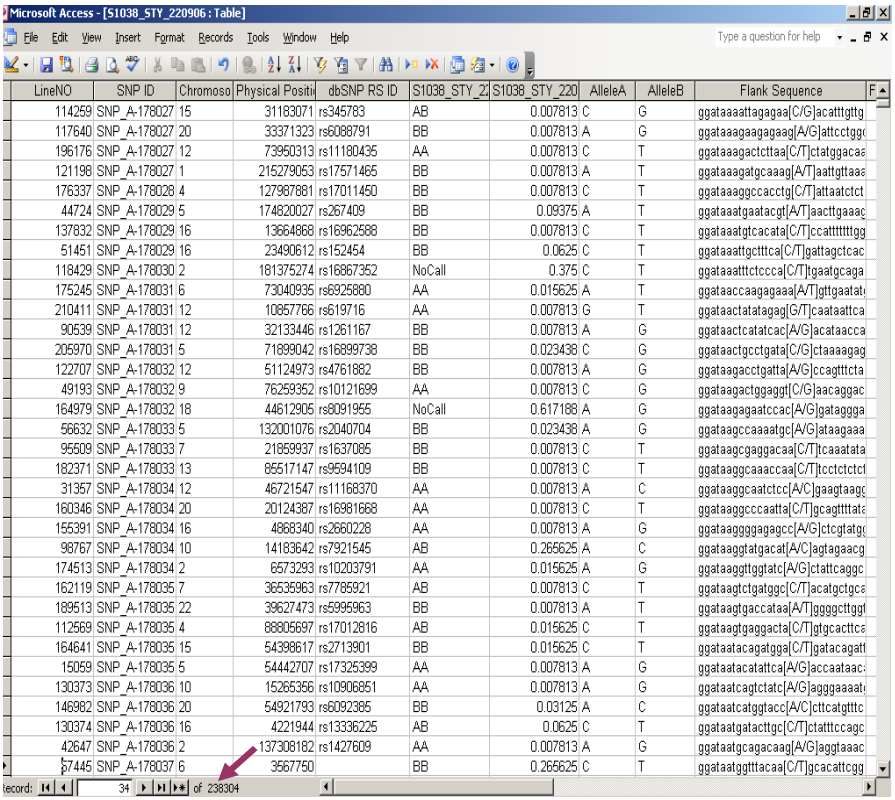


Figure 33.8. Shown above is a table illustrating how the data was presented in the Microsoft® Office Access software. The table represents one sample with the arrow at the bottom of the table indicating the amount of SNP data generated by the Affymetrix® genotyping method. The columns represent: serial number, Affymetrix SNP ID, chromosome number, database reference SNP ID, alleles call, confidence values, alleles type (A and B) and SNP flanking sequence, respectively.

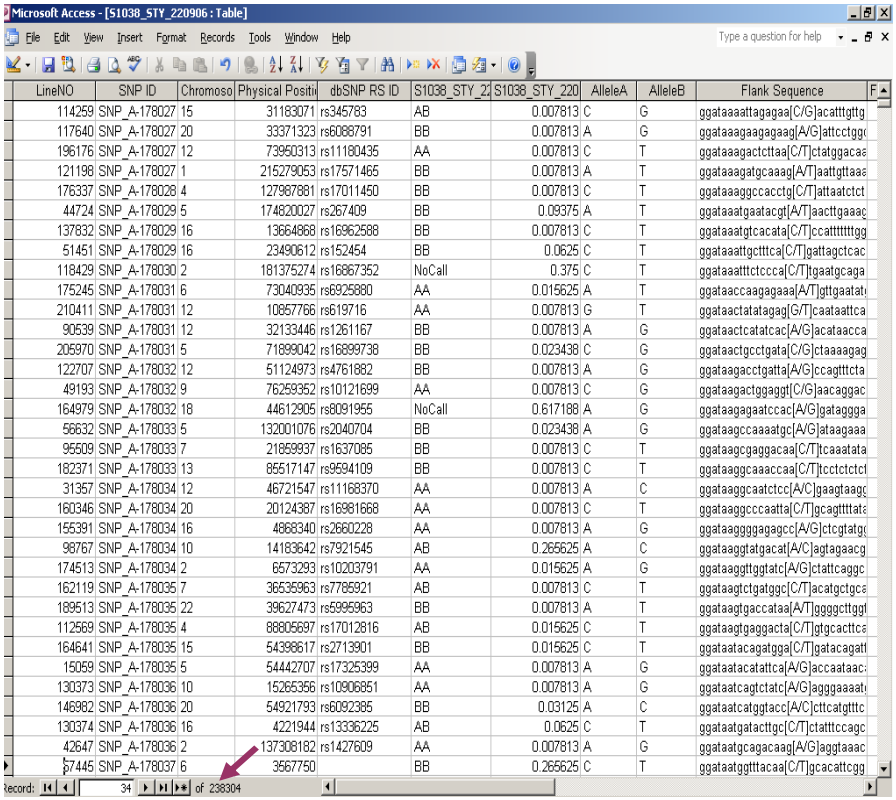


Figure 33.9. Shown above is a table illustrating how the data was presented in the Microsoft® Office Access software. The table represents one sample with the arrow at the bottom of the table indicating the amount of SNP data generated by the Affymetrix® genotyping method. The columns represent: serial number, Affymetrix SNP ID, chromosome number, database reference SNP ID, alleles call, confidence values, alleles type (A and B) and SNP flanking sequence, respectively.

2. Processing the data

The first stage was to collate the information from each chromosome, so that data from all 10 individuals would be linked (Figure 3.9).

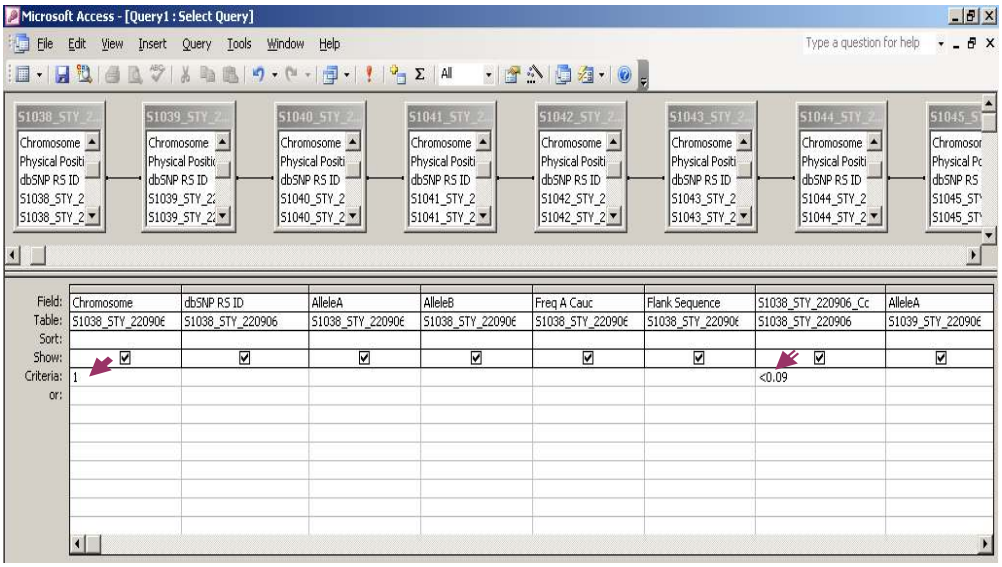


Figure 3.10. Shown above is how the 10 tables were linked together through their db SNP ID which is a part of Affymetrix® data. This allowed the 10 tables to behave as one group. The figure shows 8 tables out of the 10 due to space limitations. The arrows was showing the criteria of the confidence value (< 0.09) for chromosome number 1.

Following the linking of the data, 22 queries were carried out, representing one for each chromosome. The queries selected all data from each chromosome that displayed a confidence level of < 0.09 (greater than 91% confidence that the data is correct) which was then analysed (Figure 3.10).

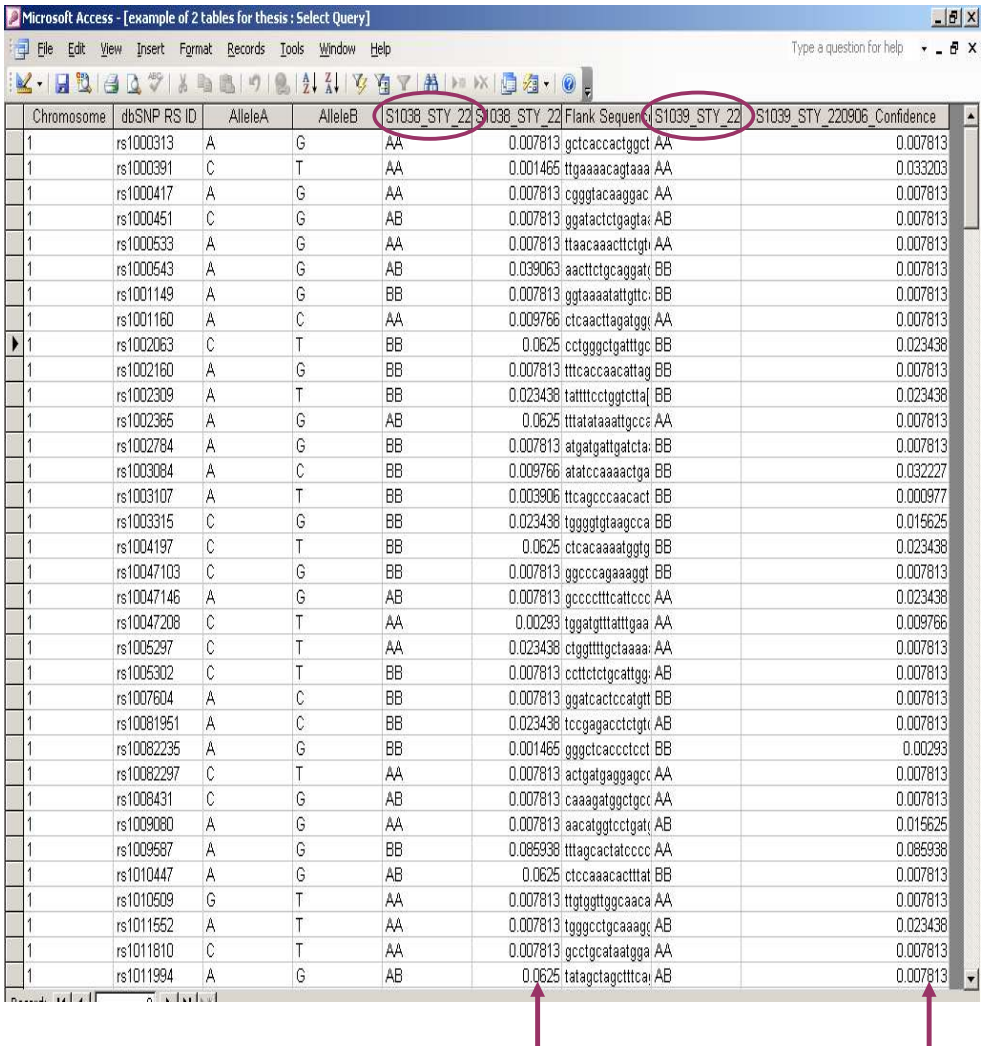


Figure 3.11. Shown above is the final output of Microsoft® Office Access. The table illustrated is for chromosome 1 and shows the data for all the 10 samples in the group. All samples share the same SNP identification through dbSNP RS ID which was set during the query design to link the samples. Arrows represent < 0.09 confidence values. The circles represent two samples.

This reduced the amount of data from the total of 229,944 SNPs which were analysed to a total of 56,826 SNPs (Table 3.3). Therefore, the final outcome of the Microsoft® Office Access process was 22 tables representing 22 autosomal chromosomes, each

with all 10 samples, ranging from 4,938 to 666 SNPs, all with confidence levels of > 91% for the data.

Table 3.3. Shown below are the different numbers of SNPs selected on different chromosomes as a result of < 0.09 confidence value was selected. Also shown is the initial number of SNPs obtained from Affymetrix®.

Chromosome Number	Initial SNPs	Selected SNPs
1	19958	4938
2	18850	4879
3	15118	3776
4	12872	3219
5	14701	3630
6	14174	3465
7	11713	2804
8	12388	3266
9	10807	2741
10	14104	3558
11	12822	3141
12	11791	3023
13	7950	2053
14	7404	1921
15	7253	1711
16	8159	2012
17	6345	1352
18	6616	1720
19	3638	696
20	6500	1561
21	3142	686
22	3647	666
Total	229944	56826

3.4.2.3. Microsoft® Office Excel

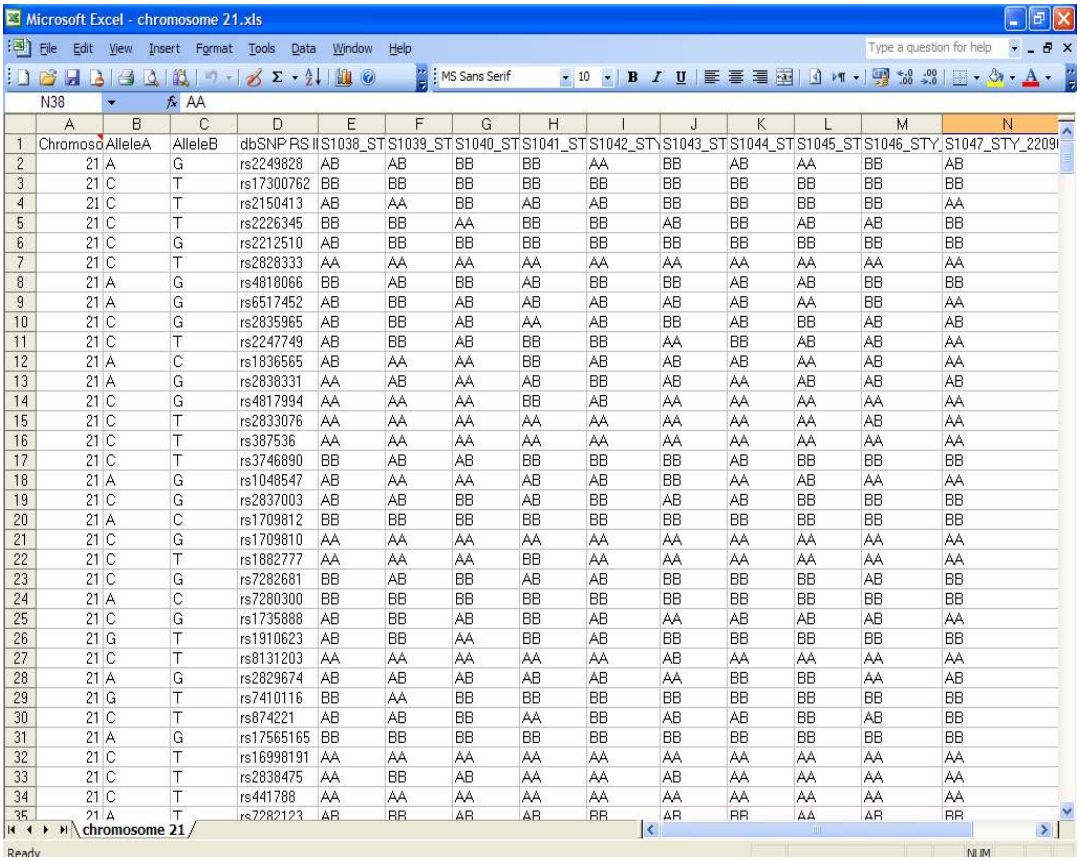
The data obtained from the Microsoft® Office Access queries were found suitable for importation into the Excel sheets. Each of the chromosome tables were imported into an

excel sheet separately. These Excel sheets were then used as the working file for the SNP data and their analysis during the entire project, unless otherwise stated.

In order to obtain the allele frequency for the SNPs data, minor modifications to the data were required.

1. Arrangement of the data

The data sheet described in Figure 3.10 was modified for the following SNPs analysis. The confidence values were removed whilst a column designated for allele frequencies was added (Figure 3.11).



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Chromoso	AlleleA	AlleleB	dbSNP RS I	S1038_ST	S1039_ST	S1040_ST	S1041_ST	S1042_ST	S1043_ST	S1044_ST	S1045_ST	S1046_STY	S1047_STY_2209
2	21	A	G	rs2249828	AB	AB	BB	BB	AA	BB	AB	AA	BB	AB
3	21	C	T	rs17300762	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
4	21	C	T	rs2150413	AB	AA	BB	AB	AB	BB	BB	BB	BB	AA
5	21	C	T	rs2226345	BB	BB	AA	BB	BB	AB	BB	AB	AB	BB
6	21	C	G	rs2212510	AB	BB	BB	BB	BB	BB	BB	BB	BB	BB
7	21	C	T	rs2828333	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
8	21	A	G	rs4818066	BB	AB	BB	AB	BB	BB	AB	AB	BB	BB
9	21	A	G	rs6517452	AB	BB	AB	AB	AB	AB	AB	AA	BB	AA
10	21	C	G	rs2835965	AB	BB	AB	AA	AB	BB	AB	BB	AB	AB
11	21	C	T	rs2247749	AB	BB	AB	BB	BB	AA	BB	AB	AB	AA
12	21	A	C	rs1836565	AB	AA	AA	BB	AB	AB	AB	AA	AB	AA
13	21	A	G	rs2838331	AA	AB	AA	AB	BB	AB	AA	AB	AB	AB
14	21	C	G	rs4817994	AA	AA	AA	BB	AB	AA	AA	AA	AA	AA
15	21	C	T	rs2833076	AA	AA	AA	AA	AA	AA	AA	AA	AB	AA
16	21	C	T	rs387536	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
17	21	C	T	rs3746890	BB	AB	AB	BB	BB	BB	AB	BB	BB	BB
18	21	A	G	rs1048547	AB	AA	AA	AB	AB	BB	AA	AB	AA	AA
19	21	C	G	rs2837003	AB	AB	BB	AB	BB	AB	AB	BB	BB	BB
20	21	A	C	rs1709812	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
21	21	C	G	rs1709810	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
22	21	C	T	rs1882777	AA	AA	AA	BB	AA	AA	AA	AA	AA	AA
23	21	C	G	rs7282681	BB	AB	BB	AB	AB	BB	BB	BB	AB	BB
24	21	A	C	rs7280300	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
25	21	C	G	rs1735888	AB	BB	AB	BB	AB	AA	AB	AB	AB	AA
26	21	G	T	rs1910623	AB	BB	AA	BB	AB	BB	BB	BB	BB	BB
27	21	C	T	rs8131203	AA	AA	AA	AA	AA	AB	AA	AA	AA	AA
28	21	A	G	rs2829674	AB	AB	AB	AB	AB	AA	BB	BB	AA	AB
29	21	G	T	rs7410116	BB	AA	BB	BB	BB	BB	BB	BB	BB	BB
30	21	C	T	rs874221	AB	AB	BB	AA	BB	AB	AB	BB	AB	BB
31	21	A	G	rs17565165	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
32	21	C	T	rs16998191	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
33	21	C	T	rs2838475	AA	BB	AB	AA	AA	AB	AA	AA	AA	AA
34	21	C	T	rs441788	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
35	21	A	T	rs7282123	AB	BB	AB	AB	BB	AB	BB	AA	AB	BB

Figure 3.12. Shown above is an example of the data arrangement in the Excel sheet for chromosome 21. Columns [B] and [C] represent the allele call type for the particular SNP, lane [D] represents the reference identifier for the SNPs and columns [E-N] represent the type of alleles type, which were designated as A and B for each of the 10 samples under study.

2. Sorting the allele genotypes

The allele genotypes generated by Affymetrix were in the form of A and B which represent the biallelic nature of the SNPs. The allele forms were changed for simplicity to the numbers 1 and 2 and only the frequency of the A allele was calculated, since the frequency of the other allele (B) can be inferred. For this, the allele B was kept blank, AA was given the number 2 and allele A (part of AB) was given the number 1 (Figure 3.12).

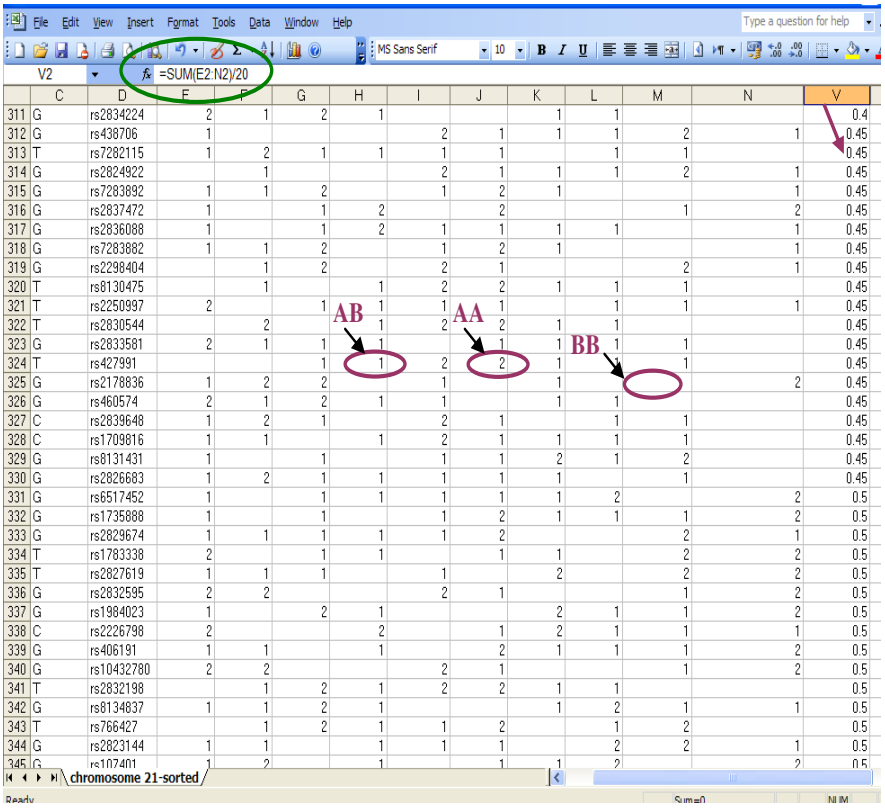


Figure 3.13. Shown above is data for chromosome 21 after the allelic designation (columns E to N represented sample 1 to 10) were changed from A and B to 1 and 2. Column V shows the ascending frequency of the alleles. The equation for calculating the frequency appears at the top of the table in a green circle.

The frequencies of the allele A were calculated using Excel and the equation (Frequency = SUM(E_n:N_n)/20) where E and N represented the cells in which the alleles were present, n represented the location of the cell in the sheet and 20 was the number of alleles under study (10 samples). The frequencies were then sorted in ascending order and the SNPs with frequencies ranging from 0.45-55 were selected and entered in a new Excel sheet (Table 3.4). A total of 4,123 SNPs were selected from 22 chromosomes. The rationale behind the selection of SNPs with allele frequencies ranging between 0.45 and 0.55 was to enhance the level heterozygosity of the selected SNPs. This in turn, will maximise the information for each SNP locus, thereby producing low match probability which is essential for forensic application (Kidd et al., 2006).

Table 3.4. Shown below are the different number of SNPs selected with frequencies ranging from 0.45- 0.55, from 22 autosomal chromosomes.

Chromosome Number	Number of SNPs
1	317
2	418
3	279
4	238
5	253
6	262
7	227
8	251
9	209
10	241
11	191
12	197
13	140
14	157
15	127
16	160
17	97
18	112
19	40
20	117
21	53
22	37
Total	4123

3.4.3. Interpretation Criteria of SNP Selection

The first step was to target SNPs that were located in the intergenic region and (Figure 3.13) shows an example of an SNP that meets the criterion of being located in such a region.

A second criterion was that SNPs should occur at a distance of at least 100 bp from any other characterised polymorphisms (Figure 3.14).

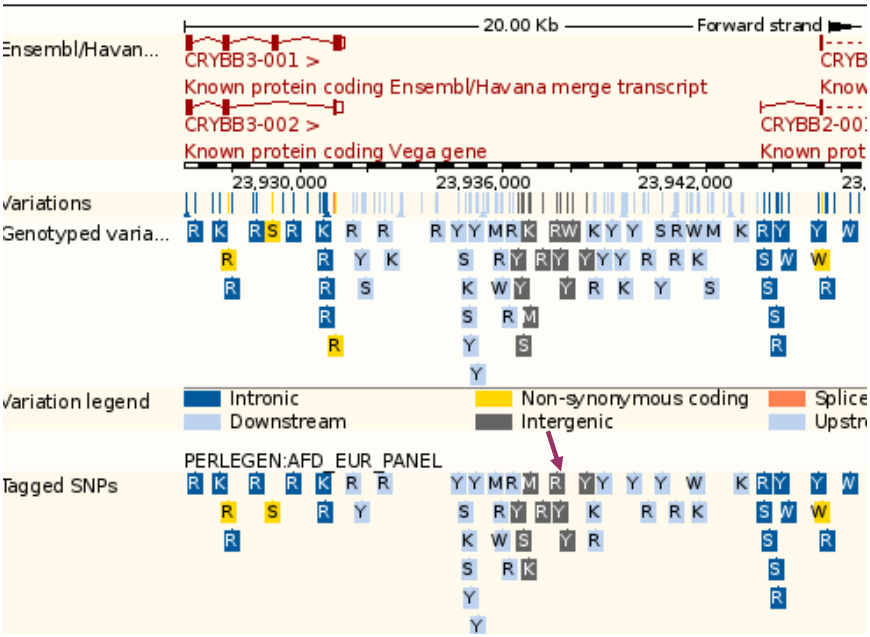


Figure 3.14. Shown above is an example of the different locations of SNPs on a chromosome. The grey colour indicates that SNPs are located in an intergenic region. Other colours indicate SNPs that are located in genic regions. The arrow represents the target SNP selected for code rs4820621 on chromosome 22.

Figure 3.15. Shown above is an example of a target SNP with no SNP within 100 bp. The arrow represents the target SNP 1-1 (rs12041851) on chromosome 1.

In other cases, such as for SNPs rs11892626, rs7573184, rs1445561, rs7858174, rs180921, rs8057434, rs17304618 and rs4820621, which occur on chromosome 2, 8, 9, 10, 16, 19 and 22 respectively, although these SNPs failed to meet the criterion of having no other SNPs within 100 bp, they were not rejected at this point (Figure 3.15). This did not have any negative impact on the results, as care was taken during primer design to avoid the overlapping SNP (Chapter 4).

No SNPs were found that were in close proximity to the commonly used forensic STRs. Some examples are shown in Table 3.5.

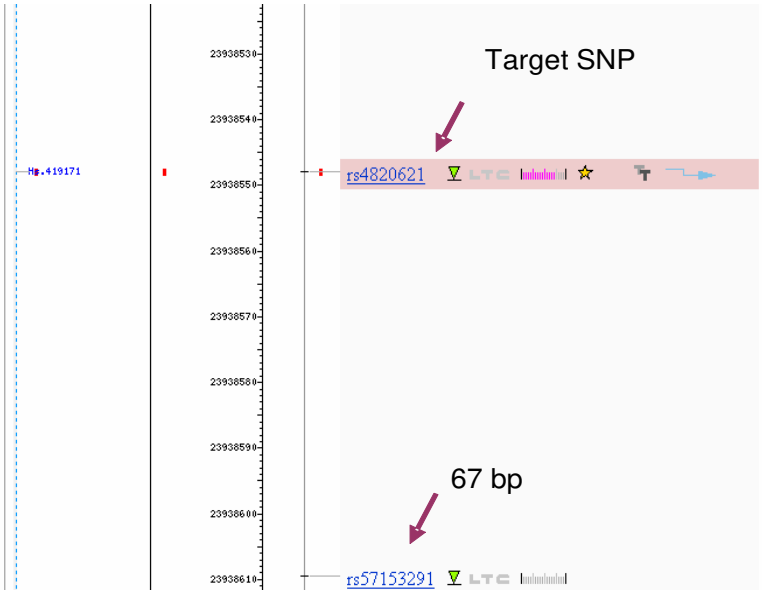


Figure 3.16. Shown above is an example of a target SNP which is located within 100 bp of other neighbouring SNPs. The figure represents target SNP 22 (rs4820621) with an SNP located 67 bp downstream of the target SNP.

Table 3.5. Shown below is an example of the positioning of SNPs and STRs that are found on the same chromosome.

Chromosome	SNP		STR	
	db SNP RS ID	position (Mb)	reference	position (Mb)
1	rs4951124	203.049170	F13B	195.3
2	rs75580941	150.753046	TPOX	1.541580
3	rs978979	56.508056	D3S1358	45.520600
4	rs4975214	130.470433	FGA	155.723730
19	rs10414856	33.595469	D19S433	35.1
21	rs8130475	33.114415	Penta D	43.9
			D21S11	19.5

3.4.4. Selection of Candidate SNP loci

The number of SNPs selected on each chromosome was proportional to the size of the chromosome. Chromosomes 1 and 2 had the greatest number of selected SNPs with 9 and 6 respectively. Most SNPs selected were from both distal regions of the p-arm and q-arm of the chromosome. Except for loci on chromosomes 13, 14, 15, 18 and 19, where SNPs were selected from the q-arm only, due to a lack of suitable loci on the p-arm. Subsequently, for initial screening, a total of 75 SNPs were selected from the 22 autosmal chromosomes (Table 3.6).

Table 3.6. Shown below are the 75 autosomal SNPs selected for analysis and their corresponding chromosomes.

No	chromosome	In-house SNP code	db SNP RS ID
1	01	1-1	rs12041851
2	01	1-2	rs10864499
3	01	1-3	rs4951124
4	01	1-4	rs4652245
5	01	1-5	rs12759915
6	01	1-6	rs1202593
7	01	1-7	rs2982742
8	01	1-8	rs576736
9	01	1-9	rs10864713
10	02	2-1	rs4832461
11	02	2-2	rs1250915
12	02	2-3	rs11892626
13	02	2-4	rs7573184
14	02	2-5	rs6542461
15	02	2-6	rs7580941
16	03	3-1	rs2649734
17	03	3-2	rs6807414
18	03	3-3	rs6793629
19	03	3-4	rs12629514
20	03	3-5	rs978979
21	04	4-1	rs1822841
22	04	4-2	rs7684079
23	04	4-3	rs2546275
24	04	4-4	rs9995245
25	04	4-5	rs4975214
26	05	5-1	rs6594747
27	05	5-2	rs7723568
28	05	5-3	rs7444492
29	05	5-4	rs4703439
30	06	6-1	rs6915280
31	06	6-2	rs17559298
32	06	6-3	rs1570281
33	06	6-4	rs3846764
34	07	7-1	rs217013
35	07	7-2	rs1525830
36	7	7-3	rs7786414
37	8	8-1	rs4105594

Table 3.6 (continued).

No	chromosome	In-house SNP code	db SNP RS ID
38	8	8-2	rs1445561
39	8	8-3	rs9297236
40	9	9-1	rs7858174
41	9	9-2	rs10491520
42	9	9-3	rs10965215
43	10	10-1	rs180921
44	10	10-2	rs555325
45	10	10-3	rs12764177
46	11	11-1	rs517679
47	11	11-2	rs2941043
48	12	12-1	rs6487665
49	12	12-2	rs10777845
50	13	13-1	rs4435117
51	13	13-2	rs4941487
52	13	13-3	rs7338627
53	13	13-4	rs2892545
54	14	14-1	rs17095615
55	14	14-2	rs11628091
56	14	14-3	rs1489870
57	14	14-4	rs10133956
58	15	15-1	rs4778706
59	15	15-2	rs3848179
60	15	15-3	rs1529883
61	16	16-1	rs8057434
62	16	16-2	rs7204754
63	16	16-3	rs1477389
64	17	17-1	rs4925075
65	17	17-2	rs2045660
66	17	17-3	rs1872236
67	18	18-1	rs4891524
68	18	18-2	rs17064977
69	18	18-3	rs9950394
70	19	19-1	rs10414856
71	19	19-2	rs17304618
72	20	20-1	rs6098780
73	20	20-2	rs745661
74	21	21	rs8130475
75	22	22	rs4820621

3.5. Discussion

The allele specific hybridisation method incorporated in the Affymetrix® microarray technique provides a reliable genotype of tens of thousands of SNPs with information such as the allele's types, the position of each SNP on the chromosome and the flanking sequence (Thompson et al., 2005).

Evaluation of Affymetrix® Results

The use of the Affymetrix® GeneChip 250K Array Sty genotyping method allowed the generation of more than 238,000 SNPs from the whole genome.

High quantity DNA (more than 250 ng/5µl) was required for the Affymetrix® genotyping method. In order to obtain such concentrations, whole genome amplification was used. For this, the double properties of the Φ29 enzyme, as a DNA polymerase and exonuclease, were employed. The DNA polymerase activity of the enzyme incorporated nucleotide bases at the 3′ end of the primer whilst its exonuclease activity cleaved nucleotides at the 5′ end of the double stranded DNA (Perez-Arnaiz et al., 2006). This double action resulting high DNA concentrations, ranging within DNA fragments from 2kb to 100 kb (Qiagen, 2005). Due to Φ29 polymerase activity, the concentration of amplified DNA can increase to more than 10 times the expected level using *Taq* DNA polymerase amplification (Schneider et al., 2004).

As the Affymetrix® technique generated large data sets powerful software such as Microsoft® Office Access was needed. This allowed the data to be analysed and stored in tables which were then exported into Microsoft® Office Excel. The use of publicly available GenBank sites such as HapMap, NCBI, and Ensemble, and the criteria formulated in this study (Section 2.3.4) allowed the selection of 75 SNPs to be further characterised for forensic applications.

Comparison with other SNP Methods

In order to evaluate the identification of SNP results obtained in this study by using the Affymetrix® GeneChip® method, some autosomal SNPs generated with different SNP typing methods were assessed.

Inagaki et al. (2004) developed a 39-plex autosomal SNP including the amelogenin locus. The multiplex was based on SBE reactions in 5 tubes using the SNaPshot™ method. The 39 SNPs were selected from different SNP databases including, Japanese SNP (JSNP) database.

Vallone et al (2005) developed 70 autosomal SNPs markers typed in 11 tubes of 6-plex and a single 4-plex reaction. The allele discrimination was performed using SNaPshot™. All SNPs were obtained using the Orchid Cellmark (Dallas, USA) Autosomal SNP Information. These 70 loci were selected from 20 autosomal chromosomes. The polymorphism loci used involved one SNP type only (C/T).

Another collection of SNPs was selected by Kidd et al (2005) from Applied Biosystems off the shelf. 19 SNPs TaqMan markers were developed in their study.

More recently, Sanchez et al (2006) developed a 52 autosomal SNP multiplex in two separate reactions, a 29-plex and 23-plex PCR and SBE for SNaPshot™ reaction. The selection was based on 'The SNP Consortium', the SNPs were selected from all autosomal chromosomes.

The 75 SNPs identefied in this study were obtained from screening Arab individuals using the Affymetrix® GeneChip® rather than selecting SNP from available GenBank databases. The objective in using this screening method was to generate SNPs markers that were obtained from individuals (UAE and Kuwait) with SNP profiles not included in the GenBank® database (at the time the research was conducted). In comparison, all

of the above described methods used SNP loci initial screened from genotyped populations available at the GenBank® databases. Also, the pre-selection of SNPs from Affymetrix® data was based on 0.45-0.55 frequencies. In addition, during the selection all the autosomal chromosomes were targeted, as were 52 SNPs developed by Sanchez et al. (2006). In this study and others, selection from entire autosomal chromosomes helped to select unlinked SNPs. Studies have reported that Linkage Disequilibrium (LD) (the association between SNPs) is reduced when SNPs are selected to be 100 kb from each other (Phillips et al., 2004; Sanchez and Endicott, 2006).

With the large amount of SNP data (238,000 SNPs per sample) obtained using the Affymetrix® screening, a significant period of time was needed in order to process the data. This was a disadvantage of screening such a high number of SNPs. Affymetrix GeneChip® of less than 250 kb could therefore be a more appropriate method for screening forensic SNP markers.

3.6. Conclusion

In conclusion, regardless of the time taken for SNP identification, a 75 autosomal SNP panel was selected. The SNPs have been selected for high heterozyosity in the target individuals. Further characterisation is to be carried out in the following Chapters to select the best SNPs for forensic applications.

CHAPTER 4

ANALYSIS of SNPs using SNaPshot

4.1. Overview

Completion of the Human Genome Project provided billions of base pairs of DNA sequence to the scientific community (Reich et al., 2001). This work identified the positions of more than 5 million SNPs, providing more understanding and information for the study of human genetics (Sachidanandam et al., 2001; Venter et al., 2001; Collins et al., 2004) and another tool for forensic applications (Budowle, 2004).

4.2. Aims of this Chapter

- To design a series of assays to evaluate the utility of SNPs identified in Chapter 3, as markers for forensic applications. Essentially, this involves amplifying these SNPs on a PCR amplicon followed by genotyping using a single base extension method (SNaPshot);
- To perform a concordance study between SNapShot™ and Affymetrix® genotypes.

4.3. Results

4.3.1. Assessment and Evaluation of SNPs

The SNaPshot™ kit was used to characterise individual SNPs which were selected after analysis using the Affymetrix® GeneChip® 250K Array Sty, as described in Chapter 3. Accurate primer design and rigorous purifications were necessary steps for characterisation in order to achieve unambiguous results when using the SNaPshot method (Sanchez and Endicott, 2006). The procedure was conducted according to the manufacturer's protocol (Figure 4.1).

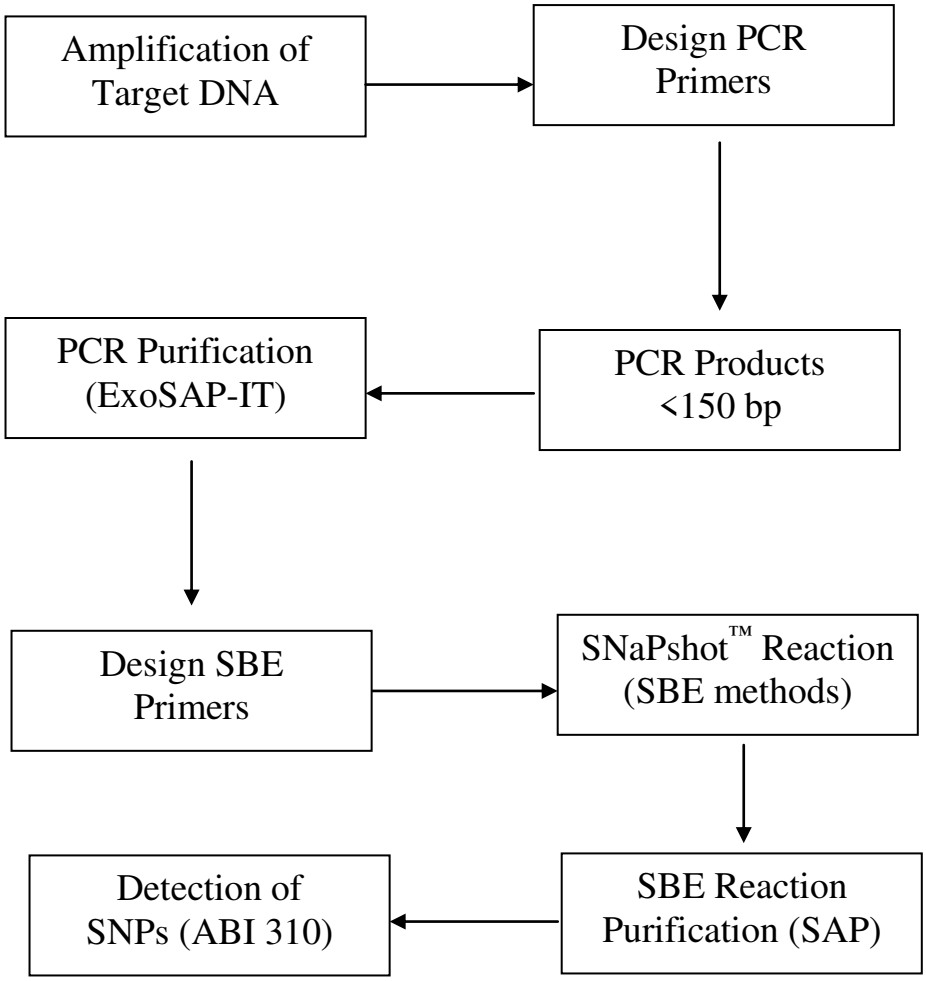


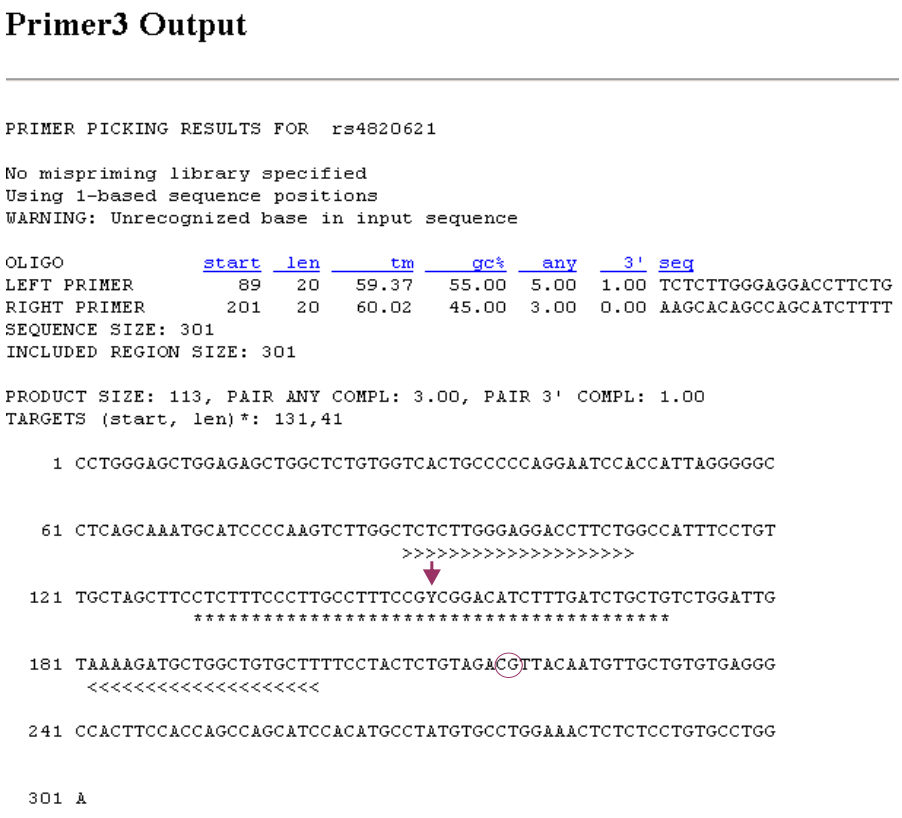
Figure 4.1. Shown above is a flow diagram describing the steps in the SNaPshot™ protocol. Template DNA was amplified producing PCR products less than 150 bp long. Excess primers and dNTPs were removed by the addition of ExoSAP-IT enzyme. Purified PCR products were then analysed by a SNaPshot™ reaction in the presence of SBE primers, followed by a final purification step in which Shrimp Alkaline Phosphatase (SAP) was added to remove unused ddNTPs. Finally, the ABI Prism 310 Genetic Analyser was used to detect SNPs.

4.3.1.1. **PCR Primer Design**

In total, 150 PCR primers were designed, fulfilling the criteria described in Section 2.3.5. Primer 3 (http://www.fro.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) was used to aid primer design, (Figure 4.2) the sequence data flanking the SNP was imported

from the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov).

The primer sequences were checked using Oligonucleotide Properties (http://www.basic.nothwestern.edu/biotools/oligocalc.html), which analysed these sequences for putative secondary structure. Finally, the primer sequences were checked using the Basic Local Alignment Search Tool (BLAST), to check for both specific and non-specific binding sites. Ultimately, 75 primer pairs were designed (Table 4.1).



Primer3 Output

PRIMER PICKING RESULTS FOR rs4820621

No mispriming library specified
Using 1-based sequence positions
WARNING: Unrecognized base in input sequence

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	89	20	59.37	55.00	5.00	1.00	TCTCTTGGGAGGACCTTCTG
RIGHT PRIMER	201	20	60.02	45.00	3.00	0.00	AAGCACAGCCAGCATCTTTT

SEQUENCE SIZE: 301
INCLUDED REGION SIZE: 301

PRODUCT SIZE: 113, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 1.00
TARGETS (start, len)*: 131,41

```
    1 CCTGGGAGCTGGAGAGCTGGCTCTGTGGTCACTGCCCCCAGGAATCCACCATTAGGGGGC

   61 CTCAGCAAATGCATCCCCAAGTCTTGGCTCTCTTGGGAGGACCTTCTGGCCATTTCCTGT
                                 >>>>>>>>>>>>>>>>>>>

  121 TGCTAGCTTCCTCTTTCCCTTGCCTTTCCGYCGGACATCTTTGATCTGCTGTCTGGATTG
            ****************************************

  181 TAAAAGATGCTGGCTGTGCTTTTCCTACTCTGTAGACGTTACAATGTTGCTGTGTGAGGG
       <<<<<<<<<<<<<<<<<<<

  241 CCACTTCCACCAGCCAGCATCCACATGCCTATGTGCCTGGAAACTCTCTCCTGTGCCTGG


  301 A
```

Figure 4.2. Shown above is PCR primer design for SNP code 22. The arrow shows the position of the target SNP and the circle shows the position of another unrelated SNP.

Table 4.1. Shown below are 75 PCR primers sorted by chromosome position. Each primer set consists of a forward and reverse primer. The predicted annealing temperature and amplicon length is shown.

Chromosome	In-house SNP code	NCBI ref	PCR Primer[a]	PCR annealing temp (°C)	Amplicon length (bp)
01	1-1	rs12041851	CCTGATTTATGAGAGGAGCTGA	60	137
			GCCTGCACTGCACATTCTA	57	
01	1-2	rs10864499	GATCAAAGGGGAGAGCACAC	60	127
			CAAGGAGTAGGCCAGGTTCC	63	
01	1-3	rs4951124	GACGACAAGTTACCTGCCTGA	61	144
			TCAGGGGTCGAACTAGACCTT	61	
01	1-4	rs4652245	GGAAGAAATGGAGTAAGGATGA	60	124
			CACCCCCTTCAACTCAGTCT	58	
01	1-5	rs12759915	AGTGCAAATGGGAAGAAAGG	56	113
			CAGAAAGTGTCAGGAGGGCTA	61	
01	1-6	rs1202593	GCAATGGGCAGTAGATCAAG	58	122
			AGGGCAGCATCTGGAATAAC	58	
01	1-7	rs2982742	GGACACACTCTAATTTCTCCATGT	62	115
			CAAAGGAGTTAATAGTCCCATTGT	60	
01	1-8	rs576736	CTCACTTAGCCTCACAACAACC	62	140
			TGGGTGAGTCTCCTTGTTCA	58	
01	1-9	rs10864713	CAATATCCAATCCACCAGCA	56	96
			GGACTAAGGTTCCTGCCAGA	60	
02	2-1	rs4832461	CACCGATCTCAGCCTGGTAA	58	108
			CATATCTTTGGAGCCCTGGA	60	
02	2-2	rs1250915	GCAAATAATCTGGTGGCTGAG	60	115
			TCCAGGTTCAAACCGAATGT	56	
02	2-3	rs11892626	AGATGCACCCTCCTAGAGCA	60	112
			TCAGAGTGAGGGGAATAGCTG	61	
02	2-4	rs7573184	TCCCAGATGACCAGAAACCT	58	120
			GAGCCTTGTCTTCTTTCCACA	60	
02	2-5	rs6542461	TTTAAGCCCTTGGTTCATGTG	57	130
			CCAGTGTTCTGATTCCAGCA	58	
02	2-6	rs7580941	CTTTCCTTCTGGCTTCTTGG	58	121
			ATGAGAAGTCTGCCAAAGCAA	57	
03	3-1	rs2649734	GAATGGCACTCTGGTGGAGT	60	139
			AGGACTGAGAGAGGGACACCT	63	
03	3-2	rs6807414	TGAAAGAGAAAGATGGTGTGAAA	58	107
			TGGAACACCAACAGTGTATGC	60	
03	3-3	rs6793629	AGACTGATTCTCTAGGCAGAGC	62	116
			CACAGTGTCCTCTTGAACACG	61	
03	3-4	rs12629514	TTGGCAGATAGCATTATCAGGA	58	136
			AGGCCACTGTTCATTTCCAG	58	
03	3-5	rs978979	TTGCCACTTCCTAATTGTCTGA	58	128
			TTTATCATTTCTCTTCCCTTCCA	58	
04	4-1	rs1822841	CCAAACTTCCGCTTAATGTTACC	61	129
			GCAAAGCTCATGTATGTAGAA	58	
04	4-2	rs7684079	CATTCTACCCTGGCCTGAGC	63	130
			ACCAGAAAGAGGAGGGAGGA	60	

Table 4.1 (continued)

Chromosome	In-house SNP code	NCBI ref	PCR Primer[a]	PCR annealing temp (°C)	Amplicon length (bp)
04	4-3	rs2546275	AGGACAGTTGGCCAAATACAAT	58	120
			CACAGGTTCATCCAAGAGCA	58	
04	4-4	rs9995245	CAGGTGAAACAAATAGCCAGAA	58	90
			GAGAAGCTTCCACCTGAATTTG	60	
04	4-5	rs4975214	GATGGGTAGGTTTATCCAAGG	60	124
			TTGACAGAGCATTACTGGTTCTT	59	
05	5-1	rs6594747	GGAAAAGCAAGTGCCATTATTTA	58	140
			GCCTCAGGGCTCTATTCTTTG	61	
05	5-2	rs7723568	GTGGAGTGAAGCCCTGAATG	60	129
			ACAGATGGCAGAAGGCAGAG	60	
05	5-3	rs7444492	GGGTTAAACAAAGGAGAAATGC	58	130
			AATCACTTGCCCAAGGTCAC	58	
05	5-4	rs4703439	CTGTGGGAAGTGGATGCTG	60	123
			ACTCCGAGCTCTTCCTCTGA	60	
06	6-1	rs6915280	TGGACACTTACTGAGTTCCTCTTT	62	111
			TTCACCGTTATTCCGAGAGC	58	
06	6-2	rs17559298	ACCCCGTGTCCACATAGTCT	60	98
			ACAGTTTCCAAAGCCAGAGC	58	
06	6-3	rs1570281	GGGATTTGATCTGCTTTATTCTC	59	116
			ATCTGCCAGCCATTGTCTTC	58	
06	6-4	rs3846764	TCTAGGTAATAAACTGGGTTTCCA	60	106
			GAGGTAAAAGCTGCCCTTGA	58	
07	7-1	rs217013	GCAGCGAATACCAGGCTC	60	120
			GCAGCAAGGTAAGAAAAAGCA	57	
07	7-2	rs1525830	CCTTCTTATCATGTCACGTTGG	60	118
			AAAGGTCACATGACGGTGGA	58	
07	7-3	rs7786414	GGGGTCTTGAGATGTTGCAG	60	119
			GCTGTGGTTCTTGGTGACCT	60	
08	8-1	rs4105594	GGGTCGGCTTATTTCTCACA	58	102
			CATTTCCCCAGCTATGGTGT	58	
08	8-2	rs1445561	TGCCAGAGGAAGGTGTATCA	58	112
			GCTGTAGACATTAGGGCACCA	61	
08	8-3	rs9297236	AGACTGGGAAACTGAAGTGTGA	60	105
			CAGGGGAAGTAGGGCTAGAAA	61	
09	9-1	rs7858174	GGTCAAATGCCAAGTGAAGC	58	106
			CCCTTCTCAAGACCACCTGA	60	
09	9-2	rs10491520	CCTTCCCCCTTAATCTGTCC	60	119
			GGCTATGCCCCTTTTGCTAT	58	
09	9-3	rs10965215	TCCTGATGGAATGTTTAGTCTGA	59	135
			CAGCATGGACACCAATATTCTC	60	
10	10-1	rs180921	GTATCCTGGGGGCAATTTCT	58	131
			TGATCTGCTTTTACGTCTTATCTCC	63	
10	10-2	rs555325	ACTGCAGGTGCTCGTTGTCT	60	104
			CTGATCCCCTTCCCTCTCTT	60	
10	10-3	rs12764177	TTGTAGCCAGGAATCTGGTTG	60	118
			CTTCAGGTTCTCTAGGGTGGA	61	

Table 4.1 (continued)

Chromosome	In-house SNP code	NCBI ref	PCR Primer[a]	PCR annealing temp (°C)	Amplicon length (bp)
11	11-1	rs517679	GCCAGATGAGGACTGTGTTG	60	120
			TGAGCTGCTACAGATTTATGCTACA	63	
11	11-2	rs2941043	CCTCTAGGATGCCAAGCAGT	60	117
			CTTTGGTTCTTCGACCTGTAAA	58	
12	12-1	rs6487665	GGGCCTGAGTCAATTTTCAG	58	119
			TGAAGAAGGACTAAGGGAATCA	58	
12	12-2	rs10777845	CCCTTGAATCCTCATGGAGTT	60	108
			CACAACATTATTGGGCGGCTA	60	
13	13-1	rs4435117	AGTTCCTGCCTAACATTCCTG	60	119
			AGATCAGTTCCACCTCCCACT	61	
13	13-2	rs4941487	ATGGCCACCTAGGGAAACTT	58	126
			TCCTCTTTTGTTGACACCTTG	57	
13	13-3	rs7338627	ACACAGCTGCCCAGGAAAAG	60	112
			TGCTGCTAACTCTGGACTGG	60	
13	13-4	rs2892545	ATCTGCATGAGTTCCTTTCAA	57	142
			GTACGGTGGGTCCTCGAAAA	60	
14	14-1	rs17095615	GCTCCCTCGACCGATTTTAT	58	117
			AACCTAACCCCCAAGGCAAT	58	
14	14-2	rs11628091	TCCCTCACTCCTGGAAACAC	60	118
			ATGAGGAGGGACCAACCAAG	60	
14	14-3	rs1489870	TCATGTTCTCAGGGTACTTGGTT	61	113
			TGCAGCAATCCAGACTGAAC	58	
14	14-4	rs10133956	AGCAGAGTTGCGTAAAGCAG	58	95
			GAACTCGAATCCAGGTCTCC	60	
15	15-1	rs4778706	AGCCCCACGCAAATGTATGT	58	120
			TTGAAGGAGGCAGTTGATCTC	60	
15	15-2	rs3848179	GTCAGGCTGGAAATGGTAAGA	60	139
			TGACTCATCCGACTTTACTTTTCT	60	
15	15-3	rs1529883	GGTCATCCTCCAAAGAACACA	60	126
			TGGCACTTCATTGCTGACTC	58	
16	16-1	rs8057434	GCCATCACTGTGTGAGCAAG	60	127
			CCATGCTTTCCATTTCTACTCC	60	
16	16-2	rs7204754	CAAGCTAAATAAATGGCCAAGG	58	133
			AGAGAGATCTTGGGGGAC GT	61	
16	16-3	rs1477389	CATGGCAGTTTCTTATTTCTGG	58	120
			GAGCTCCAATTTAACGCCATC	60	
17	17-1	rs4925075	TTGATTTTTGGCTAGCATTTAGG	58	119
			GGATGACTCCAGACCAATGC	60	
17	17-2	rs2045660	CCATCCCCAGCCTACCTA	58	144
			GCAGCATTTAAACAGGCTTTCT	58	
17	17-3	rs1872236	GCTCCGAGTCAGGTCTTGAA	60	147
			GGAAGAAGAGCCGACATCCT	60	
18	18-1	rs4891524	TGAGGCCAATCTTATCTTCTTGA	58	108
			GAGTAACCTGCGTGGAAGGA	60	
18	18-2	rs17064977	GAACACCTGGGGAAAGAACA	58	109
			AATGCCCAGGACCTCACTTT	58	

Table 4.1 (continued)

Chromosome	In-house SNP code	NCBI ref	PCR Primer[a]	PCR annealing temp (°C)	Amplicon length (bp)
18	18-3	rs9950394	TGCTGTTCCCATGGTAGTGA	58	119
			GGGGAAGGAAAACAAGTACC	61	
19	19-1	rs10414856	TAGCAAGGTGCACATGAAGC	58	129
			TGCAGTTATTGGGGTCTATGC	60	
19	19-2	rs17304618	TTCAGTGTTCTTGGGCACAG	58	110
			ATTAGGCATCCAAGACCGCATA	60	
20	20-1	rs6098780	TGAGCATCCCTTACTTCTCCA	60	123
			GGCCATTCGGAAAGAACTGT	58	
20	20-2	rs745661	TGGGTGCAGTGAGGTAGCTT	60	110
			CTTGTTGCTCCACCTTCCTT	58	
21	21	rs8130475	TCCTCTCACAACTTGCTTGG	58	92
			TGCATGACAGTGGAAGACCA	58	
22	22	rs4820621	TCTCTTGGGAGGACCTTCTG	60	113
			AAGCACAGCCAGCATCTTTT	56	

[a] primer sequences are shown from 5′ to 3′ orientation

The predicted annealing temperature for the 150 primers was 60 °C ± 3 °C except for primers 1-5, 1-9, 2-2 and 22 where it was 56 °C. During optimisation, these primers produced acceptable amplification products at an annealing temperature of 60 °C (Figure 4.3). The G+C contents for each primer were kept between 35-60%. Moreover, at least 1 but not more than 4 G/Cs were present within the first 7 nucleotides from the 3′ end of the primer pair. Exceptions occurred with the forward primers for SNPs rs6594747 and rs17095615 (SNP codes 5-1 and 14-1 respectively). G/C bases were included in the 3′ portion of the primers to increase hydrogen bonding, which in turn enhances the specificity of the primer (Dixon et al., 2005a; Dieffenbach and Dveksler, 2003). The primer size was kept less or equal to 25 bp in length. For successful amplification of targeted regions it was found that all the 75 primer pairs could be amplified at an annealing temperature of 60 °C. Also, for the optimal performance of primers, the magnesium chloride ($MgCl_2$) concentration was adjusted to 2.5 mM.

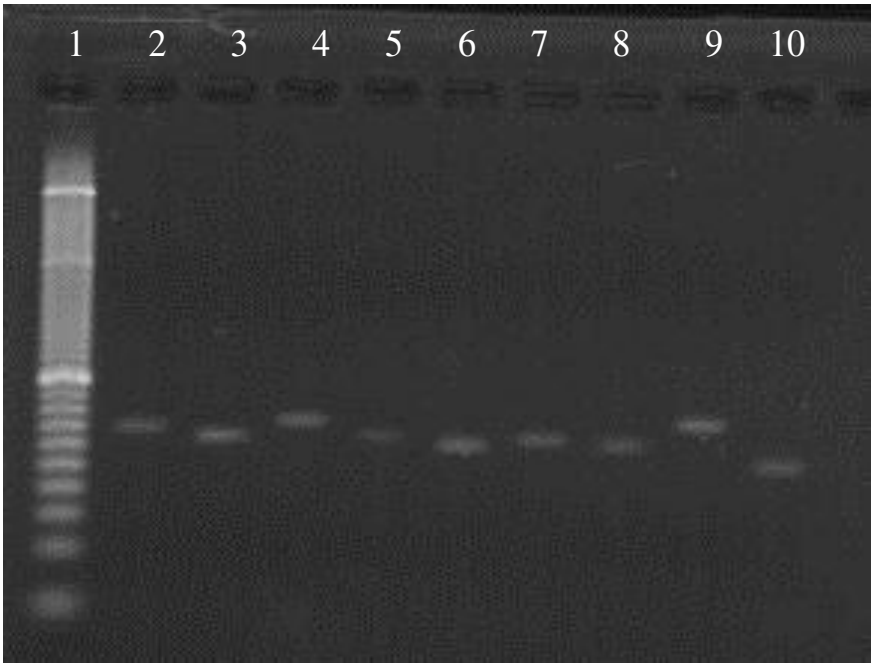


Figure 4.3. Shown above is an example of annealing temperature optimisation for chromosome 1. Lane 1 represents a 20 bp allelic ladder, lanes 2 to 10 represent an annealing temperature of 60 °C for SNP codes 1-1 to 1-9. The optimisation products were run on a 2.5% (w/v) agarose gel.

4.3.1.2. SBE Primers

The 75 single base extension primers (SBE) (Table 4.2) were designed to anneal 1 bp upstream (3′ end) to the SNP. The steps for primer characterisations were similar to the PCR primers described in Section 4.3.1.1.

The orientation of each primer was given in respect to the SNP position on the target strand (forward or reverse). Also, the orientation was dependent on the most suitable primer that met the criteria for SBE as described in Section 2.3.11. Poly-thymidine (poly-T) tails were included in some primers at the 5′ end to increase their length. The addition of poly-T tail does not have any significant effect on the annealing temperature (Dixon et al., 2005a). This was important because the properties of the fluorescent dye

used to label the ddNTPs was observed to have a more pronounced effect on the electrophoretic mobility of shorter primers in comparison to larger primers which were more than 25 bp. Therefore, all SBE primers were designed to be more than 25 bp except primers 1-1 and 15-1 which were 20 bp and 22 bp respectively. The effect of the dye on the sequence electrophoretic mobility was within the range that was expected.

Table 4.2. Shown below are 75 SBE primer sequences and their direction of orientation. F represents forward orientation and R represents reverse orientation.

Chromosome	In-house SNP Code	NCBI ref	SBE primer[a]	Direction	SNP allele
01	1-1	rs12041851	CCCTGGAGTTGGCCAAAAGA	F	A/G
01	1-2	rs10864499	TGCCCCCTCTTTCATCCACC	F	C/T
01	1-3	rs4951124	AGGGACTGGGCCTCAAGTA	R	C/T
01	1-4	rs4652245	CCAAGGTTATATTTTACAGAAACAGCTAG	R	G/T
01	1-5	rs12759915	TAACAGCATTCCAGATTTCAG	R	A/G
01	1-6	rs1202593	ACAGATGCAGGCCTGAGTCAT	R	A/G
01	1-7	rs2982742	GTCAGTCCACATCTAGAGTATC	R	A/C
01	1-8	rs576736	CCTCTTCAAATCTTAAGTTGCTAG	R	A/G
01	1-9	rs10864713	AGAGAAGAGGCGCATTTGAG	R	C/T
02	2-1	rs4832461	TCCAAATGGCTCTGGGTCAC	F	A/C
02	2-2	rs1250915	ACAGAGAAGTGGTTTTAGAAGG	F	A/T
02	2-3	rs11892626	CTCCTCGATTCTCTTCTAACAAG	F	G/T
02	2-4	rs7573184	TGGGACTGTTGCATTTGTTTCTT	F	C/G
02	2-5	rs6542461	CATGAAGCATTTTAAGACACTGGA	F	A/G
02	2-6	rs7580941	CATCAATAGGTGTAGCCCAC	F	A/G
03	3-1	rs2649734	CATGCTCCTTGATGTTCTCTCAA	F	C/T
03	3-2	rs6807414	ACAGTAAGGGTTAACACATGCT	F	A/G
03	3-3	rs6793629	TGTTCCTAGGCTTGAAACTAGAA	R	A/G
03	3-4	rs12629514	TCATCAGAAAGCATGCAGAGTTG	F	C/T
03	3-5	rs978979	ACCACTCTAAGACGCATACTTTT	R	A/G
04	4-1	rs1822841	ATCAACCAAATTGTTCTACCACGA	F	C/T
04	4-2	rs7684079	CCACCTGCAAGGGAAGATGT	F	A/C
04	4-3	rs2546275	TCATTAGCTGTTAACAATTCCAG	F	C/T
04	4-4	rs9995245	AGTACATCAAAGCAGGTAGCATA	F	A/G
04	4-5	rs4975214	TGTGGCATCTCTCTCTGGCA	R	C/T
05	5-1	rs6594747	AGGCTTATTTTCTTGCTGCTGA	R	C/T
05	5-2	rs7723568	CGGCAAATGAGACTCGTTCC	F	A/C
05	5-3	rs7444492	CCTCATAACAATAAGGTGACACA	F	C/T
05	5-4	rs4703439	AAGGACCGAGAGGTGATTGA	F	C/T
06	6-1	rs6915280	TCGTGCTGGGTATGTTGCTAAG	F	A/T
06	6-2	rs17559298	TCTCTAATGAGGGTGGCTTG	F	C/T
06	6-3	rs1570281	GCTTCCAGAACAGTACCAGGA	F	A/C
06	6-4	rs3846764	ACTTCATCTTGTAACGAGACTTTG	R	G/T
07	7-1	rs217013	TGGTTGACTGCATTTCTTGGCTT	F	A/G
07	7-2	rs1525830	CTGAGCCAAGCGATCCAAAC	R	C/T
07	7-3	rs7786414	GATCCCAAGACTTTCACCAAAG	R	C/T
08	8-1	rs4105594	TCCCACTTCAAGCCCACAAT	F	A/C
08	8-2	rs1445561	AGGAAGAAGGACTCACACCC	F	A/G
08	8-3	rs9297236	GATTAATAACAGTGCTACCAAAAGTC	F	A/G

Table 4.2 (continued).

Chromosome	In-house SNP code	NCBI ref	SBE Primer[a]	Direction	SNP allele
09	9-1	rs7858174	TTGGGTTCAGCAACTTGGAAGTG	F	C/T
09	9-2	rs10491520	GTTTGTCTGTCTACCAACCTATCT	F	C/G
09	9-3	rs10965215	GTTTTGCAGGACTATTTGCCAC	F	A/G
10	10-1	rs180921	GTGGCAGGCAGTACTTGACCT	F	C/G
10	10-2	rs555325	CACCATTTGTCACCCACTTTCT	F	C/T
10	10-3	rs12764177	ACCTCAGGCAAAGAGCTTAGCT	R	A/C
11	11-1	rs517679	TTGAAATTAGGCACCTGTCCACT	F	C/T
11	11-2	rs2941043	GGTATGAAAGGCCGTGTGAAAAT	R	A/G
12	12-1	rs6487665	TCTCATTCATTGACGTGTTTAGG	F	C/T
12	12-2	rs10777845	ACTTGCCACATACTGCTCGTC	F	C/T
13	13-1	rs4435117	CTAAATCTAGACTGCAGTTT	R	A/G
13	13-2	rs4941487	CTAACATGTTAGCTTCAAGGCTT	R	A/G
13	13-3	rs7338627	TTCAATCACTTGTGCCAGATGT	R	A/C
13	13-4	rs2892545	AGAAGTCATGCTTTCAGTTA	F	C/T
14	14-1	rs17095615	TTGGAAAATCAGTGATCCTCAACTG	R	A/G
14	14-2	rs11628091	GCTTTGATGTCCCGAGTCCA	F	A/C
14	14-3	rs1489870	GTATGGTTTTTCTAAGGAACAGA	F	A/G
14	14-4	rs10133956	CGCCTCCATTGAATTGGCTC	F	C/G
15	15-1	rs4778706	CCCTGTTGCAAAGTAAAAGCCT	F	A/T
15	15-2	rs3848179	CTCCTTTGCTTGGCCTGATAG	R	C/T
15	15-3	rs1529883	ACTCACATTTATCTCATGGTTAGTTAT	R	C/G
16	16-1	rs8057434	AAATGGAGTGTAAACTGCAAACGT	F	C/G
16	16-2	rs7204754	AAGTGTTGTGTTAATTTGGCTCCAT	R	A/T
16	16-3	rs1477389	TAGCTTCTGGGCATGTGACA	F	C/G
17	17-1	rs4925075	CTGGCTGGATGCCCACTTAG	F	A/G
17	17-2	rs2045660	AAGGCAGCAGGAAAAGGCTCA	F	C/T
17	17-3	rs1872236	TTCCTTCTTCAATTTAGGGGTTGA	F	A/C
18	18-1	rs4891524	ATTACAGCATGTTCTCCTGAGCA	F	A/C
18	18-2	rs17064977	AAGTTGGAAGAGGAGCGACTC	F	C/T
18	18-3	rs9950394	ATAAGCTGGCAGGAGAGCAAG	R	A/G
19	19-1	rs10414856	GAAGAGTTCCCCCAAGCAA	F	C/T
19	19-2	rs17304618	TGTGCCTGTGGAGTCACTC	F	A/G
20	20-1	rs6098780	CGAACTGCATTTCACATCACTCT	F	C/G
20	20-2	rs745661	CTCTGTGTTCTCTCTATTCCATC	F	A/T
21	21	rs8130475	GAAAGGTTGGCTAATAGTCAGGT	R	C/T
22	22	rs4820621	CTCTTTCCCTTGCCTTTCCG	F	C/T

[a] SBE primer for SNaPshot™ analysis are listed from 5′ to 3′.

4.3.1.3. Evaluation of SBE Primers

The primers were evaluated to ensure that they produced the expected results and that no artefact peaks, that would interfere with the target peaks, were generated. To achieve this aim, a SNaPshot™ reaction was set up except that, instead of the DNA template, 1 µl of dH_2O was added to the reaction, as described in Section 2.3.14. Certain non-specific peaks were observed in the green dye electropherograms, as in primer codes 13-

1 and 14-1 (Figure 4.4). These peaks could have originated from the addition of ddATP to the 3′ end of none target SBE primers. Since the electrophoretic mobility and their fluorescent dyes of these primers were constant, the non-specific peaks were identified (Figure 4.5).

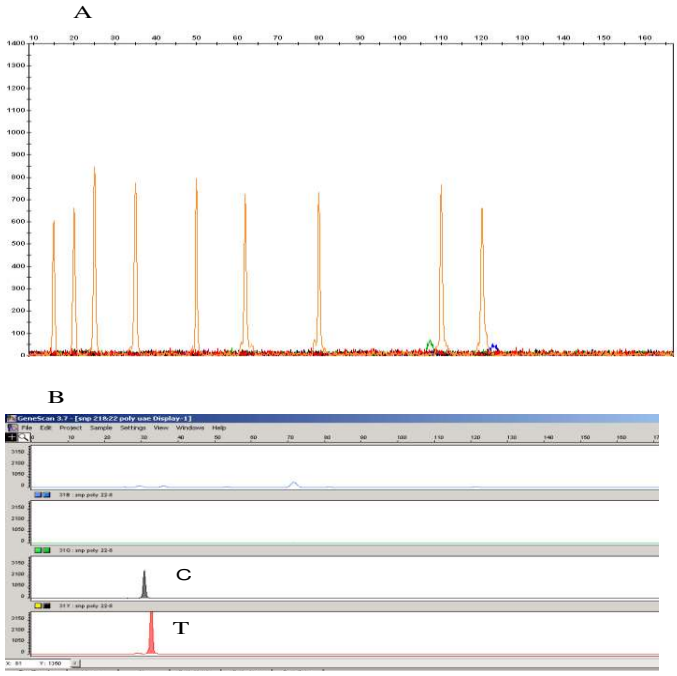


Figure 4.4. Shown above is an example of SBE evaluation. [A] represents an electropherogram of the internal standard Liz -120 without any artefact peaks. [B] represents the SNaPshot™ reaction with the same SBE primer with presence of template amplicon; two clear peaks are produced that represent the expected alleles. The figure is for SNP code 22.

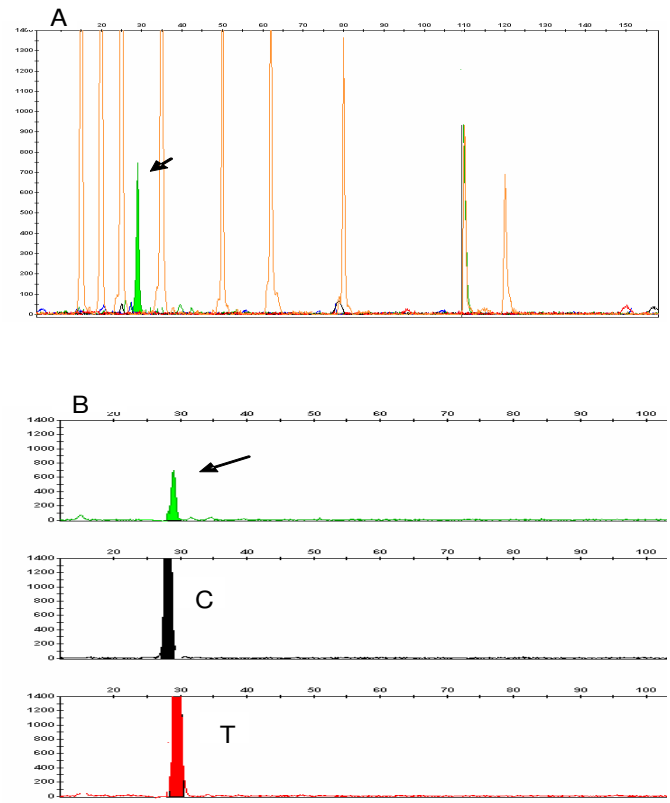
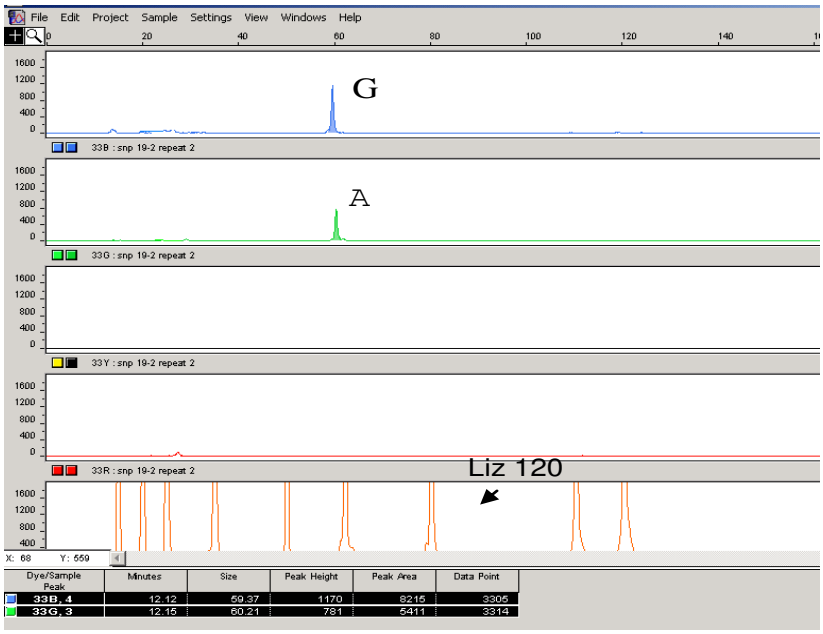


Figure 4.5. Shown above are electropherograms representing SBE primer evaluation for SNP 13-1. [A] represents the SNaPshot™ reaction without the DNA template and the 9 peaks of GeneScan™ LIZ-120 size standard. [B] represents the SNaPshot™ reaction with DNA template and the SNP target CT. The arrows represent the extra peak observed due to the non-target SBE primer peak, which can be differentiated from the true allele peaks.

4.3.1.4. Performance of the SBE Primer Reactions

To evaluate the performance, reproducibility and specificity of the designed SBE primers, the reactions were performed in triplicate (Figure 4.6).

A



B

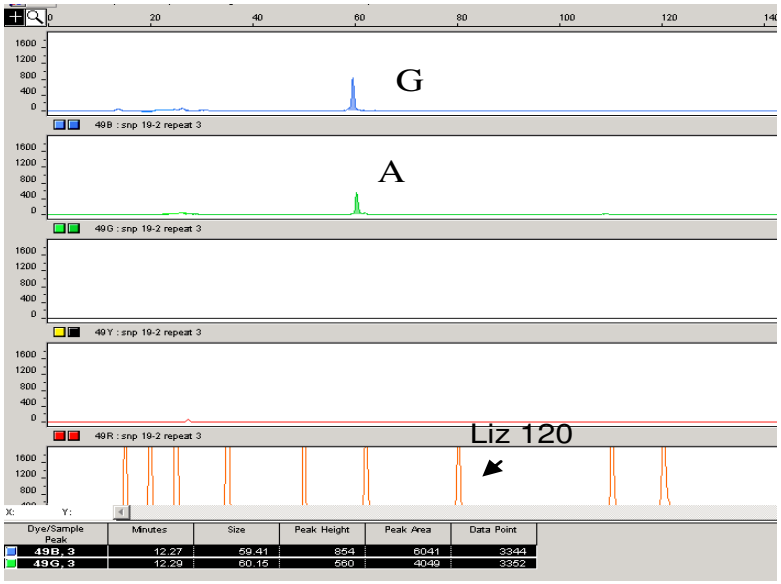


Figure 4.6. Shown above are Electropherogram A and B, which represent repeat 2 and 3 respectively for SNP code 19-1.

Each replicate was compared to the expected size of the SNP. No allele varied by more than 4 bp from the true allele size when the SBE primer was 25 bp or more. Also, each replicate was determined to have the correct genotype: homozygote loci appeared as a single peak and heterozygote loci as two peaks. In relation to the actual SBE primer

size, it was found that the peak signal size from a ddG incorporation showed no significant difference. It was also found that ddA incorporation led to an allele size 1 bp bigger than that of ddG, whilst that of ddT was 2-3 bp higher than ddG. The biggest allele size difference was observed for ddC, which was up to 4 bp higher than that of ddG. These differences in peak signal size were pronounced in primers shorter than 30 bp (Figure 4.7).

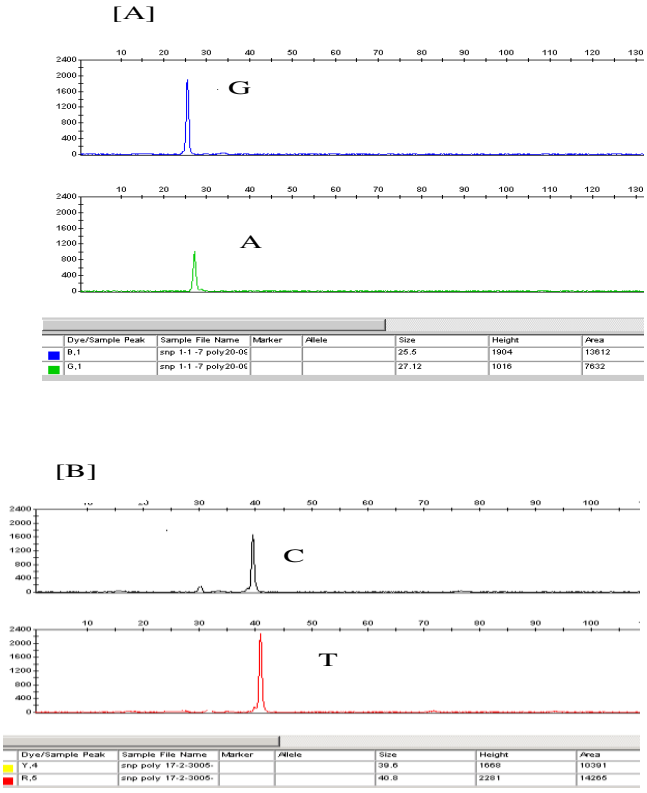


Figure 4.7. Shown above are electrophoretic peaks of SBE primer reaction. [A] represents SNP code1-1 (actual SNP size 21 bp) with heterozygote alleles AG and giving sizes of 26 bp and 27 bp respectively. [B] represents SNP code 17-2 (actual 38 bp) with heterozygote alleles CT with allele sizes of 39 bp and 40 bp respectively.

In addition, the replicates were determined to have a minimum threshold of 100 relative fluorescent units (RFUs) and the peak ratio of heterozygote alleles at each locus was recorded and calculated according to the dye signal effect observed on each of the SNP types. The maximum peak ratio for heterozygote alleles was 4:1 – this is due to the variation in signal strength from the four ddNTPs. All SBE primers were observed to have the correct sizes and genotypes except for the primers 4-5, 7-3, 10-1, 10-3, 16-3 and 20-1, which showed extra peaks that could interfere with the legitimate SNP peak (Figure 4.8). In addition, during the analysis SNPs 6-3, 17-1 and 17-3 were observed to be intronic. These intronic SNPs along with those that produced artefacts were rejected, and the number of candidate SNPs was reduced to 66 (Table 4.3).

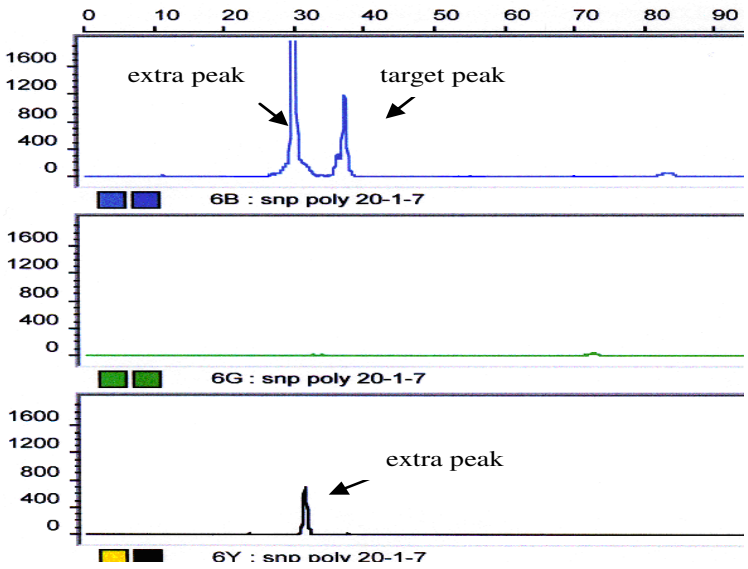


Figure 4.8. Shown above is an example of incorrect genotype observed due to the impurity of the SBE primer. The electropherogram represents primer 20-1 with unrelated heterozygote G/C (blue/green) peaks. The target peak is homozygote GG. This SNP was rejected.

Table 4.3. Shown below are data for the 66 SNPs that produced clear results after SBE: The average size and standard deviation (s.d.) for each triplicate are shown. The highlighted figures indicate slight increases in s.d.

In-house Code	SNP Genotype[a]	Alleles	SNP size	Allele A		Allele B	
				Average	s.d.	Average	s.d.
1-1	AG	GG	21			26.59	0.29
1-2	CT	TT	25			29.80	0.07
1-3	CT	TT	29			34.56	0.09
1-4	GT R	AA	37	40.44	0.25		
1-5	AG R	CC	33	34.39	0.15		
1-6	AG R	CC	41	43.57	0.18		
1-7	AC R	GG	45	46.51	0.21		
1-8	AG R	TT	49	52.01	0.23		
1-9	CT R	AA	53	56.43	0.16		
2-1	AC	CC	28			28.83	0.226
2-2	AT	AT	27	33.47	0.59	33.09	0.60
2-3	GT	GT	28	30.69	0.14	32.90	0.42
2-4	CG	GC	28	32.41	0.28	33.60	0.56
2-5	AG	AG	29	30.71	0.35	32.06	0.44
2-6	AG	GG	29			29.80	0.94
3-1	CT	CT	28	31.61	0.31		
3-2	AG	GG	27			32.02	0.57
3-3	AG R	CC	28	32.02	0.58		
3-4	CT	TT	28			32.35	0.40
3-5	AG R	CT	28	30.87	0.27	32.48	0.31
4-1	GT	TT	29			31.82	0.12
4-2	AC	CC	29			31.53	0.12
4-3	CT	CT	28	30.85	0.266	32.25	1.08
4-4	AG	AG	28	29.25	0.10	30.09	0.09
5-1	CT R	AG	27	31.41	0.06	33.04	0.01
5-2	AC	AC	29	31.74	0.14	31.18	0.12
5-3	CT	TT	28			30.94	0.71
5-4	CT	CT	29	31.76	0.02	33.61	0.04
6-1	AT	AA	27	34.24	0.01		
6-2	CT	CC	29	30.64	0.26		
6-4	GT R	AC	29	33.47	0.02	32.17	0.04
7-1	AG	AG	28	31.82	0.37	33.69	0.05
7-2	CT R	GG	29			29.73	0.26
8-1	AC	AC	29	30.87	0.07	30.10	0.08
8-2	AG	AG	29	29.14	0.14	30.77	0.02
8-3	AG	AG	31	31.72	0.03	32.76	0.02

Table 4.3 (continued).

In-house Code	SNP Genotype[a]	Alleles	SNP size	Allele A		Allele B	
				Average	s.d.	Average	s.d.
9-1	CT	TT	28			33.77	0.01
9-2	CG	CG	29	31.15	0.02	32.17	0.04
9-3	AG	AG	27	29.04	0	30.48	0.05
10-2	CT	TT	27			31.39	0.13
11-1	CT	CC	28	29.63	0.04		
11-2	AGR	CC	28	30.92	0.02		
12-1	CT	TT	28			34.14	0.01
12-2	CT	CC	26	28.33	0.05		
13-1	AG R	CT	25	28.63	0.47	29.79	0.55
13-2	AG R	CT	32	34.52	0.08	35.52	0.03
13-3	AC R	GT	35	37.27	0.13	39.73	0.29
13-4	CT	TT	37	41.10	0.39		
14-1	AG R	CT	46	49.81	0.12	49.44	0.31
14-2	AC	CC	45	47.52	0.20		
14-3	AG	GG	52	54.50	0.89		
14-4	CG	CC	53	54.94	0.19		
15-1	AT	AT	23	27.69	0.89	27.52	0.76
15-2	CT R	GG	26	29.85	0.56		
15-3	CG R	CC	32	35.17	0.13		
16-1	CG	CG	37	38.38	0.22	39.20	0.30
16-2	AT R	AA	30	34.63	0.05		
17-3	AC	AA	42	46.02	0.30		
18-1	AC	AC	46	48.67	0.19	48.25	0.20
18-2	CT	CC	50	51.78	0.07		
18-3	AG R	CT	54	56.01	0.64	56.67	0.08
19-1	CT	CT	58	58.83	0.06	59.49	0.03
19-2	AG	AG	58	59.27	0.07	60.08	0.07
20-2	AT	TT	40	44.22	0.12		
21	CT	AA	28	33.43	0.03		
22	CT	CC	27	30.83	0.45		

[a] The SNP genotypes are arranged in forward sequence as in the NCBI database. R represents the reverse sequence used during SBE primer design.

4.3.2. Multiplexing

For this study to assess the potential for combining the primer sets, 6 loci were selected to represent the developed 66 SNPs markers (Table 4.4). These markers were selected to have different lengths of PCR products, ranging from larger PCR product; 142 bp-147 bp, medium,; 110 bp-119 bp; and small; 90 bp-92 bp. Therefore, two triplex sets were used.

The triplex optimisation and genotyping was performed as described in Sections 2.4.10 and 2.4.11.

Table 4.4. Shown below are the PCR and the SBE primers in the triplex sets with their SNP reference and position.

SNP code	SNP ref	SNP genotype	Position	PCR size (bp)	SBE size (bp)
		Triplex 1			
4-4	rs9995245	A/G	4	90	28
19-2	rs17304618	A/G	19	110	58
13-4	rs2892545	C/T	13	142	37
		Triplex 2			
21	rs8130475	A/G*	21	92	28
18-3	rs9950394	C/T*	18	119	54
17-3	rs1872236	A/C	17	147	42

* Genotypes are for the reverse sequence

The annealing temperature that was designed for singleplex (Section 2.3.8), gave almost the same results for both triplex sets. At 60 °C, DNA bands were observed in agarose gels (2.5% w/v) (Figure 4.9).

However, the concentration of PCR primers varied slightly from the concentration used in the singleplex reaction, ranging from 0.2 to 0.4 µm, with the addition of 1.5 mM $MgCl_2$ used to make the final concentration to 3.0 mM in the amplification reaction (Table 4.5).

SBE optimisation was found to be the same as the SNaPshot™ singleplex condition except that all of the primer concentrations were reduced to 0.2 µm for both triplex sets.

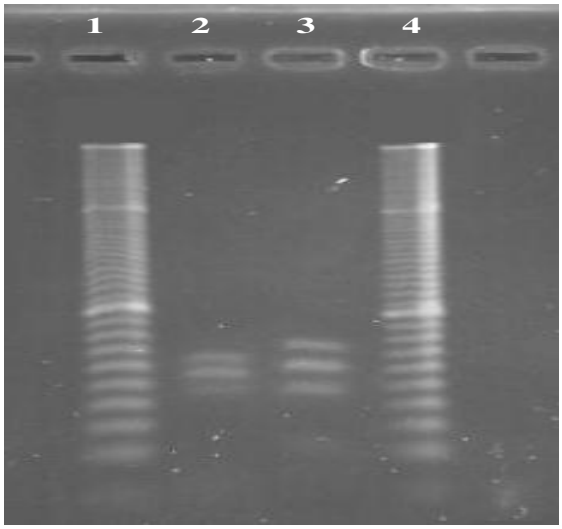


Figure 4.9. Shown above are the results from the optimised triplexes, run on a 2.5% agarose gel. The primer concentration ranged from 0.2 µm to 0.4µm, $MgCl_2$ was 3.0 µm and annealing temperature at 60 °C. Lanes 1 and 4 are for 20 bp ladder; lane 2 represents triplex set 1 and lane 3 represents triplex set 2. The full conditions are shown in Table 4.5.

Table 4.5 Shown below are the optimised primer concentrations (µm) for the PCR triplex sets at an annealing temperature of 60 °C and 3.0 µm of $MgCl_2$.

Triplex 1 SNP Code	NCB ref	PCR primer concentrations (µm)	Triplex 2 SNP Code	NCBI ref	PCR primer concentrations (µm)
4-4	rs9995245 A/G	0.2	21	rs8130475 A/G R	0.2
19-2	rs17304618 A/G	0.4	18-3	rs9950394 C/T R	0.4
13-4	rs2892545 C/T	0.2	17-3	rs1872236 A/C	0.2

The multiplexes were used to assess the effectiveness of SNPs in SNaPshot on real and simulated forensic casework (Chapter 6).

4.3.3. SNaPshot™ vs. Affymetrix® Genotype

A comparison between the Affymetrix® and SNaPshot™ systems was carried out to evaluate the SNP genotype results from each method. One Kuwaiti sample from the ten samples that were used for Affymetrix® screening in Section 2.3 was selected for this study. DNA extraction and purification was performed according to the procedures described in Section 2.2.1.

25 SNP loci from the 22 autosomal chromosomes were selected randomly to represent the 66 SNPs. Chromosomes 1, 2, and 3 contributed two loci each, when one SNP was selected from each of the other chromosomes. SNaPshot™ singleplex reactions were performed. The data were collected and compared with those generated from Affymetrix® screening.

The results obtained from the concordance study between Affymetrix® and SNaPshot™ showed an agreement in all 25 primers. However, the SNP code 22 (rs4820621) showed a deviation with homozygote TT for SNaPshot™ from the Affymetrix® AG (R) heterozygote (Table 4.6). A reassessment of the primer design and the result obtained during the triplicate genotyping of the primers with another different sample showed that the expected results were obtained – heterozygous genotypes were also detected at this locus. This difference could be explained by the sample possibly having a mutation in the forward strand at the SNP rs4820621 site; the Affymetrix® data generated from this sample used the reverse primer. However, the most likely explanation is that this non-concordance is that this datum from the Affymetrix® was incorrect. However, for more confirmation the sample could be sequenced to check my mutation present at the primer site.

Table 4.6. Shown below are SNPs genotypes obtained from concordance study between Affymetrix® and SNaPshot™. [F] represents the forward primer sequence and [R] represent the reverse primer sequence. The highlighted SNP represents the homozygote genotype TT, which deviated from the Affymetrix® result.

SNPcode/ ID	Genotype	Chromosome	Affymetrix®	SNaPshot™
1-2/ rs10864499	C/T	1	CC F	CC F
1-9/ rs10864713	C/T	1	CT F	GA R
2-1/ rs4832461	A/C	2	AC	AC F
2-5/ rs6542461	A/G	2	GA	GA F
3-1/ rs2649734	C/T	3	CT	CT F
3-5/ rs978979	A/G	3	GG	CC R
4-4/ rs9995245	A/G	4	GA	GA F
5-3/ rs7444492	C/T	5	CC	CC F
6-4/ rs3846764	G/T	6	TT	AA R
7-1/ rs217013	A/G	7	CC R	GG F
8-1/ rs4105594	A/C	8	TT R	AA F
9-1/ rs7858174	C/T	9	GG R	CC F
10-2/ rs11259108	C/T	10	GA R	CT F
11-1/ rs 517679	C/T	11	GG R	CC F
12-1/ rs6487665	C/T	12	GA R	CT F
13-1/ rs4435117	A/G	13	AA F	TT R
14-/ rs11628091	A/C	14	AA F	AA F
15-1/ rs4778706	A/T	15	AT F	AT F
16-1/ rs8057434	C/G	16	CC R	GG F
17-3/ rs1872236	A/C	17	CC F	CC F
18-/ rs17064977	C/T	18	CT F	CT F
19-/ rs17304618	A/G	19	AG F	AG F
20-2/ rs745661	A/T	20	AA F	AA F
21/ rs8130475	C/T	21	CT F	A/G R
22/ rs4820621	CT	22	AG R	TT F

4.4. Discussion

The potential of SNPs as a forensic tool has been widely acknowledged over the last few years. In this respect, the most attractive feature of SNPs is their short amplicon and suitability for degraded DNA detection (Inagaki et al., 2004; Budowle, 2004).

SNP Identification

This chapter demonstrated that a careful selection from a genome screen (autosomal) identified candidate SNPs that could later be validated for forensic applications.

Careful primer design, with annealing temperatures of 60 °C ± 3 °C enabled them to all be efficiently amplified at 60 °C. A uniform annealing temperature minimised the number of thermal cycler parameters, which in turn saved time and reduced variations as all reactions were carried out in the same thermal cycler under the same conditions. Moreover, equal annealing temperatures for PCR primers is an advantage when producing a multiplex system, as will be described below.

In order to achieve correct SNP genotyping, an assessment of parameters, such as the amount of sample to be used in both the PCR and SNaPshot™ reactions, the SNP type, the SNP length and the presence of any ambiguous peaks, was required at the start of the SNP development. If the amount of sample is very high, or very low and unrelated peaks are present, then, collectively, this can lead to drop-in/drop-out alleles and unrelated SNP peaks. In turn, this can lead to a misinterpretation of the results, especially when handling samples such as those that are degraded or those of low concentrations. Moreover, false results can affect the statistical parameters that will be applied later for SNP characterisations, such as allele heterozygosity. For this, each SNP locus was assessed through triplicate analysis. This allowed the selection of 66 SNP

candidates. Additionally, the assessment for SNP genotyping was carried out in the presence of negative and positive controls (Applied Biosystems). This allowed for any ambiguous results relating to the reaction set up to be eliminated. However, during the assessment, it was found that one of the PCR primers (reverse) of SNP code 19-1 was within the region of another non target SNP at -45 bp of the target SNP. To remedy this, careful design of the SBE primer for that specific position was undertaken. This was confirmed by a successful result obtained from the assessment of the SNaPshot™ reaction.

Multiplexing

The objective of this research was to identify SNP candidates that could be useful for forensic application and that might in future be multiplexed. Therefore, formation of large multiplex such as that developed by Sanchez et al. (2006) is essential for typing forensic casework. However, in this study only a few SNPs were multiplexed to assess the potential for combining the primer sets. The careful optimisation of both PCR and SBE primers helped in the development of the triplex sets without significant complications. All the PCR primers in the triplexes produced acceptable results at an annealing temperature of 60 °C. Six SNP loci in two triplexes were chosen for further SNPs assessment (Chapter 6). The SNapShot technique allows up to approximately 25 loci to be analysed in a single reaction, however, the results generated from such a large set loci are often difficult to interpret. Development of such a large multiplex was not attempted as part of this project.

Concordance Study

The results obtained from the concordance study between Affymetrix® and SNaPshot™ genotyping provided an additional assessment of the selected SNPs. The 25 SNPs genotyped using SNaPshot™ showed full concordant with the Affymetrix® results except at locus 22 when a homozygote for TT allele was observed. The most likely cause of this non-concordance is that the result from the Affymetrix® was incorrect. In the context of this study the non-concordance is not a problem, as long as it has not led to the selection of monomorphic SNP loci. It would only be problematic if the genotype data from forensic samples analysed in different laboratories did not produce the same results.

Comparison between SNaPshot™ and other SNP Genotyping Methods

There are various SNP genotyping application in genetic field such as TaqMan® SNP Genotyping Assays, SNPlex™ Genotyping System (Vega et al., 2005), GenPlex SNP Genotyping System (Musgrave-Brown et al., 2008) and Affymetrix® GeneChip® Technique (Matsuzaki et al., 2004). These applications vary in cost, number of SNPs can be detected and DNA sample quantity. TaqMan® SNP Genotyping Assays require a single enzymatic step and a large number of validated off-the-shelf assays that make the application simple and low of cost. The assay for the TaqMan® SNP Genotyping is however limited to the detection of 2 SNPs (Vega et al., 2005). In forensic casework this would require the setting up of 30 to 40 separate assays in order to analyse the required number of SNPs, and in many forensic cases there would be

insufficient DNA in the sample. The SNPlex™ and Genplex Genotyping Systems are highly automated, and designed for high throughput SNP application. The GenPlex system is a modification of the SNPlex™ system (Phillips et al., 2007). The SNPlex™ system begins with a multiplex oligo- ligation assay (OLA) reaction that is followed by PCR reaction of the ligation products. The GenPlex system begins with PCR amplification of the template DNA followed by an OLA reaction. The assays require special instruments for the SNPs detection such as 3130 or 3730 DNA Genetic Analyser which are not available in all forensic labs (Vega et al., 2005).

Affymetrix® GeneChip® Technique is also designed for high throughput application, but the assay requires special instruments for the detection and also a large amount of sample is required. This type of technique can be useful for screening purposes such as in clinical tests or for association studies. The same applies to other similar platforms provided by Illumina. All these high throughput methods require large amounts of DNA, which are not commonly found when analysing forensic samples.

4.5. Conclusion

SNaPshot™ Genotyping Assay in comparison to the above assays is robust and convenient as it can be performed in simple instrument (310 Genetic Analyser) that can be available in most forensic labs. The assay is sensitive with 0.5 ng/µl of sample detected, and suitable for high throughput application (Sanchez et al., 2006). The limitation of this technique was found to be in the dyes that are associated with the ddNTPs. Future study is needed to overcome the influence of the dyes on the detected SNPs.

CHAPTER 5

CHARACTERISATION

of SNPs

5.1. Overview

When introducing any new marker for forensic applications, it is a prerequisite to assess the marker's utility by testing parameters associated with that marker. Accordingly, the SNP candidates that were identified in Chapter 4 were analysed for such parameters, including: allele frequency, heterozyosity, match probability and discrimination power.

In addition, forensic samples are often limited in quantity and typing the low amounts of these samples can cause incomplete DNA profiling or failure altogether. Low levels of DNA template can increase the stochastic effects of PCR (Krenke et al., 2002), resulting in heterozygote allele imbalance and also allele dropout. This can greatly be influenced the successful profiling of DNA. Therefore the performances of the selected SNPs were assessed using low-levels of DNA.

5.2. Aims of this Chapter

The objectives of this chapter are:

- To generate allele frequencies using UAE individuals for the 66 SNPs detected in Chapter 4.
- To determine the threshold sensitivity of the SNPs to generate full DNA profiles.

5.3. Generation of Allele Frequencies

5.3.1. Samples

Dried blood samples on FTA card® from 100 UAE individuals were used. The samples were collected by the Dubai Police Crime Laboratory, which were received with informed consent, and were anonymised upon receipt (Section 2.1).

5.3.2. DNA Extraction and Quantification

DNA extraction was carried out using organic extraction and was followed by phenol chloroform purification. These procedures were carried out as described in Chapter 2 (Sections: 2.2.1.1, 2.2.1.2 and 2.4.2).

The estimation of DNA concentration was determined using the Quantifiler™ Human DNA Quantification Kit (Applied Biosystems) and the ABI 7500 real time PCR (Applied Biosystems). These procedures were performed as described in Section 2.2.2.1.

DNA concentration were found to range between 27 ng /µl and 0.39 ng/µl. Based on the results obtained from DNA quantification, 25 samples with DNA concentrations greater than 3 ng/µl (in a total volume of 20 µl) were selected to represent UAE individuals for the study of allele frequency see (Appendix A1).

5.3.2.1. Amplification and Genotyping of SNPs

In order to generate allele frequencies for UAE individuals, the 66 SNPs that were identefied in Chapter 4 were genotyped. Each of the 25 UAE samples were tested using the 66 SNPs in singleplex reactions, resulting in 1650 singleplex SNP amplifications

and singleplex SNaPshot reactions, performed using the 66 PCR primer pairs and 66 SBE primers respectively. Each SNP amplification was carried out using 0.5 ng/µl of DNA sample.

5.4. Results

5.4.1. Statistical Analyses

5.4.1.1. Alleles Frequencies Distribution

Since the selected SNPs are biallelic markers, a smaller number of samples are required to provide an accurate allele frequency compared to other markers such as STRs (Vaarno et al., 2004; Sanchez et al., 2006). Therefore, 25 samples from the UAE population were used to determine the allele frequencies of the SNPs. A total of 3300 alleles were observed for the 66 loci.

The results have shown that the 66 SNP loci were polymorphic with minimum observed heterozygosity of 20 % and a minimum allele frequency of 0.14 (Table 5.1).

Table 5.1. Shown below are the allele frequencies observed for each of the 66 SNP loci for 25 UAE individuals listed with their genotypes.

In-house Code	Alleles (1, 2)	Frequency of Allele 1	Frequency of Allele 2	In-house Code	Alleles (1, 2)	Frequency of Allele 1	Frequency of Allele 2
1-1	A, G	0.32	0.68	8-1	A, C	0.62	0.38
1-2	C, T	0.46	0.54	8-2	A, G	0.66	0.34
1-3	C, T	0.14	0.86	8-3	A, G	0.38	0.62
1-4	C, A	0.36	0.64	9-1	C, T	0.4	0.6
1-5	T, C	0.5	0.5	9-2	C, G	0.38	0.62
1-6	T, C	0.3	0.7	9-3	A, G	0.5	0.5
1-7	T, G	0.28	0.72	10-2	C, T	0.4	0.6
1-8	T, C	0.64	0.36	11-1	C, T	0.68	0.32
1-9	G, A	0.6	0.4	11-2	T, C	0.5	0.5
2-1	A, C	0.44	0.56	12-1	C, T	0.4	0.6
2-2	A, T	0.3	0.7	12-2	C, T	0.7	0.3
2-3	G, T	0.66	0.34	13-1	T, C	0.46	0.54
2-4	C, G	0.32	0.68	13-2	T, C	0.42	0.58
2-5	A, G	0.6	0.4	13-3	T, G	0.24	0.76
2-6	A, G	0.64	0.36	13-4	C, T	0.52	0.48
3-1	C, T	0.32	0.68	14-1	T, C	0.4	0.6
3-2	A, G	0.22	0.78	14-2	A, C	0.34	0.66
3-3	T, C	0.52	0.48	14-3	A, G	0.46	0.54
3-4	C, T	0.36	0.64	14-4	C, G	0.52	0.48
3-5	T, C	0.28	0.72	15-1	A, T	0.34	0.66
4-1	G, T	0.46	0.54	15-2	G, A	0.66	0.34
4-2	A, C	0.42	0.58	15-3	G, C	0.32	0.68
4-3	C, T	0.42	0.58	16-1	G, C	0.56	0.44
4-4	A, G	0.66	0.34	16-2	A, T	0.24	0.76
5-1	G, A	0.52	0.48	17-3	A, C	0.32	0.68
5-2	A, C	0.4	0.6	18-1	A, C	0.5	0.5
5-3	C, T	0.48	0.52	18-2	C, T	0.28	0.72
5-4	C, T	0.48	0.52	18-3	T, C	0.36	0.64
6-1	A, T	0.76	0.24	19-1	C, T	0.54	0.46
6-2	C, T	0.78	0.22	19-2	A, G	0.14	0.86
6-4	C, A	0.36	0.64	20-2	A, T	0.24	0.76
7-1	A, G	0.3	0.7	21	G, A	0.38	0.62
7-2	G, A	0.6	0.4	22	C, T	0.72	0.28

5.4.1.2. Hardy-Weinberg Equilibrium (HWE)

Observed heterozygosity within the population was measured to indicate departure from HWE expectation; the test was applied using the Markov chain method with 10000 permutations (Arlequin v. 3.1). Three of these SNPs showed significant departure ($p = 0.043$, $p = 0.014$ and $p = 0.011$) from HWE at p values < 0.05, as shown in (Table 5.2).

This deviation was expected (5%) as a result of multiple tests (1000 dememorization steps), which yield significant levels of false results (Rice, 1989). The Bonferroni correction at p>0.0008 (0.05 divided number of loci (66)) was applied to correct the results. After employing the Bonferroni correction, these observations were not significant. This indicates that the observed heterozygosity in all 66 loci is in equilibrium with HW heterozygosity expectation.

5.4.1.3. Linkage Disequilibrium

The loci data were tested for genotypic disequilibrium using the pairwise test with p values < 0.05. A total of 10100 pair wise comparisons for all loci were performed to check any correlation between alleles at any of the pairwise comparisons of the 66 loci using Arlequin v. 3.1 software. Most of the loci in the data behaved as expected with no linkage disequilibrium. However, 4 loci on different chromosomes: (7-2, 13-4), (11-2, 15-1) and (14-4, 15-1) were observed to be significant at p < 0.05 (0.00000), some departure was expected as this occurs by chance (Gill et al., 2003; Kidd et al., 2006). However, as the number of loci affected was small, and within the levels expected for such a large number of loci, the affected loci were not rejected based on these results.

Table 5.2. Shown below are the observed (Obs.) and expected (Exp.) heterozygosities for the 66 SNPs typed in 25 individuals. The highlighted numbers show significant deviation from HWE at $p < 0.05$.

No.	In-house Code	Obs. Het	Exp. Het	P-value	s.d.
1	1-1	0.560	0.444	0.366	0.005
2	1-2	0.360	0.507	0.216	0.004
3	1-3	0.200	0.246	0.378	0.005
4	1-4	0.480	0.47	1	0
5	1-5	0.680	0.497	0.098	0.003
6	1-6	0.440	0.458	1	0
7	1-7	0.560	0.411	0.13	0.003
8	1-8	0.640	0.47	0.08	0.003
9	1-9	0.520	0.429	0.378	0.004
10	2-1	0.720	0.503	0.043	0.002
11	2-2	0.600	0.429	0.063	0.002
12	2-3	0.600	0.458	0.178	0.004
13	2-4	0.320	0.444	0.198	0.004
14	2-5	0.560	0.49	0.665	0.005
15	2-6	0.480	0.47	1	0
16	3-1	0.480	0.444	1	0
17	3-2	0.400	0.327	0.551	0.005
18	3-3	0.480	0.509	1	0
19	3-4	0.640	0.47	0.091	0.003
20	3-5	0.261	0.433	0.124	0.004
21	4-1	0.360	0.507	0.218	0.004
22	4-2	0.520	0.497	1	0
23	4-3	0.440	0.497	0.689	0.005
24	4-4	0.360	0.458	0.389	0.005
25	5-1	0.400	0.509	0.416	0.006
26	5-2	0.4	0.49	0.432	0.004
27	5-3	0.56	0.509	0.702	0.005
28	5-4	0.48	0.509	1	0
29	6-1	0.4	0.372	1	0
30	6-2	0.44	0.35	0.313	0.005
31	6-4	0.56	0.47	0.401	0.005
32	7-1	0.6	0.429	0.062	0.003
33	7-2	0.4	0.49	0.418	0.005

Table 5.2 (continued)

No.	In-house Code	Obs. Het.	Exp. Het.	P-value	s.d.
34	8-1	0.44	0.481	0.69	0.005
35	8-2	0.52	0.458	0.659	0.005
36	8-3	0.44	0.481	0.695	0.005
37	9-1	0.56	0.49	0.671	0.004
38	9-2	0.36	0.481	0.225	0.004
39	9-3	0.44	0.51	0.69	0.005
40	10-2	0.4	0.49	0.425	0.005
41	11-1	0.48	0.444	1	0
42	11-2	0.6	0.51	0.442	0.005
43	12-1	0.5	0.496	1	0
44	12-2	0.36	0.429	0.636	0.005
45	13-1	0.44	0.51	0.684	0.004
46	13-2	0.68	0.497	0.087	0.003
47	13-3	0.48	0.372	0.267	0.004
48	13-4	0.24	0.509	0.014	0.001
49	14-1	0.52	0.481	1	0
50	14-2	0.64	0.47	0.095	0.003
51	14-3	0.44	0.507	0.684	0.004
52	14-4	0.48	0.509	1	0
53	15-1	0.52	0.458	0.665	0.005
54	15-2	0.52	0.458	0.662	0.004
55	15-3	0.32	0.372	0.585	0.005
56	16-1	0.44	0.497	0.688	0.005
57	16-2	0.24	0.372	0.102	0.003
58	17-3	0.4	0.444	0.655	0.005
59	18-1	0.68	0.51	0.122	0.003
60	18-2	0.4	0.411	1	0
61	18-3	0.48	0.47	1	0
62	19-1	0.44	0.481	0.698	0.005
63	19-2	0.28	0.301	1	0
64	20-2	0.4	0.372	1	0
65	21	0.52	0.481	1	0
66	22	0.24	0.411	0.055	0.002

5.4.2. Forensic Statistics

The 66 SNPs were analysed in order to assess the utility of the SNPs for forensic application. The PowerStats V.12 program was used to test the classical forensic parameters: power of discrimination and match probability. The tests were carried out independently for each locus.

The selected SNPs possessed an average observed heterozygosity of 0.47. The probability that two individuals would have the same genotype profile (match probability) was found to be 3.058×10^{-25}. Whilst the probability that two individuals are different (a combined power of discrimination) was found to be 0.999999999 (99.9999999%) with a combined power of exclusion of 99.9999999% (Table 5.3). This indicated that the SNPs could be useful for forensic samples identification.

5.4.3. SNPs Performance Evaluation

5.4.3.1. Sensitivity Study

Four SNPs from loci on different chromosomes were selected to represent the 66 SNP markers. To ensure all genotypes are present in the study, the SNPs were selected to exhibited the 4 possible genotypes (G, A, C, and T) (Table 5.4).

In this assessment, two template samples from different individuals were included; the procedure was carried out as described in Section 2.4.5. The basis for selecting more than one sample was to achieve better assessment and analysis of the results obtained from the samples. Moreover, the use of two samples would increase the number of SNP genotypes that lead to more variation in the generated data. The major concern during analysis of the genotypes was the effect on heterozygote loci peak height that were obtained in different dilutions.

Table 5.3. Shown below are the final 66 SNP locus selected from the autosomal chromosomes according to their forensic parameters. The results were obtained using

PowerStats software. Hom; represent homozygosity, Het; represents heterozygosity.

In-house Code	Match Probability	Power of Discrimination	Power of Exclusion	Frequency of Allele A	Hom.	Het.
1-1	0.3376	0.662	0.091	0.32	0.64	0.56
1-2	0.3376	0.662	0.091	0.46	0.64	0.36
1-3	0.6192	0.381	0.030	0.14	0.8	0.2
1-4	0.4048	0.595	0.171	0.36	0.52	0.48
1-5	0.5264	0.474	0.398	0.5	0.32	0.68
1-6	0.4016	0.598	0.142	0.3	0.56	0.44
1-7	0.5072	0.493	0.246	0.28	0.44	0.56
1-8	0.5136	0.486	0.342	0.64	0.36	0.64
1-9	0.4656	0.534	0.206	0.6	0.48	0.52
2-1	0.565	0.435	0.460	0.44	0.28	0.72
2-2	0.52	0.48	0.291	0.3	0.4	0.6
2-3	0.4912	0.509	0.291	0.66	0.4	0.6
2-4	0.3984	0.340	0.072	0.32	0.68	0.32
2-5	0.4304	0.57	0.246	0.6	0.44	0.56
2-6	0.4048	0.595	0.171	0.64	0.52	0.48
3-1	0.4304	0.57	0.171	0.32	0.52	0.48
3-2	0.5392	0.493	0.0.91	0.22	0.64	0.36
3-3	0.3664	0.634	0.171	0.52	0.52	0.48
3-4	0.5136	0.486	0.342	0.36	0.36	0.64
3-5	0.4177	0.582	0.049	0.28	0.74	0.26
4-1	0.3376	0.662	0.092	0.46	0.64	0.36
4-2	0.3984	0.602	0.206	0.42	0.48	0.52
4-3	0.3632	0.637	0.140	0.42	0.56	0.44
4-4	0.3856	0.614	0.091	0.66	0.64	0.36
5-1	0.3408	0.659	0.114	0.52	0.6	0.4
5-2	0.36	0.640	0.114	0.4	0.6	0.4
5-3	0.3856	0.557	0.206	0.48	0.48	0.52
5-4	0.3664	0.634	0.171	0.48	0.52	0.48
6-1	0.4752	0.525	0.114	0.72	0.6	0.4
6-2	0.5072	0.493	0.140	0.78	0.56	0.44
6-4	0.4496	0.55	0.246	0.36	0.44	0.56
7-1	0.52	0.48	0.291	0.3	0.4	0.6
7-2	0.36	0.64	0.114	0.6	0.6	0.4
8-1	0.3792	0.621	0.140	0.62	0.56	0.44
8-2	0.4368	0.563	0.206	0.66	0.48	0.52
8-3	0.3792	0.621	0.140	0.38	0.56	0.44
9-1	0.4304	0.559	0.246	0.4	0.44	0.56
9-2	0.3632	0.621	0.091	0.38	0.64	0.36
9-3	0.3504	0.65	0.140	0.5	0.56	0.44
10-2	0.36	0.669	0.114	0.4	0.6	0.4
11-1	0.4304	0.57	0.171	0.68	0.52	0.48
11-2	0.44	0.56	0.291	0.5	0.4	0.6

Table 5.3 (continued)

	Match	Power of	Power of	Frequency	Hom.	Het.

In-house Code	Probability	Discrimination	Exclusion	of Allele A		
12-1	0.389	0.611	0.188	0.4	0.5	0.5
12-2	0.414	0.586	0.091	0.7	0.64	0.36
13-1	0.350	0.650	0.140	0.46	0.56	0.44
13-2	0.526	0.474	0.398	0.42	0.32	0.68
13-3	0.501	0.499	0.170	0.24	0.52	0.48
13-4	0.347	0.653	0.042	0.52	0.76	0.24
14-1	0.414	0.586	0.206	0.4	0.48	0.52
14-2	0.514	0.486	0.342	0.34	0.36	0.64
14-3	0.354	0.646	0.140	0.46	0.56	0.44
14-4	0.366	0.634	0.171	0.52	0.52	0.48
15-1	0.437	0.563	0.206	0.34	0.48	0.52
15-2	0.437	0.563	0.206	0.66	0.48	0.52
15-3	0.469	0.531	0.072	0.32	0.68	0.32
16-1	0.363	0.637	0.140	0.56	0.56	0.44
16-2	0.482	0.518	0.042	0.24	0.76	0.24
17-3	0.405	0.595	0.114	0.32	0.6	0.4
18-1	0.514	0.486	0.390	0.5	0.32	0.68
18-2	0.437	0.563	0.114	0.28	0.6	0.4
18-3	0.405	0.595	0.171	0.36	0.52	0.48
19-1	0.379	0.521	0.140	0.54	0.56	0.44
19-2	0.542	0.458	0.056	0.14	0.72	0.28
20	0.475	0.525	0.114	0.24	0.6	0.4
21	0.414	0.586	0.206	0.38	0.48	0.52
22	0.443	0.557	0.042	0.72	0.76	0.24
Total	3.05794E-25	>99.9999999%	99.9999999%		0.54	0.47

Table 5.4. Shown below are the chromosome, SNP type and PCR length for each of the 4 SNP loci used in the sensitivity study.

In-house Code	SNP ref	Chromosome	SNP genotype	PCR length (bp)
4-2	rs7684079	4	A/C	130
12-1	rs6487665	12	C/T	119
17-3	rs1872236	17	A/C	147
19-2	rs17304618	19	A/G	110

The genotypes and the RFU values for each homozygote and heterozygote peaks in each of the 9 dilutions were observed and assessed. Each replicate was checked for the correct SNP and the genotypes were noted as partial profiles (pp) when one allele

dropped below the 100 RFU threshold. Normalised RFU was calculated for all alleles; the homozygote signals were divided into two (Table 5.5 to Table 5.9).

Table 5.5 Shown below are the RFUs generated from different DNA dilution for individual 1.Each SNP locus was tested in triplicate and the results are before normalisation of RFUs. [pp] represents partial profile.

DNA concentrations (pg)	100	200	300	400	500	1000	2000	4000	8000	
SNP locus										Genotype
12-1 /CT	358	1150	515	496	798	1421	1362	2913	4542	TT
(119 bp)	598	897	515	924	605	1836	1172	2500	2763	TT
	770	1429	579	668	1236	1550	1919	2858	4178	TT
17-3 A/C	533	1144	2022	2979	1517	6398	7435	7462	7358	AA
(147 bp)	445	1252	2819	2997	1122	5110	7366	7280	7328	AA
	590	1447	1035	2605	1242	6106	7278	7139	7347	AA
19-2 A/G	214	293	1025	664	682	1929	3116	5418	7138	A
	pp	pp	546	370	333	635	950	1729	3236	G
(110 bp)										
	188	263	826	398	560	1655	3023	6178	7154	A
	pp	pp	279	385	275	563	906	1952	3982	G
	182	285	935	470	695	1902	3192	6597	6741	A
	pp	100	165	283	335	638	955	2137	2131	G
4-2 /AC	456	647	486	1000	2392	4590	7173	7159	7179	CC
(130 bp)	516	701	1129	1007	2377	4753	7415	7352	7319	CC
	541	731	1171	1460	2409	3992	6855	7283	7369	CC

Table 5.6 Shown below are the normalised RFUs generated from different DNA dilution for individual 1. [pp] represents partial profile.

DNA concentration (pg)	100	200	300	400	500	1000	2000	4000	8000
SNP									
12-1C/T	179	575	257.5	248	384	460	681	1456.5	2271
	179	575	257.5	248	384	460	681	1456.5	2271
	299	448.5	257.5	312	302.5	368.5	586	1250	1381.5
	299	448.5	257.5	312	302.5	368.5	586	1250	1381.5
	385	714.5	289.5	334	618	461	959.5	1429	2089
	385	714.5	289.5	334	618	461	959.5	1429	2089
17-3A/C	266.5	572	1011	989.5	758.5	3199	3717.5	3731	3679
	266.5	572	1011	989.5	758.5	3199	3717.5	3731	3679
	222.5	626	1409.5	989.5	561	2555	3683	3640	3664
	222.5	626	1409.5	989.5	561	2555	3683	3640	3664
	295	723.5	517.5	802.5	621	3053	3639	3569.5	3673.5
	295	723.5	517.5	802.5	621	3053	3639	3569.5	3673.5
19-2A/G	214	293	1025	3262	682	1929	3116	5418	7138
	pp	pp	546	1356	333	635	950	1729	3236
	188	263	826	2886	560	1655	3023	6178	7154
	pp	pp	279	2167	275	563	906	1952	3982
	182	285	935	3560	695	1902	3192	6597	6741
	pp	100	165	1133	335	638	955	2137	2131
4-2A/C	228	323.5	243	500	1196	2295	3586.5	3579.5	3589.5
	228	323.5	1816	1846.5	1196	2295	3586.5	3579.5	3589.5
	258	350.5	564.5	503.5	1188.5	2376.5	3707.5	3676	3659.5
	258	350.5	1453	1495.5	1188.5	2376.5	3707.5	3676	3659.5
	270.5	365.5	585.5	730	1204.5	1996	3427.5	3641.5	3684.5
	270.5	365.5	1659	2407.5	1204.5	1996	3427.5	3641.5	3684.5

Table 5.7 Shown below are the RFUs generated from different DNA dilution for individual 2. The results are before normalisation of RFUs. [pp] represents partial profile.

DNA concentrations (pg)	100	200	300	400	500	1000	2000	4000	8000	
SNP										Genotype
12-1 /CT	pp	118	351	357	507	438	421	887	3160	C
(119 bp)	242	240	462	1176	1076	1187	1050	2071	7413	T
	pp	112	180	430	348	277	471	1682	2697	C
	242	351	357	1016	771	574	1177	4574	7043	T
	pp	101	206	281	269	214	557	942	1416	C
	102	240	358	685	687	742	950	2418	3718	T
17-3 A/C	2719	4051	1007	1788	7490	7372	7380	7005	6960	A
(147 bp)	260	439	374	467	2536	3994	5900	6456	6228	C
	2779	4680	1197	1720	7244	7193	7155	7000	6896	A
	166	742	220	784	2596	3997	4500	6524	6261	C
	3591	4382	1427	2009	7339	7201	7334	7116	6883	A
	415	809	416	870	4225	3708	5506	6540	6010	C
19-2 A/G	350	411	776	1173	1441	5094	7353	7245	6880	G
(110 bp)	324	127	598	1553	974	2082	6701	7071	7105	A
	106	837	882	789	3156	2244	7298	7187	7189	G
	164	499	801	1305	2640	1551	7284	7099	6963	A
	701	1083	660	820	3093	4240	7281	7174	7059	G
	185	486	570	1104	1260	2903	7258	7115	6817	A
4-2 /AC	439	664	2454	2505	3850	3077	7633	5179	7396	A
(130 bp)	272	514	1516	1293	2267	1978	4400	2695	3735	C
	417	1189	1804	2390	2814	6215	7596	7593	7433	A
	812	486	915	1508	1695	2431	4870	5142	7202	C
	333	663	2334	2791	3299	5427	7635	7551	7519	A
	391	651	932	1537	1193	3029	4487	6459	7193	C

Table 5.8 Shown below are the normalised RFUs generated from different DNA dilution for individual 2. [pp] represents partial profile.

DNA concentrations (pg)	100	200	300	400	500	1000	2000	4000	8000
SNP 12-1 /CT	pp	118	351	357	507	438	421	887	3160
	242	240	462	1756	1076	1187	1050	2071	7413
	pp	112	180	430	348	277	471	1682	2697
	242	351	357	1016	771	574	1177	4574	7043
	pp	101	206	281	269	214	557	942	1416
	102	240	358	685	687	742	950	2418	3718
17-3 A/C	2719	4051	3552	3517	7490	7372	7380	7005	6960
	260	439	1808	1578	2536	3994	5900	6456	6228
	2779	4680	3041	3404	7244	7193	7155	7000	6896
	166	742	1290	1754	2596	3997	4500	6524	6261
	3591	4382	2517	3324	7339	7201	7334	7116	6883
	415	809	2884	2235	4225	3708	5506	6540	6010
19-2 A/G	350	411	3792	3742	1441	5094	7353	7245	6880
	324	127	2820	2994	974	2082	6701	7071	7105
	106	837	3741	2047	3156	2244	7298	7187	7189
	164	499	3835	2064	2640	1551	7284	7099	6963
	701	1083	1301	2430	3093	4240	7281	7174	7059
	185	486	1382	3158	1260	2903	7258	7115	6817
4-2 /AC	439	664	2454	2505	3850	3077	7633	5179	7396
	272	514	1516	1293	2267	1978	4400	2695	3735
	417	1189	1804	2390	2814	6215	7596	7593	7433
	812	486	915	1508	1695	2431	4870	5142	7202
	333	663	2334	2791	3299	5427	7635	7551	7519
	391	651	932	1537	1193	3029	4487	6459	7193

In this study, all 4 SNPs produced profiles and gave reproducible results for most of the concentrations analysed (Figure 5.1). However, for samples containing 100 pg and 200 pg of template, some expected heterozygote loci were observed as homozygotes because one allele either dropped out or was below the threshold, resulting in a partial profile. In those templates with higher concentrations, such as 4000 pg and 8000 pg, some unrelated peaks from the background were observed. The more balanced peaks and full genotypes were obtained with 300 pg to 2000 pg of template. In general, the lowest RFU in both individuals was observed to be for SNP code 12-1 genotype CT. This observation may be due to the influence of dyes in this locus. For individual 1, the locus 19-2 genotype AG exhibited a partial profile, the allele G dropped below the threshold (RFUs 100) in the dilutions: 100 pg and 200 pg. Whilst individual 2 exhibited a partial profile genotype in the 100 pg dilution at locus 12-1 CT, the allele C dropped below the threshold. The genotype A, in loci 17-3, 19-2 and 4-2, showed profiles for both individuals in all the dilutions for both the homozygote and heterozygote loci. The other dyes all displayed some drop out. The different relative fluorescence of the dyes is a limitation of this methodology.

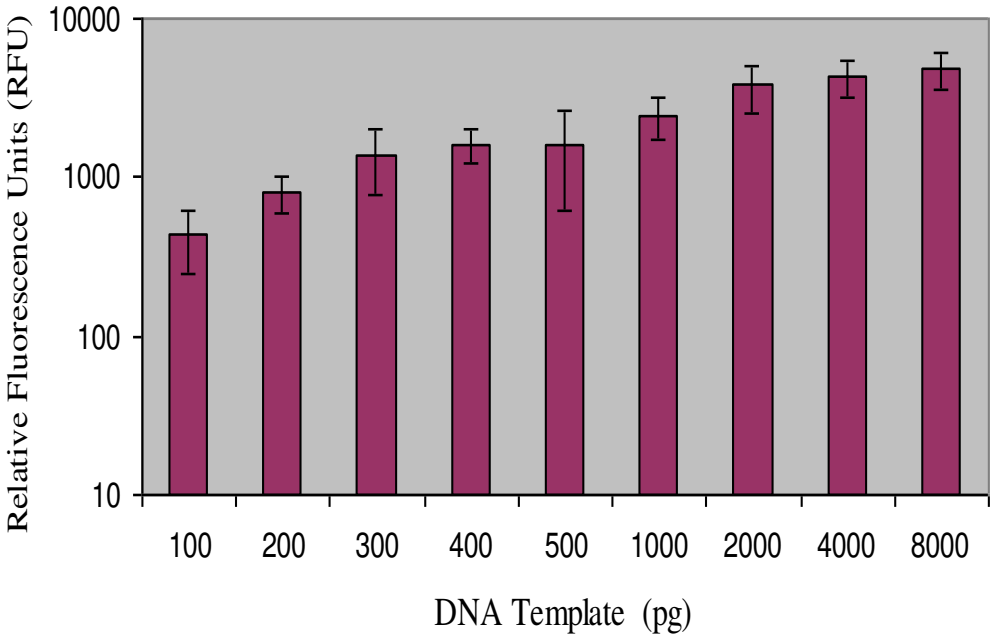


Figure 5.1. Shown above are the RFUs obtained from the sensitivity study of the 4 SNPs using two DNA samples. Normalised average RFUs are shown. The error bars indicate the standard error of the mean.

5.5. Discussion

Population Study

The 66 loci produced the genotyping results expected in accordance with HWE. To our knowledge this is the first report of allele frequencies for SNPs in the UAE population. The allele distribution of all loci proved to be polymorphic with a minimum allele frequency of 0.14, which is in good agreement with the value of 0.17 reported by Sanchez et al (2006).

Forensic Statistical Analysis

A high average heterozygosity was found with a value of 0.47 and thus, the selected 66 loci would be expected to exhibit high variability between samples. This is very valuable for forensic application as increases in heterozygosity improving individualisation of samples under comparison (Vallone et al., 2005). The value obtained for heterozygosity was not surprising, considering that one of the initial criteria for SNP selection, based on frequencies ranging 0.45-0.55, was priority to maximise the heterozygosity in the developed SNPs, albeit that the initial allele frequencies were based on only 20 alleles.

The forensic characterisation of the 66 SNP panel showed encouraging features. With 66 SNPs, the combined power of discrimination of > 0.99999999 was in the range achieved with the 52 loci (> 99.99999) reported by Sanchez et al (2006). The match probability of 3.058×10^{-25} was found to be higher than the match probability achieved with the CODIS markers 10^{-15} (Kidd et al., 2006). Although SNPs are not as polymorphic as multiallelic STRs; the biallelic SNP showed abilities to discriminate between unrelated and related individuals when a reasonable number of loci are developed.

Sensitivity Study

The SNP typing results were reproducible and sensitive. The SNP profiles obtained from all the triplicates tested for reproducibility in the 25 individuals were all concordant even when SNP profiles were obtained in samples with as little as 100 pg template DNA. However, completely balanced genotyping was obtained at 300 pg compared to 500 pg needed for STR typing (Butler et al., 2007). The 52 plex that were developed by Sanchez, et al (2005) showed complete SNP profiles from 500 pg. This demonstrated that the SNPs developed in this study are suitable to be used for forensic samples.

5.6. Conclusion

In conclusion, the studies presented in this chapter show that the developed 66 SNPs offer both the potential for genotyping with forensic samples. The sensitivity studies conducted demonstrated that the SNP loci were as sensitive, and in many cases more sensitive than STR systems. The sensitivity levels were similar with larger multiplexes (Dixon et al., 2005b; Sanchez et al., 2006).

CHAPTER 6

ANALYSIS of ARTFICIALLY DEGRADED DNA and CASEWORK SAMPLES

6.1. Overview

In many cases, forensic scientists involved in the analysis of biological materials can only generate incomplete DNA profiles (Fondevila et al., 2008) as DNA will often undergo gradual fragmentation, causing the loss of one of the PCR primer binding sites (Pang and Cheung, 2007). Amplification failure leads to the loss of vital genetic information, which can be important for identification and comparison purposes: DNA samples of this nature are classed as degraded (Bender et al., 2004).

In desert countries, such as the Gulf Region, a hot and humid environment is commonly found throughout the year; and this can be problematic when generating DNA profiles from forensic evidence. In this study the effect of two environmental factors on the degradation of DNA, the temperature and humidity, were assessed. Also, DNA samples subjected to endonuclease enzymatic degradation were included in this study.

In real casework, most of the saliva and semen samples brought to the laboratory for analysis are collected using a swab. Therefore, in order to assess the effect of different environments on biological samples, saliva and semen were applied to swabs and incubated in both controlled and different natural environments. STRs and SNPs were used to assess the effectiveness of different markers when analysing degraded DNA.

6.2. Aims of this Chapter

To test the hypotheses that:

- high temperature and humidity will increase the degradation of DNA;
- SNPs of less than 150 bp can be used efficiently to improve allele profiling of degraded DNA; and

- to assess and evaluate, the performance of SNPs on degraded samples compared to STRs that are used routinely in forensic laboratories and in particular, the AmpFℓSTR® SGM Plus® (Applied Biosystems).

6.3. Samples

Saliva and semen samples were used in this study because these types of stains are commonly encountered at crimes scenes. Also, these samples were obtained without difficulty from volunteers at the time the experiment was conducted. Saliva and seminal fluid samples were collected from two individuals. DNA extractions that were degraded using DNase 1 from different incubation periods of 10, 60 and 180 minutes were also used. Analysis of these samples were carried out in the laboratory as described by Zahra (2009). Teeth samples were obtained from 8 different human remains, all of which were greater than 4 years old.

6.4. Results

6.4.1. DNA Extraction and Quantification

Experiments to determine the effects of different environmental conditions (Table 6.1) on saliva and semen samples were performed. Extraction procedures for all saliva samples were carried out using Qiagen® QIAamp® DNA Mini Kit as described in Section 2.4.6 and DNA from semen was extracted using Qiagen® QIAamp® DNA according to the manufacturer's protocol as described in Section 2.5.6

DNA was estimated using the Quantifiler® Human DNA kit with the ABI 7500 real time PCR machine as described in Section 2.2.2.1.

The results that were obtained from the Quantifiler® DNA showed that the amount of DNA degradation was dependent on the type of sample analysed. These results are shown in Tables 6.2, 6.3 and 6.4.

Table 6.2. Indicated below are the different environmental conditions that were induced to generate degraded DNA.

Indoor environment (Saliva and Semen samples)		
100% humidity (37 °C)		Room temperature (22 °C)
Outdoor environment (saliva samples)		
UAE summer (September)	UAE Winter (December/January)	UK summer (August)

Table 6.1. Shown below are quantification results from semen and saliva samples studied at room temperature (22 °C). 50 µl of sample was added to a swab and the final extracted volume was 150 µl.

	Quantification values (ng/µl)													
	Saliva							Semen						
Days	0	3	6	9	12	15	18	0	3	6	9	12	15	18
Individual 1	1.0	1.33	2.29	1.46	1.13	3.02	1.19	4.22	12.73	9.13	7.88	16.57	14.57	17.29
Individual 2	1.75	4.09	6.29	5.04	4.38	1.38	1.63	5.67	4.59	4.30	4.85	3.40	3.02	5.22

Table 6.3. Indicated below are quantification results from semen and saliva samples studied at 100% humidity and at 37 °C. 50 µl of sample was added to a swab and the final extracted volume was 150 µl. Sample 'not available' is represented by: N/A.

	Quantification values (ng/µl)													
	Saliva							Semen						
Days	0	3	6	9	12	15	18	0	3	6	9	12	15	18
Individual 1	1.0	0.22	0.03	0.04	0.01	0.01	0.01	4.22	16.93	21.42	8.86	22.76	3.57	15.49
Individual 2	1.75	0.04	0.11	0.00	0.01	0.01	0.00	5.67	44.69	33.32	29.19	NA	21.97	11.44

Table 6.4. Indicated below are quantification results for DNA in saliva samples under natural conditions in UAE and UK environments with 50 µl samples. The final extracted volume was 150 µl. Sample 'not available' is represented by: N/A.

Quantification values (ng/µl)			
Time intervals (days)	UAE Dec/Jan 2008	UAE Sept 2008	UK Aug 2008
0	2.50	6.01	1.00
3	3.51	1.57	3.16
6	3.20	2.62	1.58
9	NA	NA	0.35
12	1.59	0.31	0.20
15	NA	NA	0.09
18	NA	0.05	0.02

The quantifications obtained from semen were variable in comparison with the reference samples for both individuals. For example in individual 1, the value for the sample incubated for 9 days was estimated at 9 ng/µl, compared to 22 ng/µl for the sample incubated for 12 days. Due to the viscosity of the semen sample a constant volume of pipetting could not be achieved. Also, a difference was observed between the amounts of DNA estimated for each individual. This could be a natural occurrence as different concentrations of DNA are produced by different individual. The same results were observed between control and degraded saliva samples, but with lesser variation than the semen samples.

6.4.2. DNA Genotyping

6.4.2.1. Performance of SNPs and STRs

Criteria for the triplex development of 6 loci were described in Chapter 2 and Chapter 4. Also, as mentioned before, each sample was amplified and genotyped three times to ensure reproducibility, and an average of the results is presented. Based on the number of alleles profiled from the two triplexes (12 alleles), the genotyping results were calculated as a percentage (%). A partial profile was designated as (pp) and no profile as (np). The reference samples for each individual were genotyped as a control, producing 12 alleles with which the subsequent profiles were compared (Figure 6.1 and Figure 6.2).

The results for SGM plus® were also calculated as a genotype percentages with the amelogenin locus emitted from the analysis, therefore 100% allele profile was estimated as the presence of all 20 alleles in the 10 loci (Figure 6.3). The amount of sample used for STR genotyping was the same to that used for SNP analysis, ranged from 0.06 ng to 0.5 ng.

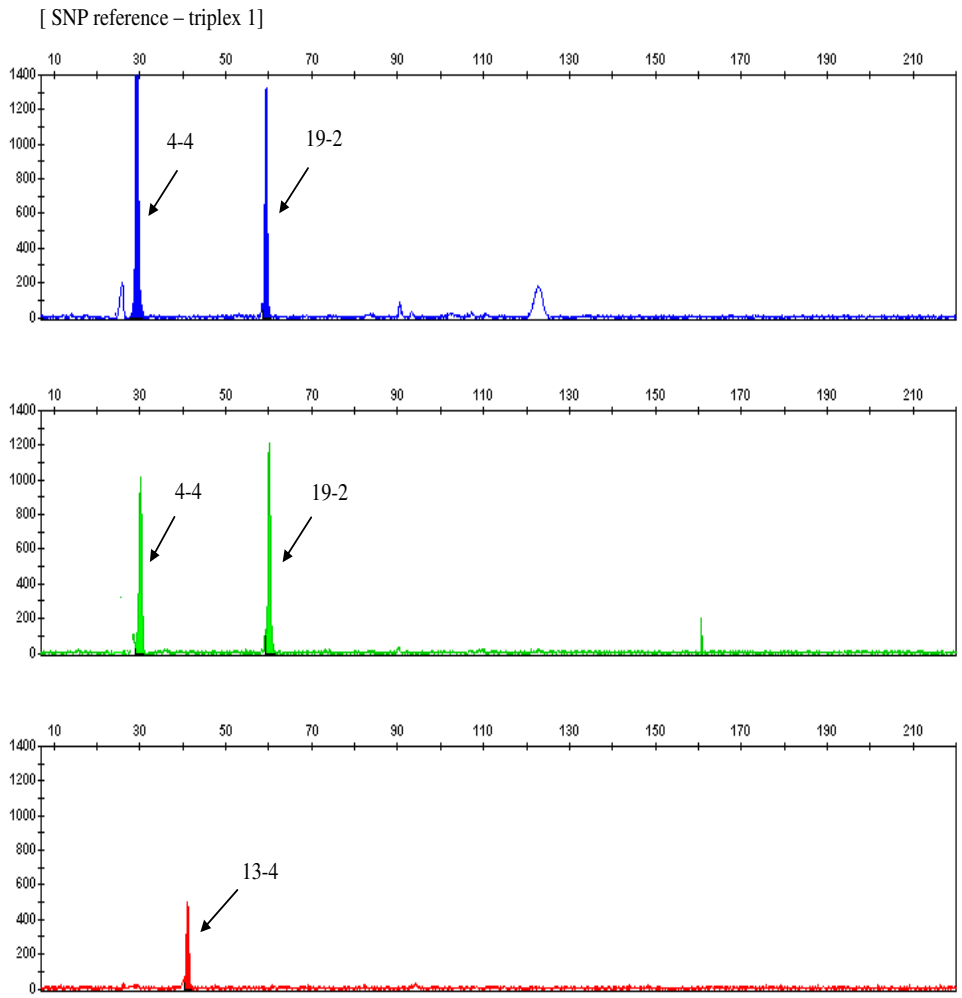


Figure 6.1. Shown above is the electropherogram for multiples 1 for the reference sample that was used as standard to assess the allele profiles.

[SNP reference-triplex 2]

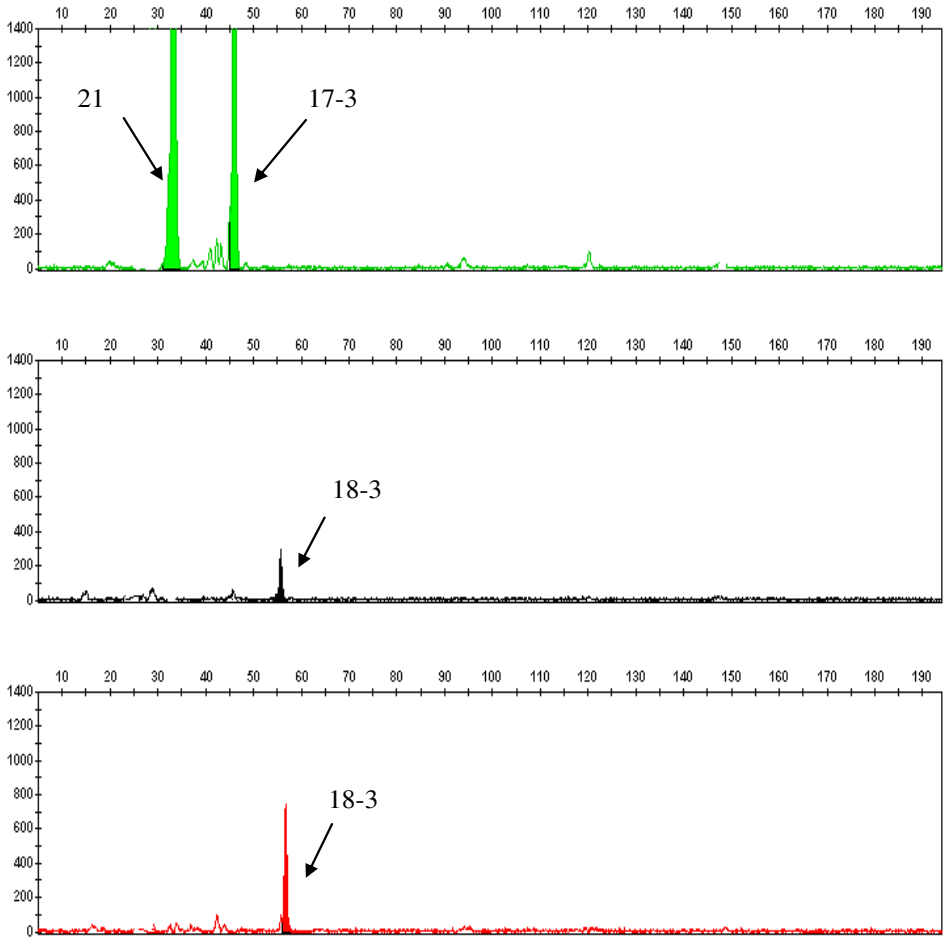


Figure 6.2. Shown above is the electropherogram for multiplex 2 for the reference sample profiles.

STR reference sample

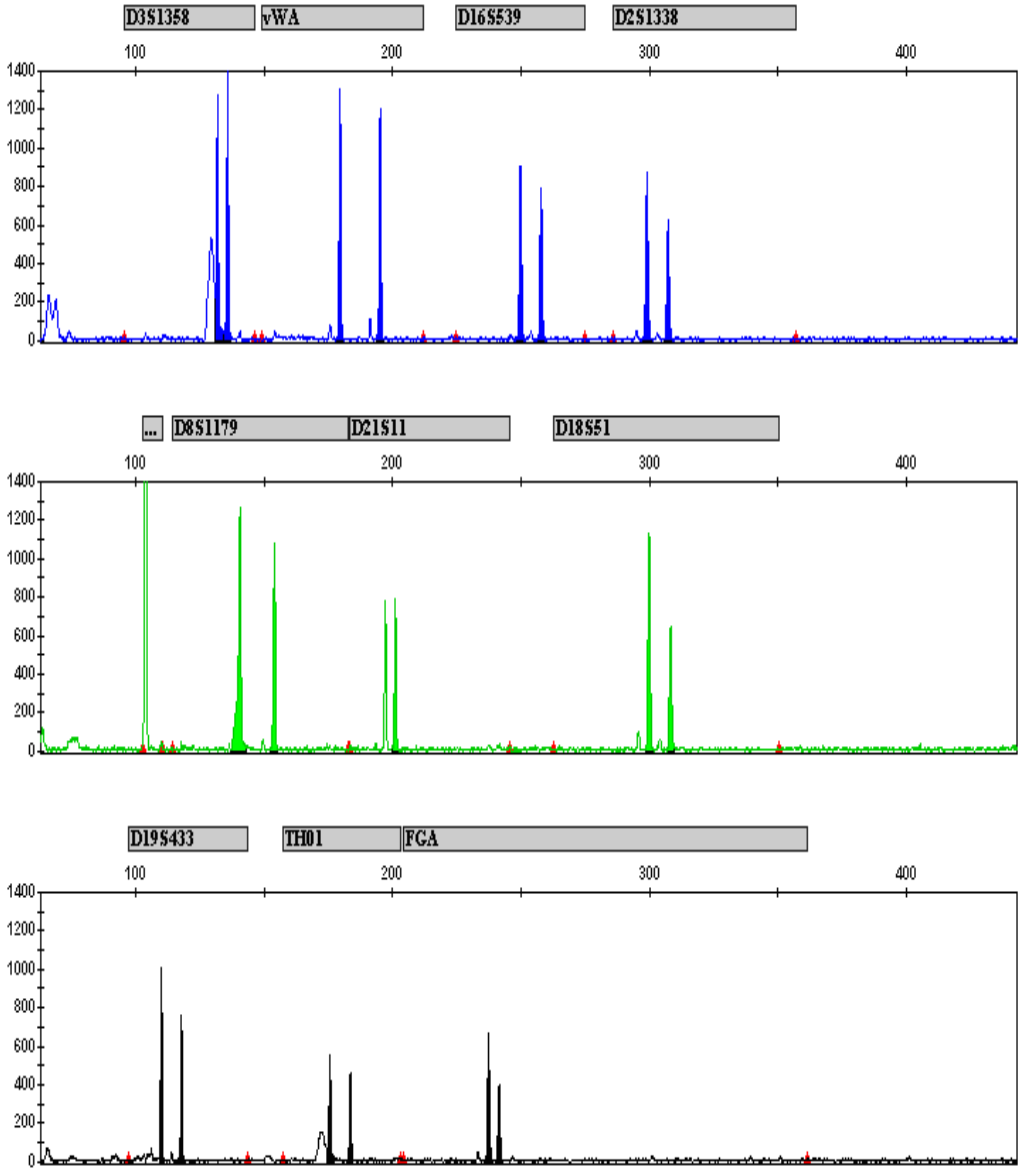


Figure 6.3. Shown above is the electropherogram for the reference sample profiled with SGM plus®.

6.4.2.2. Degradation at 37 °C and 100% Humidity

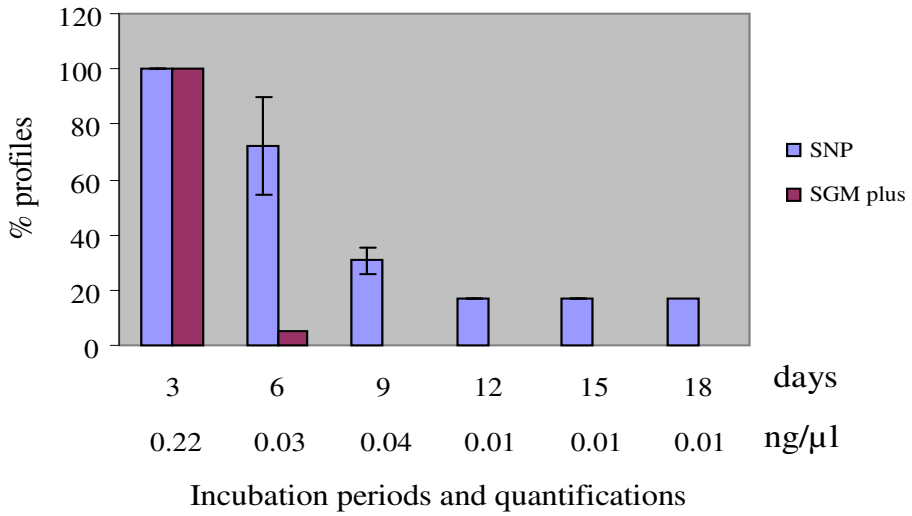
SNPs and STRs Typing of Saliva

The results of the SNP and STR typing are shown in (Figure 6.4).

In SNP typing, the signal strength obtained for each allele was dependent on the nature of the dyes incorporated for each ddNTP (Figure 6.5). The lowest peak heights were observed for ddCTP (dTAMRA™, yellow) and ddTTP (dROX™, red), which is consistent with previous observations (Vallone et al., 2004, Sanchez et al., 2006).

The amount of DNA template used was 0.5 ng for the PCR reaction whenever possible. In some reactions, a reduced amount of DNA as low as 0.06 ng in the highly fragmented DNA samples was used for amplification in both SNP and STR analysis such as, saliva sample taken in interval 9, 12, 15 and 18.

[A] Saliva- humidity/ temperature individual 1



[B] Saliva- humidity/temperature individual 2

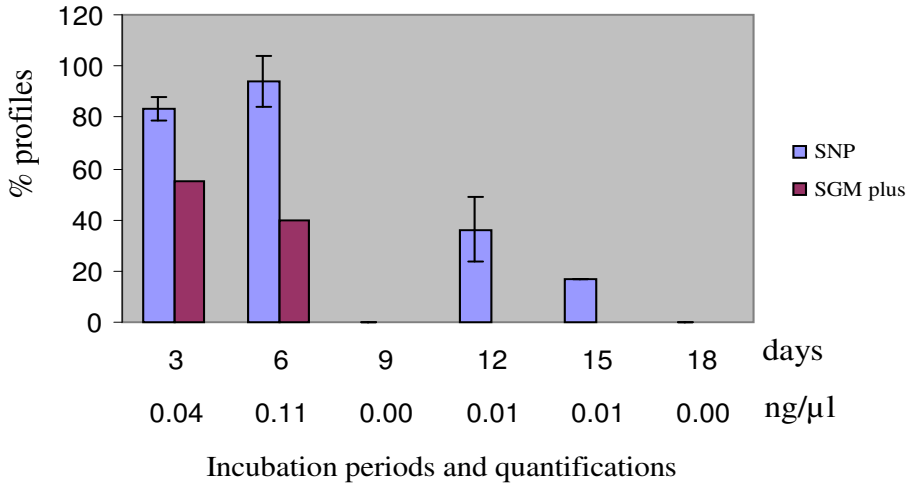


Figure 6.4. Shown above is percentage of profiles obtained from artificially degraded DNA from saliva samples under 100% humidity at 37 °C with their corresponding DNA concentrations. The results are for SNaPshot™ and SGM plus® for individual 1 (A) and individual 2 (B). The error bars indicate the standard deviation.

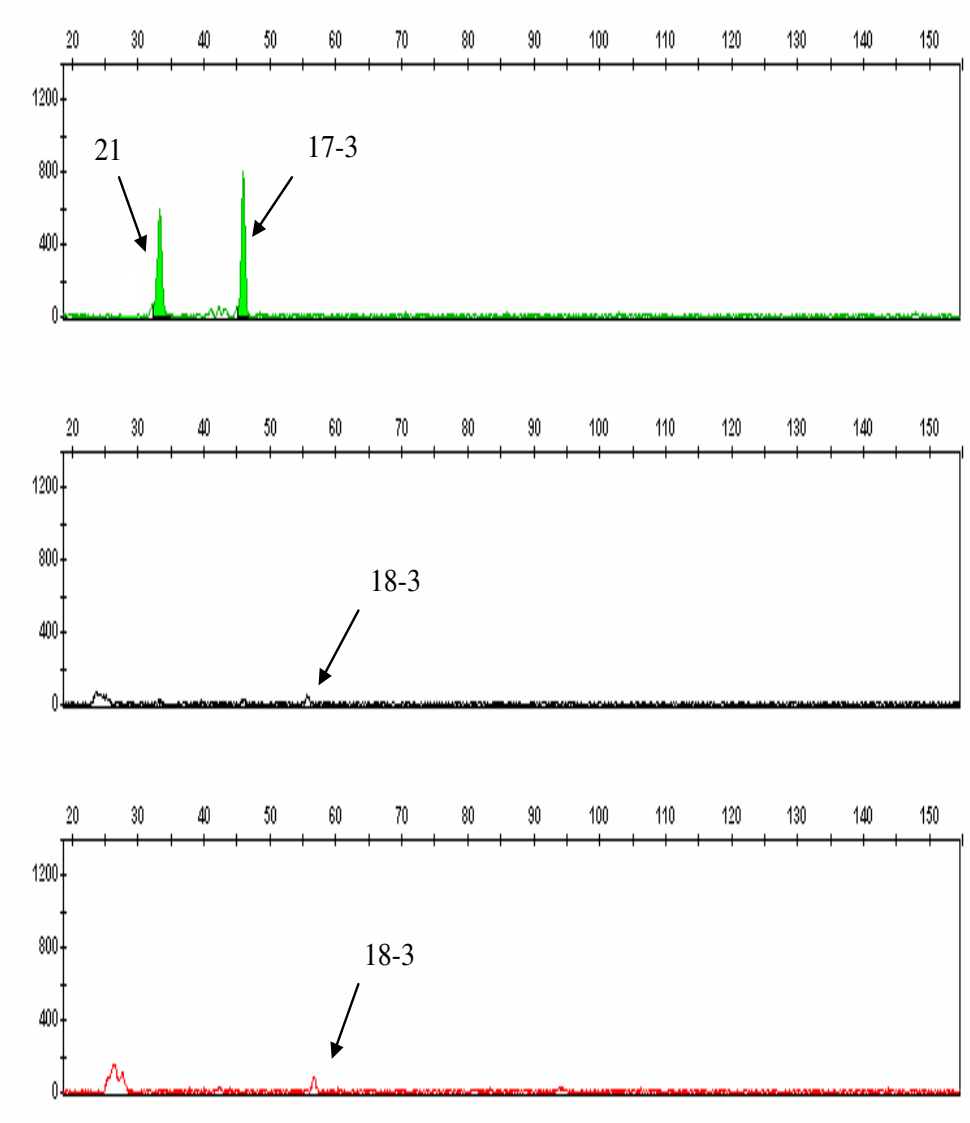


Figure 6.5. Shown above is an electropherogram of alleles below the RFU threshold (100) at C (black) and T (red) as a result of dye effect for locus 18-3. Alleles for loci 21 and 17-3 were above the threshold.

In order to evaluate the efficiency and the contribution of each locus in both triplexes, the percentage of each locus was calculated: for each locus, the total number of observed alleles in the three repeats (Appendex A2A and A2B) was divided by the total number of expected alleles. The average for both individuals was determined. SNP code 21 performed the best with 100% amplification followed by 4-4 (62%), 17-3 (59.5%), 19-2 (51.8%), 13-4 (45.2%) and 18-3 was the lowest contributor with 38.9%. Although both SNP code 21 and 4-4 are of a similar amplicon size, 4-4 showed a remarkably

lower percentage than code 21; this is because locus 4-4 for individual 1 was observed to be heterozygous (AG) and because of the difference in signal strength between the dyes (Vallone et al., 2004). Allele A was the first to dropout, giving a partial profile at day 12. Also, it could be that the template sequence for locus 4-4 was more affected by prolonged degradation (day 15 and 18) with complete allele dropout when compared to locus 21 (Dixon et al., 2005a).

The percentage of each locus for SGM plus® profiling (Appendex A3) was also calculated as for SNP profiling.

SNPs and STRs Typing of Semen

The experiment performed for semen samples from both individuals showed full SNP and STR profiles in all incubation periods (Appendix A4A, A4B and A4C). This may be because the degradation period was not long enough to affect the PCR primer target sequence of the DNA template (Figure 6.6 A and B). However, this observation was in agreement with a previous degradation experiment on semen samples where full DNA profiles were obtained after 243 days incubation at 37 °C and after 24 days at 100% humidity (Cotton et al., 2000, Dixon et al., 2005b).

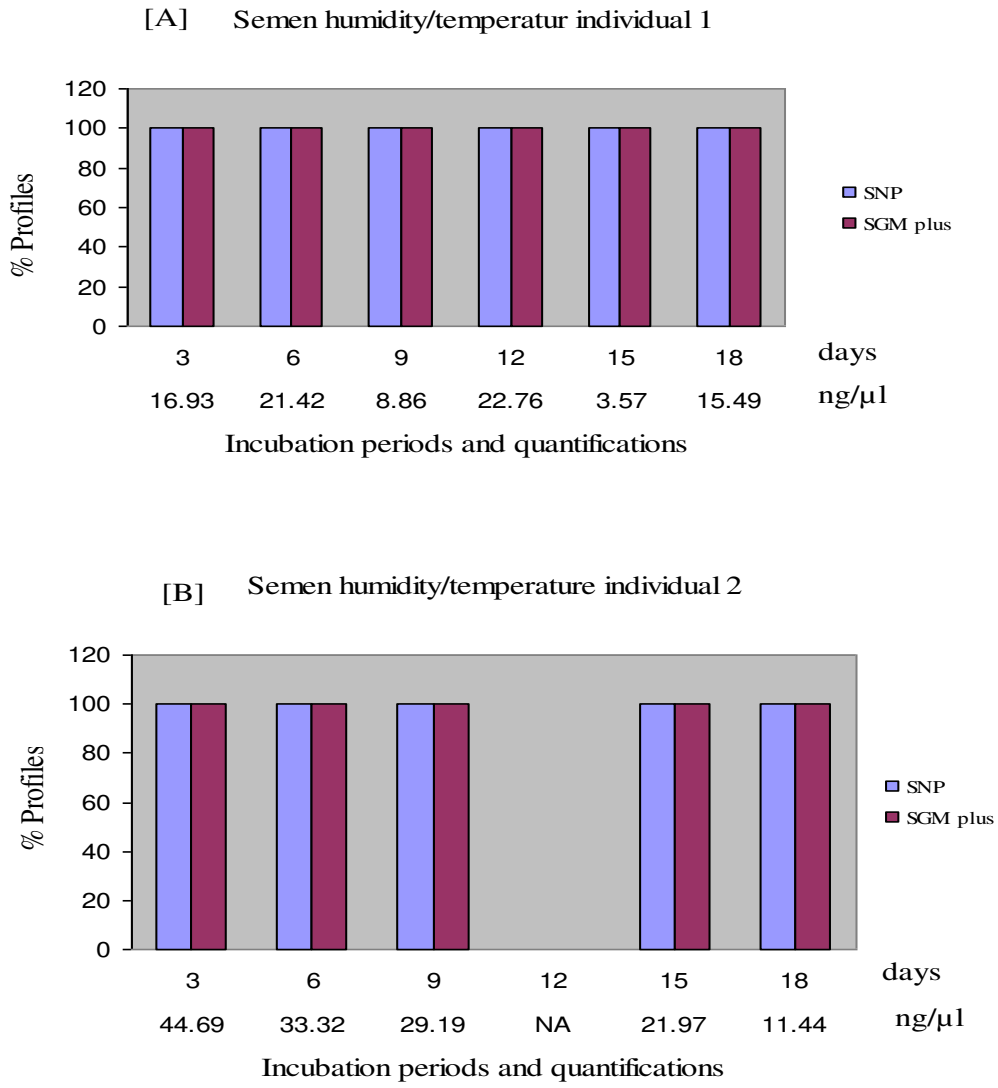


Figure 6.6. Shown above are profiles of 100% obtained from artificially degraded DNA from semen samples under 100% humidity and 37 °C with their corresponding DNA concentrations. The results are for SNaPshot™ and SGM plus® for individual 1 (A) and individual 2 (B). NA; represents not available sample.

6.4.2.3. Degradation at Room Temperature

SNPs and STRs Typing of Saliva and Semen

In order to check the effect of temperature alone, or at least reducing the influence of other weather effects such as sun radiation and humidity, saliva and semen samples were kept at average room temperature, which was recorded as 22 °C. At the time, the

experiment was conducted, the laboratory temperature was observed to be approximately 4 °C higher than the average atmosphere temperature outdoors (18 °C).

As expected from the quantification values, a full profile (100%) was obtained for saliva in the cases of both SNP and STR (appendex A5, A5B and A5C). This observation strongly indicates that an indoor temperature below 24 °C and incubation period of 18 days did not have major effects on the DNA template (Figure 6.7 A and B).

Since a full semen DNA profile was obtained in the previous experiment (100% humidity/ 37 °C temperature) at all time intervals; it was assumed that under the less stringent environmental factor (22 °C), the DNA template would also exhibit a 100% successful genotyping results. Therefore, semen DNA for this experiment was not genotyped.

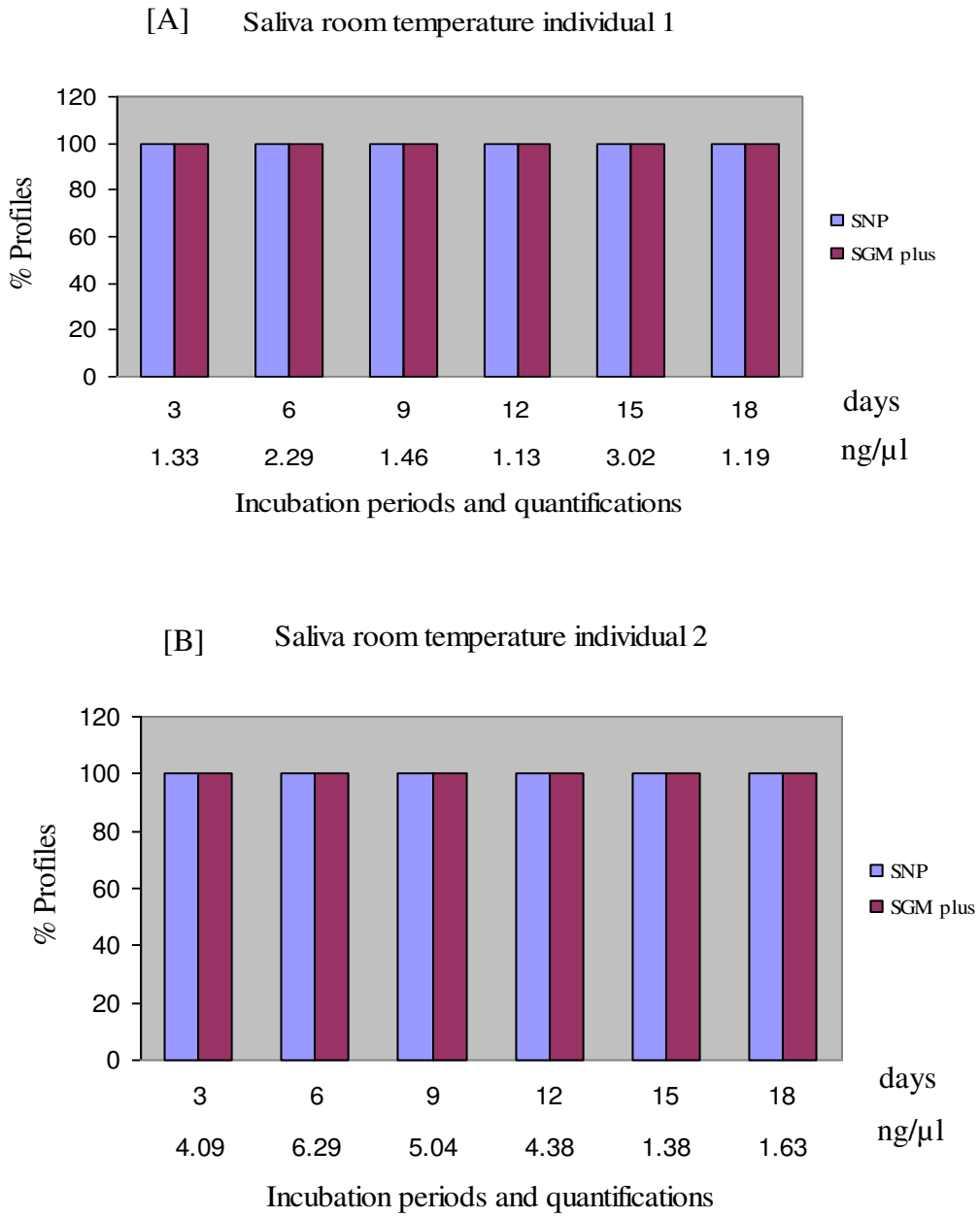


Figure 6.7. Shown above are profiles obtained from artificially degraded DNA from saliva samples under 100% humidity and 37 °C, also shown are their corresponding DNA concentrations. The results are for SNaPshot™ and SGM plus® for individual 1 (A) and individual 2 (B). The amount of DNA template used was 0.5 ng for the PCR reaction.

6.4.3. Outdoor Environment

The reason behind this methodology was to observe the effect of different temperatures and other naturally occurring weather elements on biological samples. The temperature in this study, ranged from less than 20 °C to more than 37 °C, which was classified for simplicity as cold, mild and hot temperatures. In order to achieve such ranges of temperature naturally, the sample was exposed to three different environments; the UAE environment: December 2007/ January 2008 (Figure 6.8); mild temperature up to 22 °C, partial cloud, and average relative humidity up to 50%; September/October 2008; hot with average temperatures reaching 34 °C, sunny and an average relative humidity of up to 58% (Figure 6.9). UK weather: August 2008, cold temperature less than 20 °C, raining, and average relative humidity up to 92 % (Figure 6.10). An aliquot of the same saliva sample (female) was exposed to each of the three conditions.

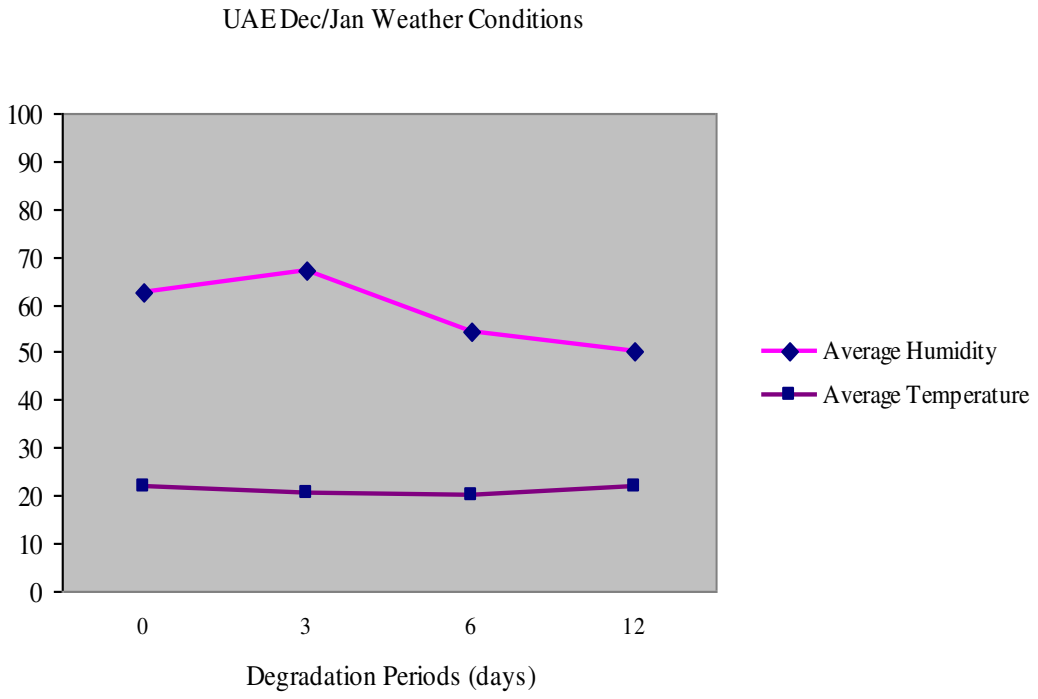


Figure 6.8. Shown above are UAE December/ January average temperatures and humidity for each of the degradation period. The average of temperature and humidity was calculated based on the hourly data (24 hours) obtained for each of degradation periods.

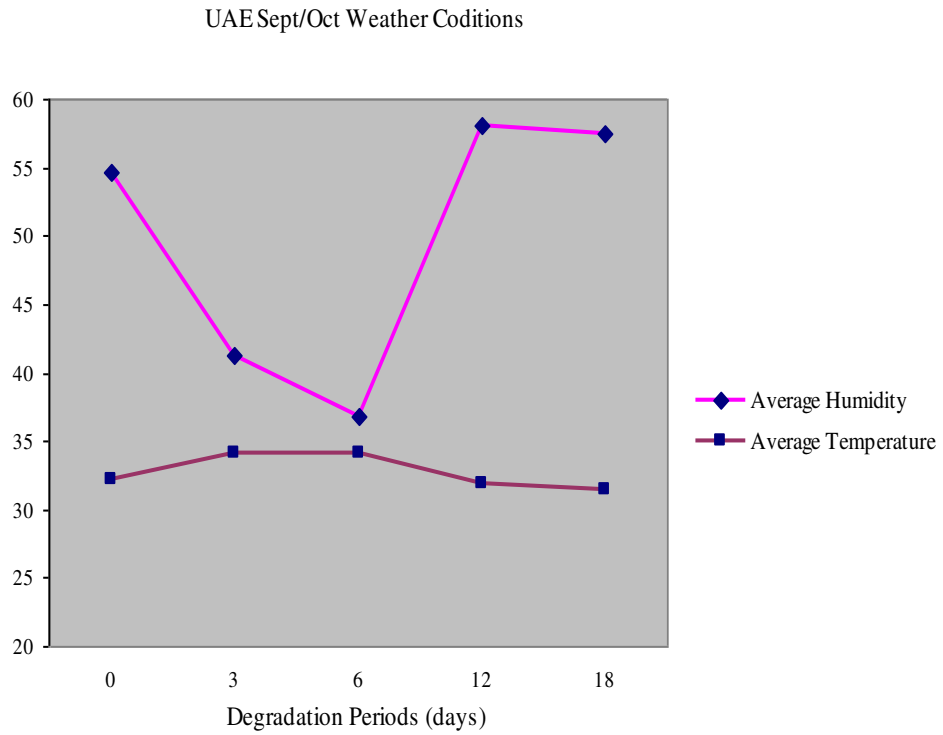


Figure 6.9. Shown above are UAE September/October average temperatures and humidity for each of the degradation period. The average of temperature and humidity was calculated based on the hourly data (24 hours) obtained for each of degradation periods.

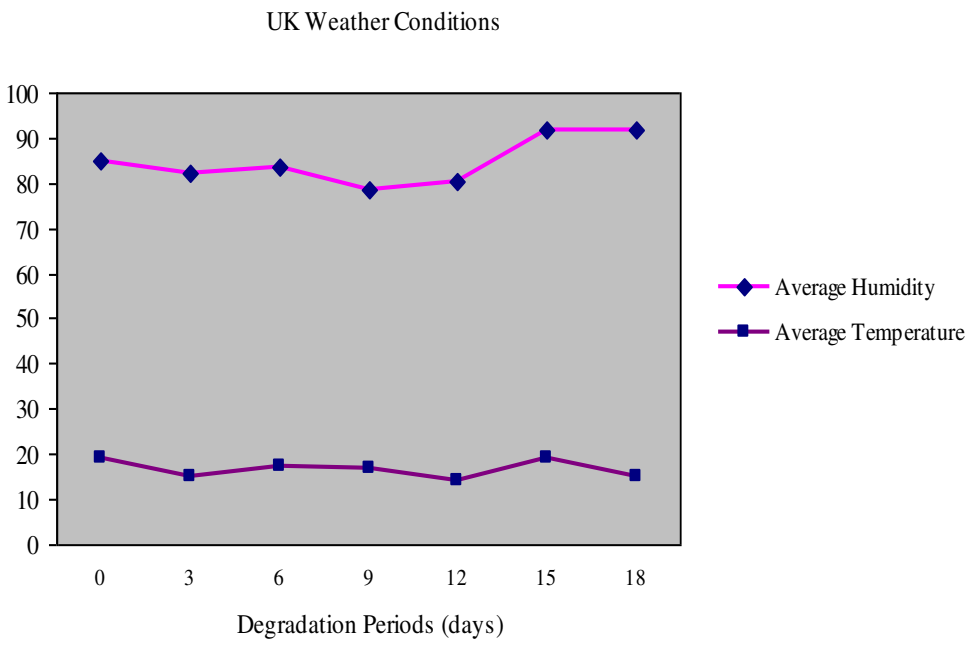


Figure 6.10. Shown above are UK August average temperatures and humidity for each of the degradation period. The average of temperature and humidity was calculated based on the hourly data (24 hours) obtained for each of degradation periods.

6.4.3.1. SNP and STR Profiles

Based on the results obtained from quantification (above Table 6.4), 0.5 ng of DNA was used for amplification in most reaction unless otherwise mentioned.

UAE- December 2007/ January 2008

SNPs and STRs Typing

The results are shown in (Figure 6.11).

As mentioned above , the amplification for SNP typing of each sample was performed in triplicate. The duration of the experiment was for 12 days due to time constrain In this experiment, the sample exhibited little degradation and full SNP profiles were observed in all time intervals except for complete dropout at locus 13-4 in the second

repeat of triplex 1 and one allele dropout at locus 18-3 in the second and third repeat of triplex 2 (Appendex A6).

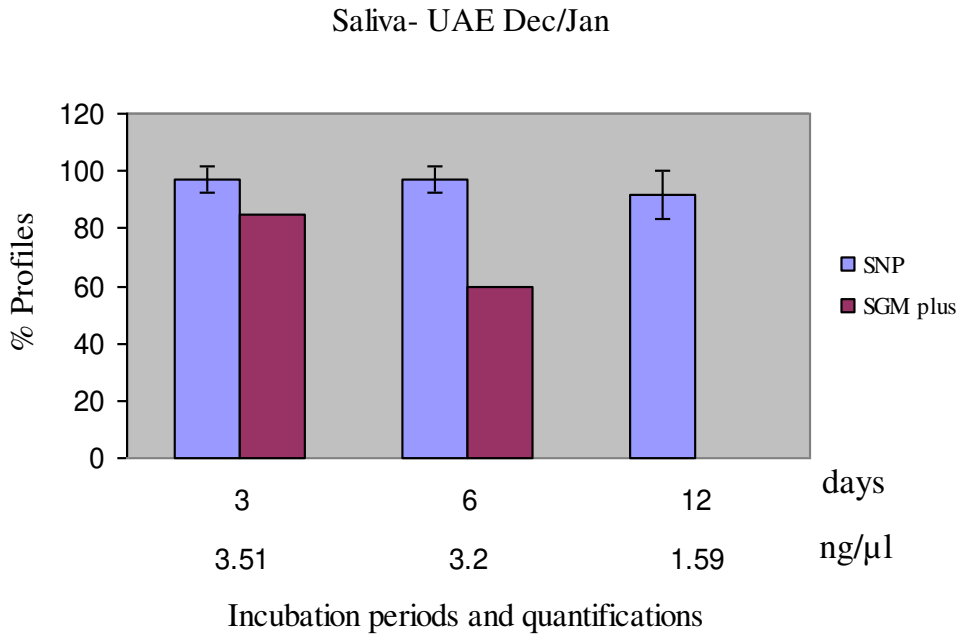


Figure 6.11. Shown above is the percentage of profiles obtained from degraded DNA from saliva samples under natural conditions of the UAE in December/January. The results are for both SNaPshot™ and SGM plus®. The error bars indicate the standard deviation.

The STR typing gave partial profiles with most affected alleles were those present in the FGA locus (Appendex A7).

UAE- September 2008

SNPs and STRs Typing

The average temperature of 34 °C and average relative humidity of 58% in this period had a high effect on the saliva samples (Figure 6.12). An average SNP profiling efficiency of 48.9% (partial profile) was observed, with the most affected locus, 18-3,

(Appendex A8) only profiling a total of 50%. Whilst the STR typing gave an average profiling efficiency of 25% (Appendex A9).

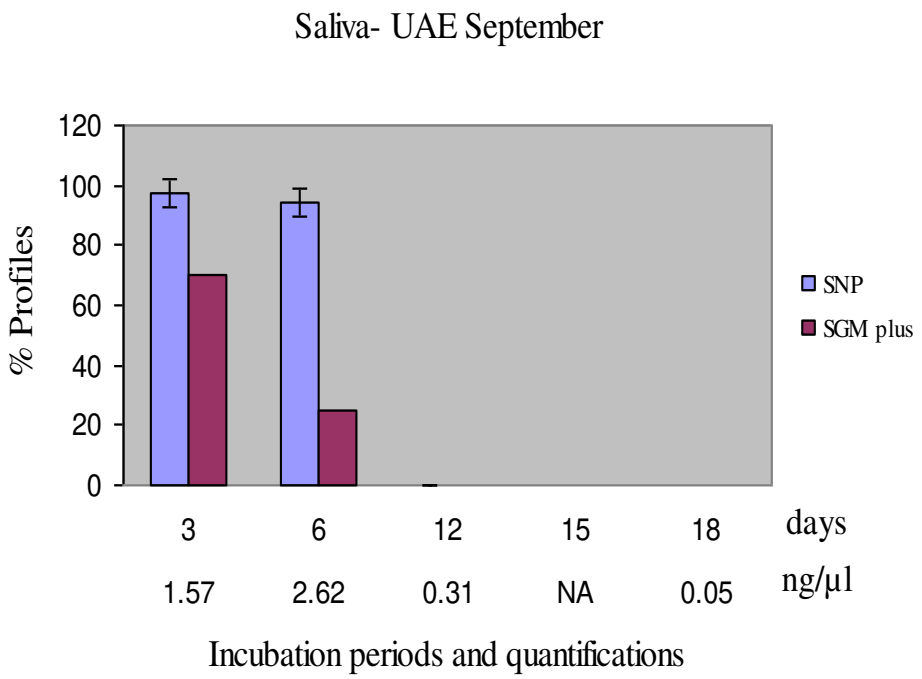


Figure 6.12. Shown above is the percentage of profiles obtained from degraded DNA from saliva samples under natural conditions of the UAE in September. The results are for both SNaPshot™ and SGM plus®. The error bars indicate the standard deviation.

UK- August 2008

SNPs and STRs Typing

The results are shown in (Figure 6.13). For the sample degraded for 18 days, the amount of DNA template used for amplification was estimated as 0.36 ng.

The effect of an average temperature of 16 °C and up to 92% average humidity varied between different time intervals (Appendex A10 and A11). The overall genotyping percentage was found to be 78.7% for SNPs and 40.8% for STRs.

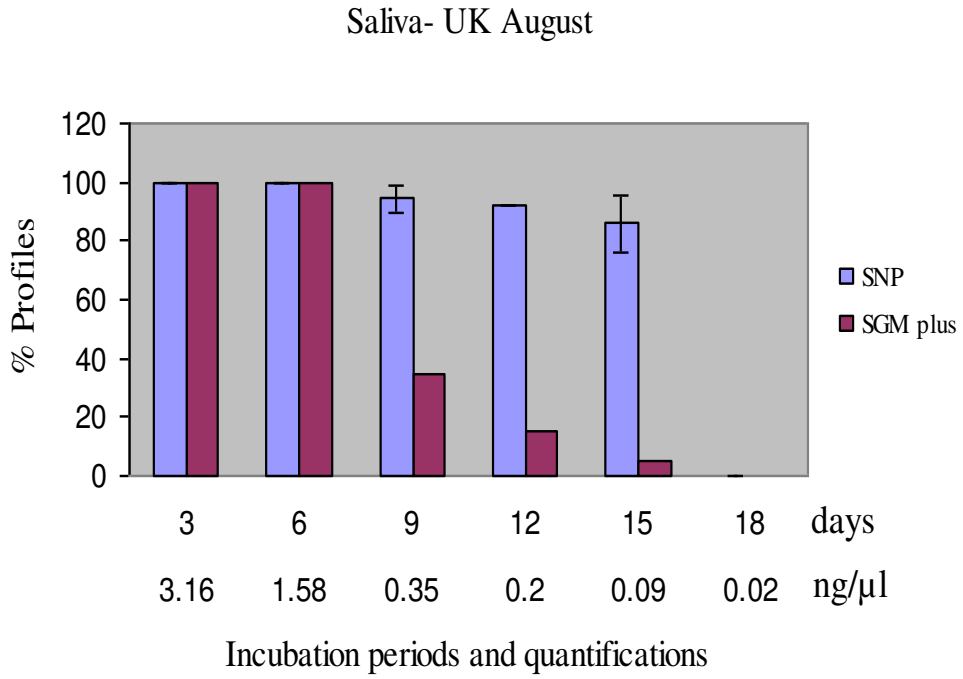


Figure 6.13. Shown above is the percentage of profiles obtained from degraded DNA from saliva samples under natural condition in the UK in August. The results are for both SNaPshot™ and SGM plus®. The error bars indicate the standard deviation.

6.4.4. Comparison between SNP and STR Profiling

In this comparison, the results obtained for degraded saliva samples incubated for 6 days under all the conditions employed are illustrated in the following figures. Whenever 2 individuals were included in the degradation experiment; samples from individual 1 were only used for the comparison.

Using the artificially degraded DNA samples, the comparisons between SNP and STR analysis showed that the amplification of severely fragmented DNA templates were

more successful using SNP genotyping. In many cases full allele profiling was obtained by using SNPs, whilst only partial profiles were sometimes recovered using STRs However, in severely fragmented DNA, dropout of alleles was observed for both systems.

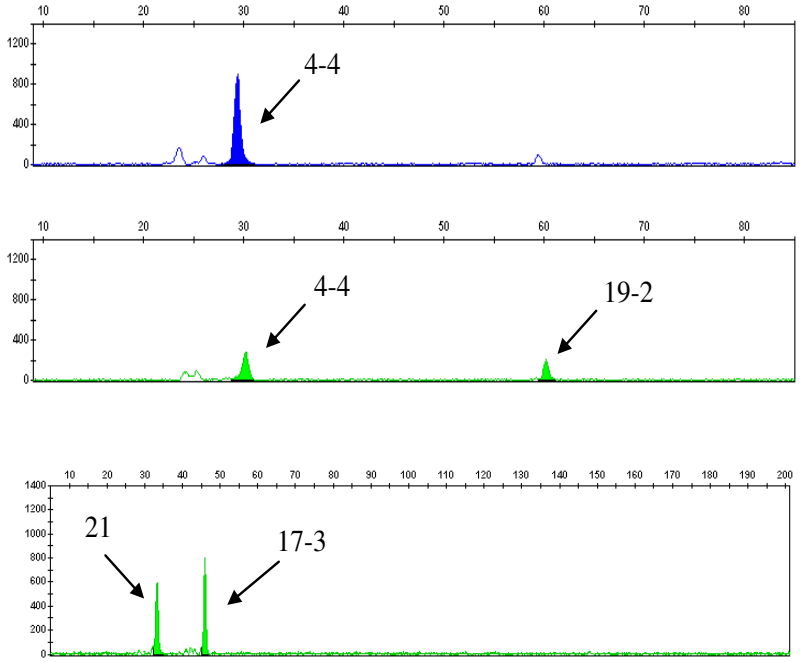
Comparing overall the percentage of genotypes obtained from all saliva samples degraded under 100% humidity and 37 °C temperature conditions (Figure 6.14 A and B), 37.7% of allele profiles were observed for SNP and 16.7% of profiles were observed for STR analysis.

In the natural environment, intact DNA was exposed to more than two factors such as wind, sun radiation (UV), humidity, moisture and temperature. Dependent on the environmental conditions, the DNA samples exhibited variation in the amount of degradation observed. The UAE samples degraded in December/ January (Figure 6.15 C and D) showed 95.4% SNP profiles and 48.3% of STR profiles, whilst samples degraded in September (Figure 6.16 E and F) gave 47.9% SNP profiles and 23.8% STR profiles. Alternatively, samples that were subjected to UK weather conditions (Figure 6.17 G and H), exhibited 77.8% for SNP profiles and 42.5% for STR profiles.

Amplification efficiency of samples that were degraded in the UAE September environment, showed the least efficiency, because the combination of >87% humidity, >37 °C and sunny conditions collectively caused the DNA to be fragmented to a greater extent than the other conditions (UAE, December/January and UK, August conditions). Also, ultraviolet radiation from the sun light could alter the primary structure of DNA strand leading to the formation of thymidine dimerization (Mitchell et al., 1992). This did not fragment the DNA, but cross-link renders the DNA inert in a PCR. Ultimately, dropout of larger alleles especially for STR analysis was exhibited; this system was approximately 19.6% less efficient than the SNP amplification.

However, although the temperature (cold-17 °C) for UK degradation conditions was much lower than the temperature observed for UAE December/January (mild-23 °C) conditions, the efficiency of the PCR primers for UK samples incubated longer than 6 days gave less efficient results than were expected. A combination of 81% relative humidity and the damp environment resulting from continuous rain could be responsible of the increased degradation effects on the DNA samples.

[A] SNaPshot triplex1 and 2



[B] SGM plus

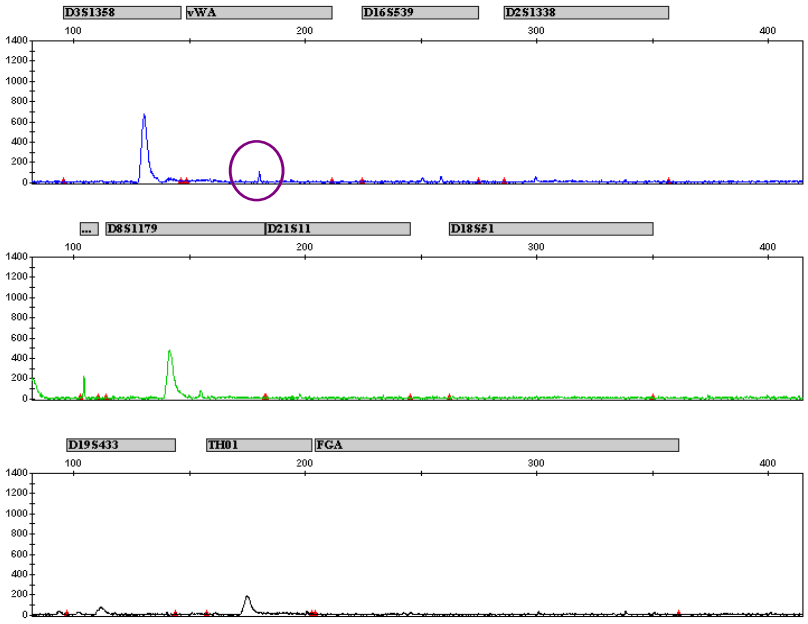
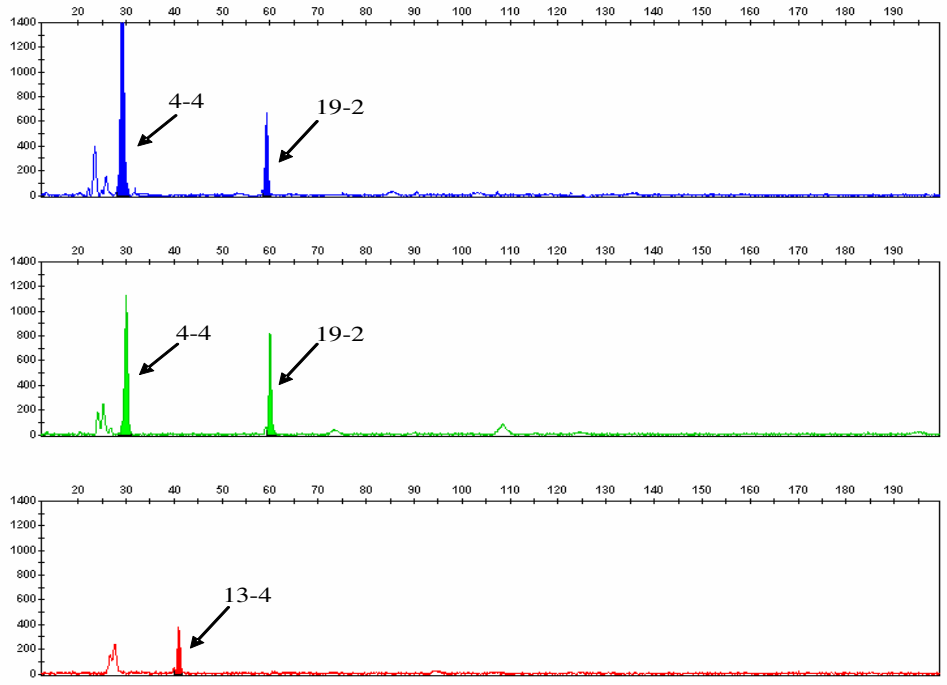
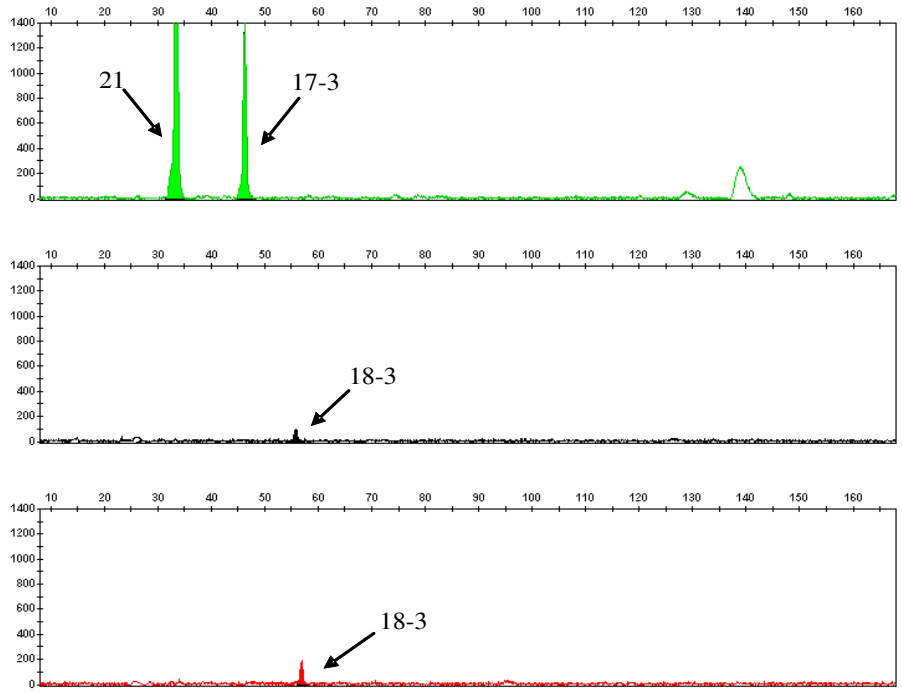


Figure 6.14. Shown above are electropherograms showing a comparison of allele genotyping that was obtained from (A) SNaPshot™ triplex and (B) from SGM plus®. 0.5ng of DNA from a sample degraded under humidity and 37 °C for 6 days for individual 1 was used for both systems. Allele profiles of 58.3% were obtained for SNP and 5% (one allele is circled) for STR.

[C-T1] SNP



[C-T2] SNP



[D] SGM plus

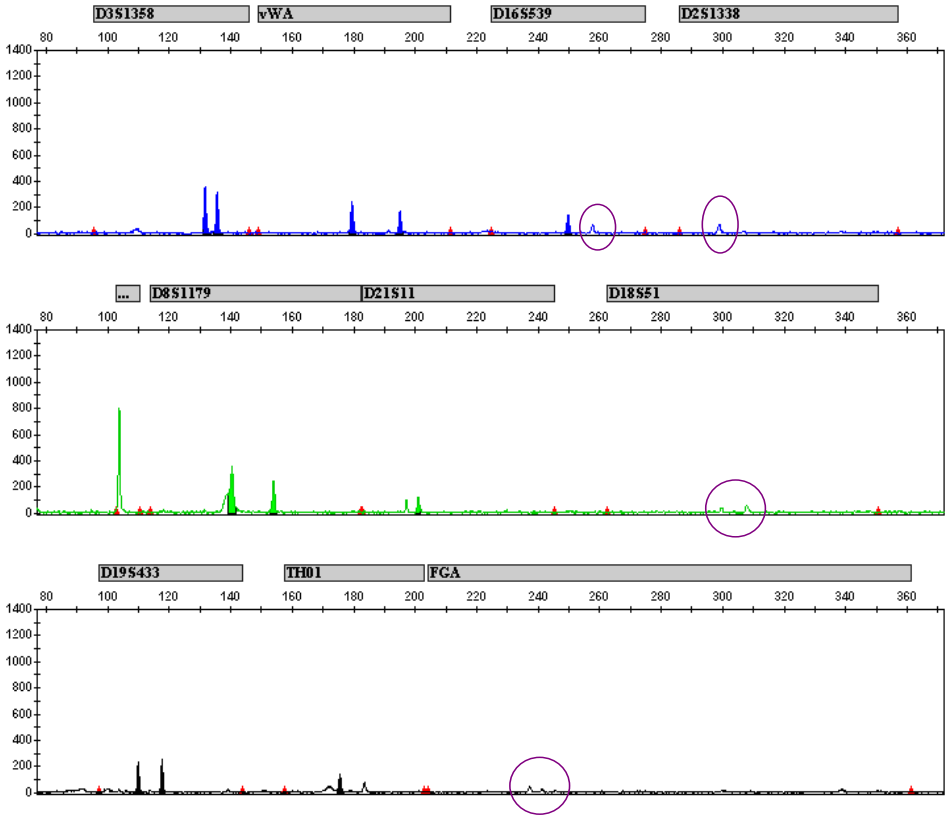
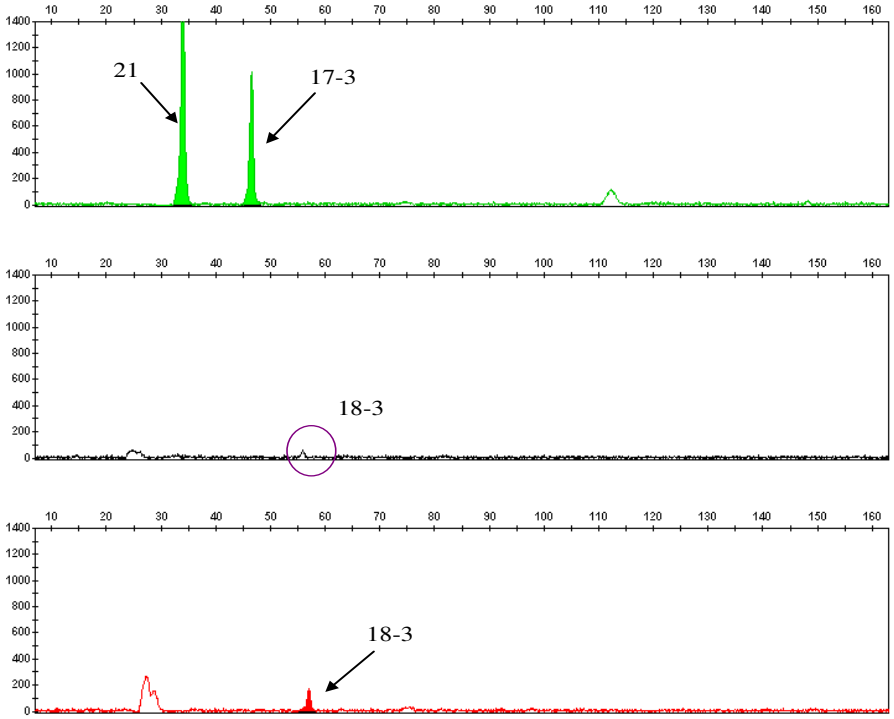
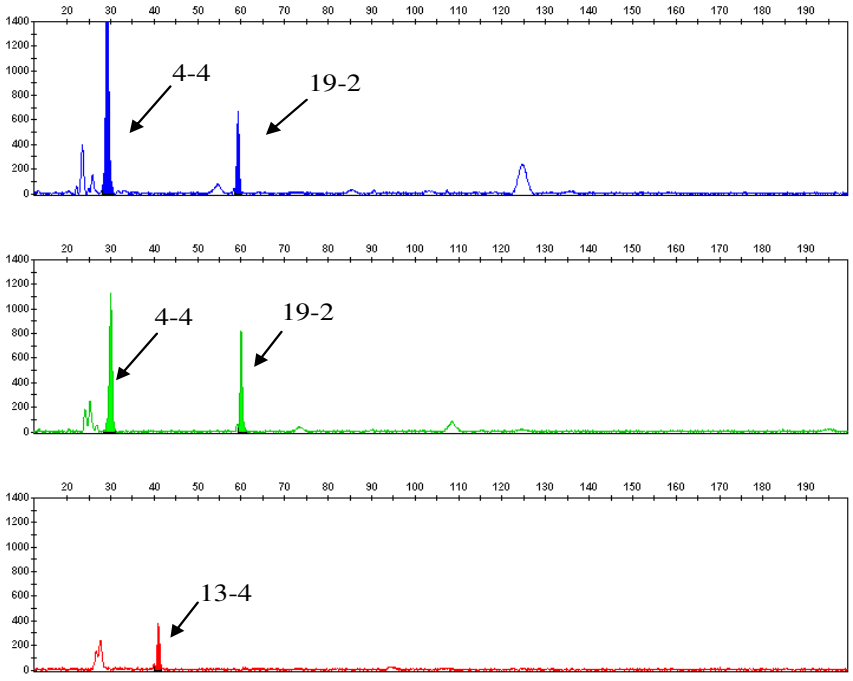


Figure 6.15. C-T1,C-T2 and D. Shown above are results for the samples at 6 day intervals obtained from UAE December/January degradation. Electropherograms C-T1 and C-T2 represent triplex 1 and triplex 2 of one of the repeats obtained from SNP genotyping with 100% profiles. D is the result for the same sample obtained from STR genotyping with 60% profiles. Arrows indicate alleles and circles indicate the partial and complete allele dropout due to degradation.

[E-T2] SNP



[E-T1] SNP



[F] SGM plus

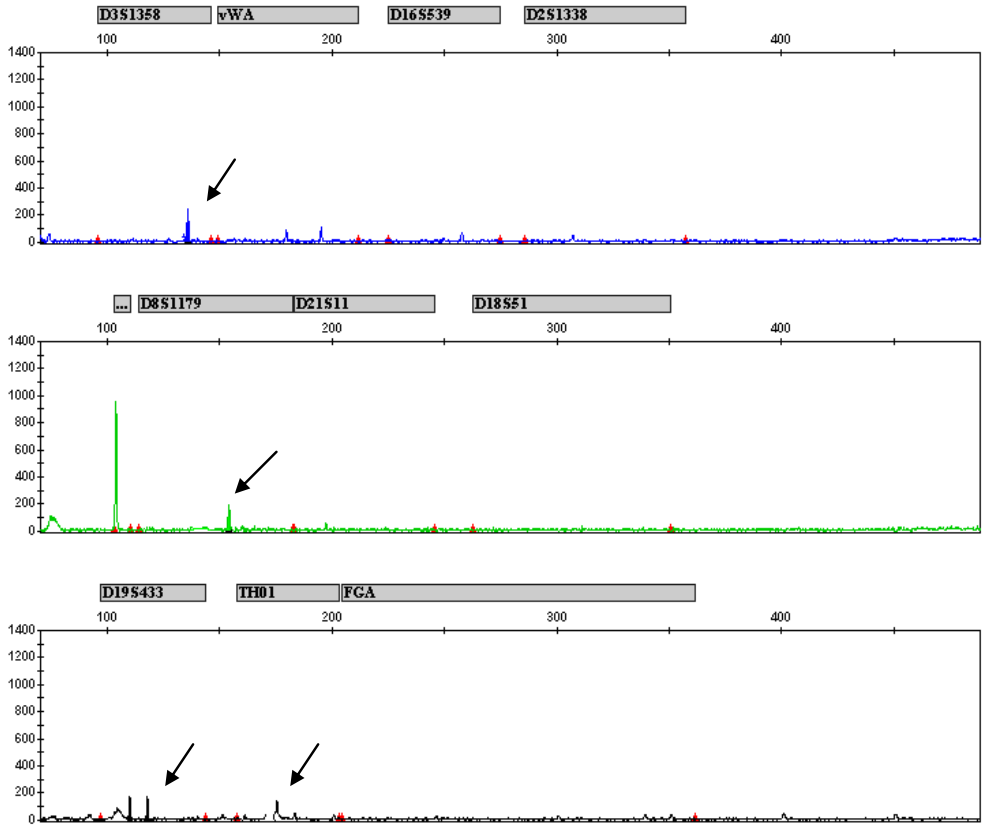
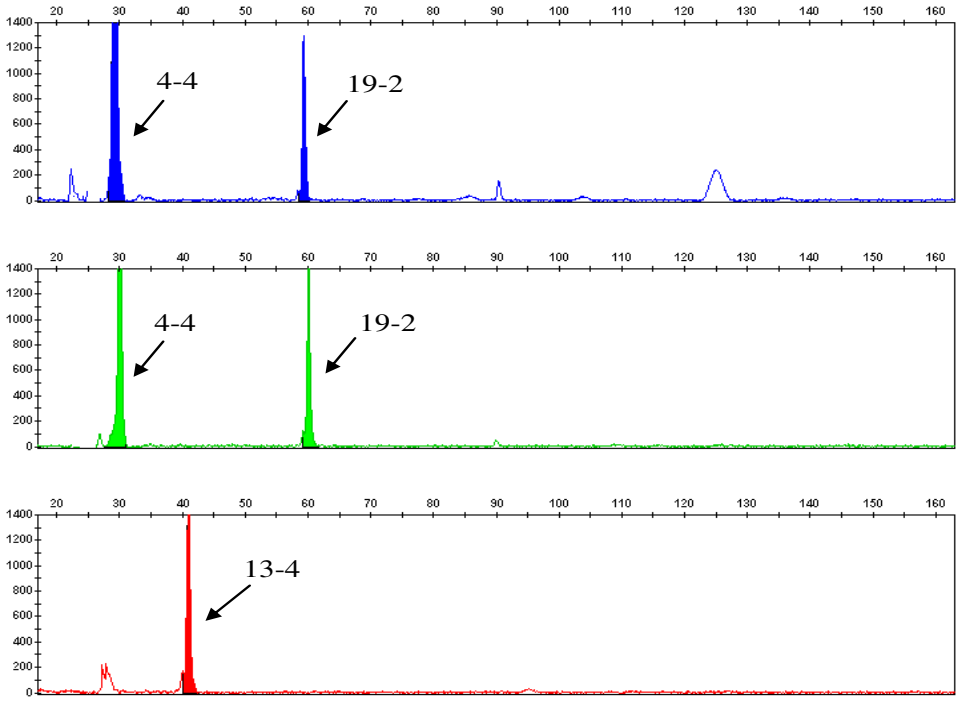
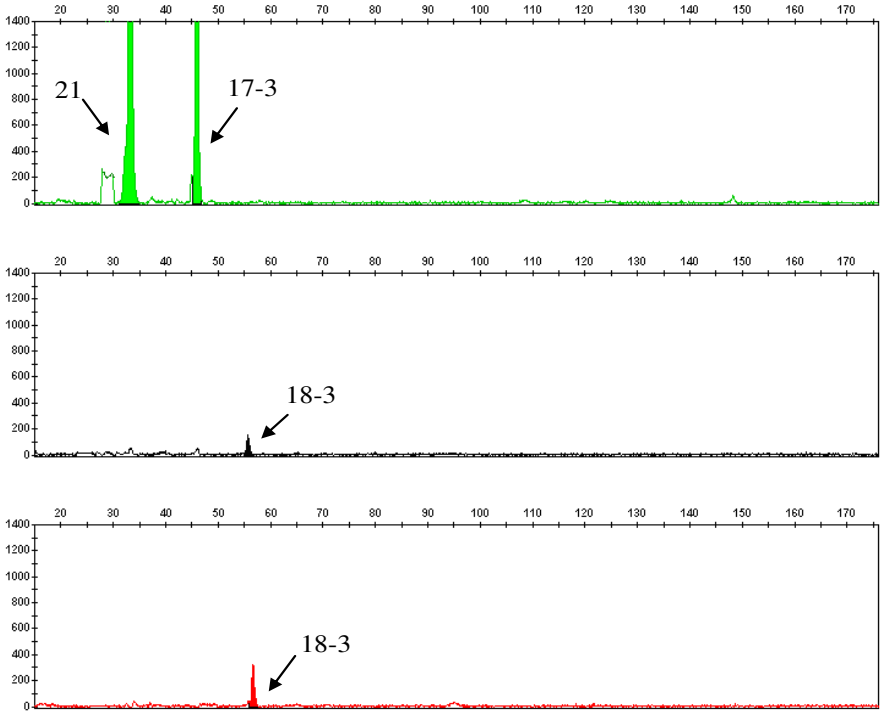


Figure 6.16 E-T1, F-T2 and G. Shown above are results for the samples at 6 days interval obtained from UAE September degradation. Electropherograms E-T1 and E-T2 are one of the repeats of triplex 1 and 2 of SNP genotyping have 100% profiles. F is the result for the same sample obtained from STR genotyping, which has 25% of alleles. Arrows indicate the allele peaks above 100 RFU and circle indicate the allele below 100 RFU.

[G-T1] SNP



[G-T2] SNP



[H] SGM plus

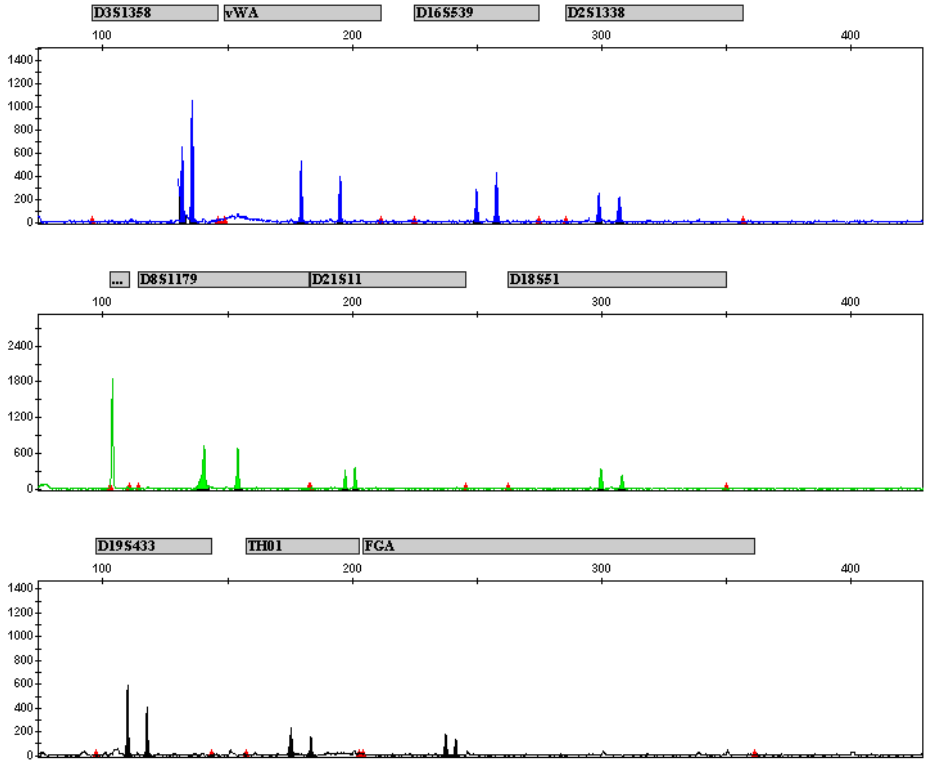


Figure 6.17 G-T1, G-T2 and H. Shown above are results for the samples at 6 day intervals obtained from UK August degradation. Electropherograms G-T1 and G-T2 are one of the repeat of SNP genotyping with 100% profiles. H is the result for the same sample obtained from STR genotyping with 100% profiles. Arrows indicate the alleles.

6.4.5. DNA Genotyping from DNase 1 Degradation

The samples (Section 6.3) were previously identified based on the profiles obtained from the genotyping of STRs. The 8 pp indicated a partial profile where 8 loci including the ameloginin were profiled, 4 pp; when 4 loci including the ameloginin were profiled and no profile when none of the loci were profiled.

The concentration of DNA in the samples (Table 6.5) were estimated using Quantifiler® Human DNA kit with the ABI 7500 real time PCR machine as described in Section 2.2.2.1.

Table 6.5. Indicated below are quantification results for DNA in DNase I samples. A partial profile is represented by pp and np represents no profile obtained in STRs.

Quantification values (ng/µl)			
Samples	8 pp	4 pp	np
Amount	0.74	0.37	0.29

6.4.5.1. SNP Profiling

Results are shown in (Table 6.6).

Table 6.6. Indicated below are SNP genotypes for samples treated with DNase 1 in both triplex. np represents no profile.

Triplex 1			
SNP code	AG 4-4	AG 19-2	CT 13-4
Samples.			
8 pp	AA	AG	CC
4 pp	AA	AG	CC
np	AA	AG	np
Triplex 2			
SNP code	AG 92 21	CT 119 18-3	AC 147 17-3
Samples			
8 pp	AG	TT	CC
4 pp	AG	TT	CC
np	AG	AG	CC

Samples 8pp and 4pp produced full loci with 100% allele profiles. Whilst sample np gave 83.3% with loss of one locus at SNP code 13-4 (Figure 6.18).

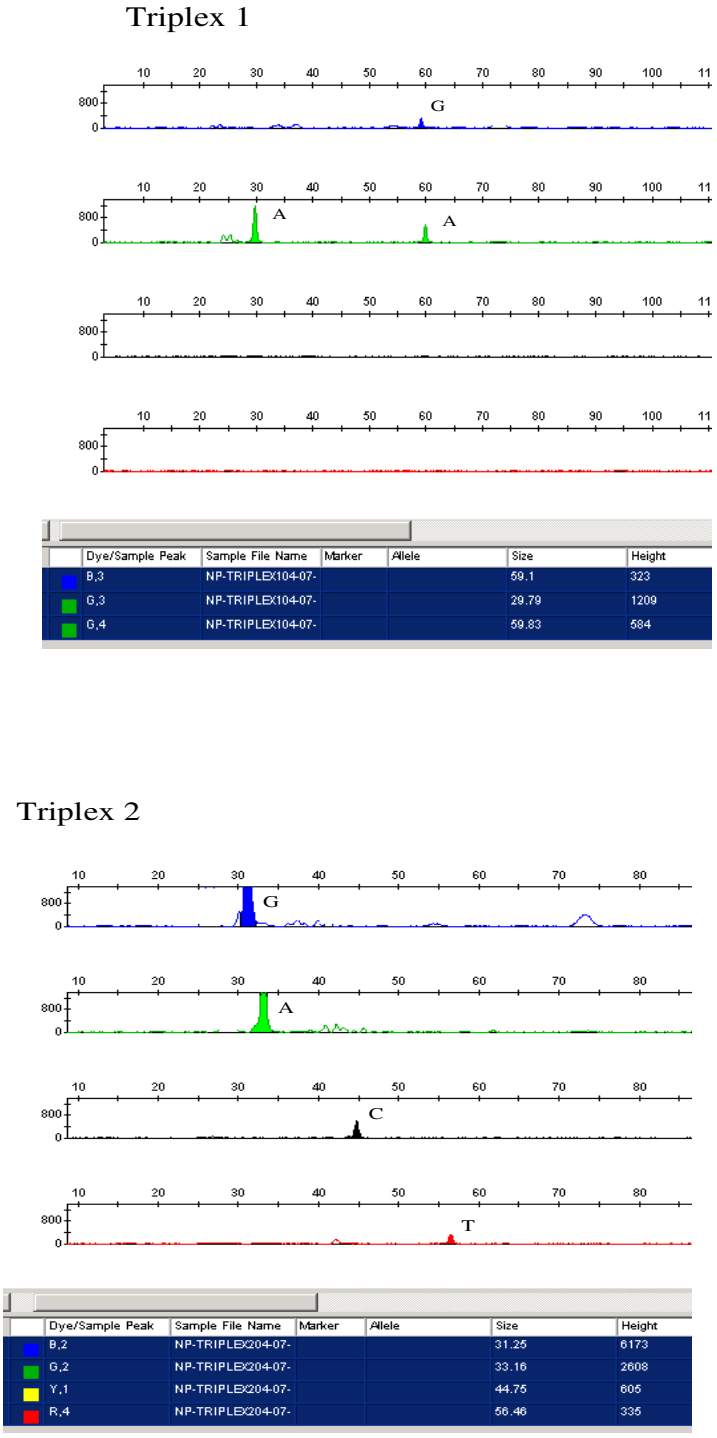


Figure 6.18. Triplex 1 and 2 electropherograms for sample NP at 100 RFU. 83.3% allele profiles was obtained due to locus 13-4 not profiling.

6.4.6. Application of developed SNP

The developed SNPs were also tested with forensic samples such as teeth extracted from human jaws.

The extraction procedure for all were carried out using Qiagen DNeasy® Blood and Tissue Kit as described in Section 2.6 (Chapter 2) and DNA was estimated using the Quantifiler® Human DNA kit with the ABI 7500 real time PCR machine as described in Section 2.2.2.1 (Table 6.7).

Table 6.7. Indicated below are results for DNA extracted from teeth samples. The quantification was carried out in duplicate for each sample. ud represent undetermined sample.

Quantification values (ng/µl)	
Samples	Amount
11	0.27
11	0.28
12	0.04
12	0.05
13	ud
13	0.02
14	0.03
14	0.05
15	0.01
15	0.02
16	0.01
16	0.02
17	0.58
17	0.55
18	0.19
18	0.14

6.4.6.1. SNP and STR Profiling

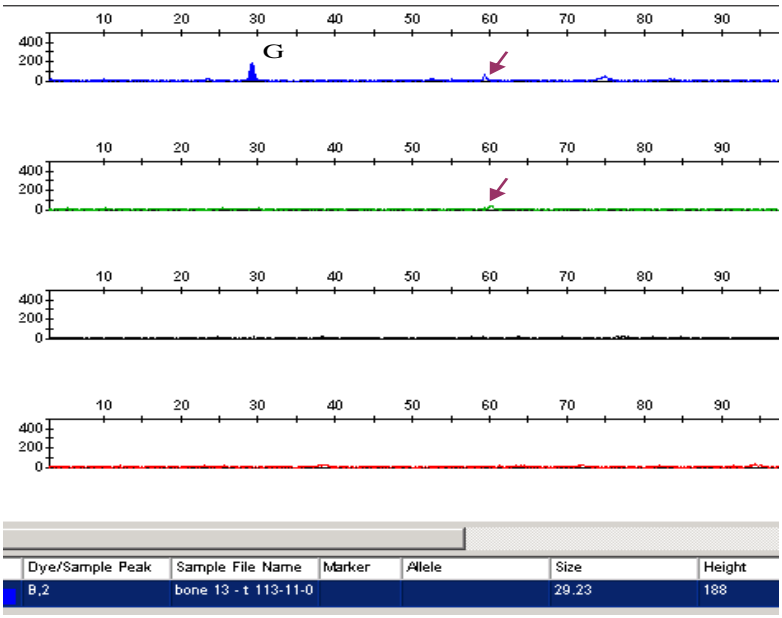
The SNP profiling results are showing in Table 6.8. Sample 13 and 14 produced 33.3% and 66.7% allele profiles, however, when the RFU thresholds was lowered to 50 (Sanchez et al., 2006) with modification, the allele profiles increased to 50% and 83%

respectively (Figure 6.19 to 6.20). This indicated that some of the allele profiles that were below the 100 RFU level were able to be pooled and identified. However, the lowest allele profiles were achieved for samples 15 and 16 with no profile suggesting that the samples were highly degraded. Maching allele profiles were observed between several of the samples. As an example: samples 11 and 12, 13 and 14, and 17 and 18, which were duplicate samples from the same individual, gave the same profiles. This provided additional confirmation for the genotyping results.

Table 6.8. Indicated below are SNP genotypes for teeth samples in both triplexes. np represents no profile.

RFU 100				RFU 50		
Triplex 1				**Triplex 1**		
SNP code	AG 4-4	AG 19-2	CT 13-4	AG 4-4	AG 19-2	CT 13-4
Samples						
11	AA	AG	CC			
12	AA	AG	CC			
13	G	np	np	G	G	np
14	G	AG	np	AG	AG	np
15	np	np	np			
16	np	np	np			
17	AA	AG	CC			
18	AA	AG	CC			
Triplex 2				**Triplex 2**		
SNP code	AG 21	CT 18-3	AC 17-3	AG 21	CT 18-3	AC 17-3
Samples						
11	AG	TT	CC			
12	AG	TT	CC			
13	GG	np	np	GG	np	A
14	GG	np	AA	GG	TT	AC
15	np	np	np			
16	np	np	np			
17	AG	TT	CC			
18	AG	TT	CC			

Triplex 1



Triplex 2

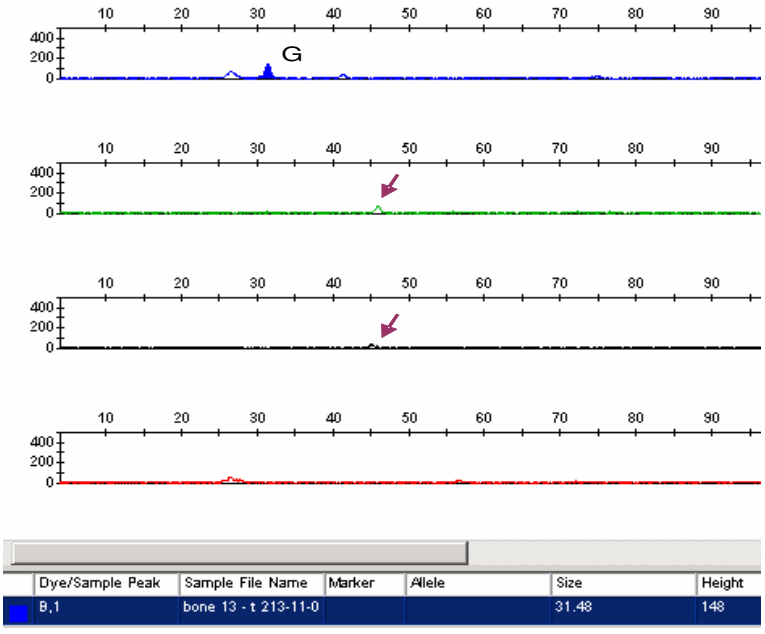
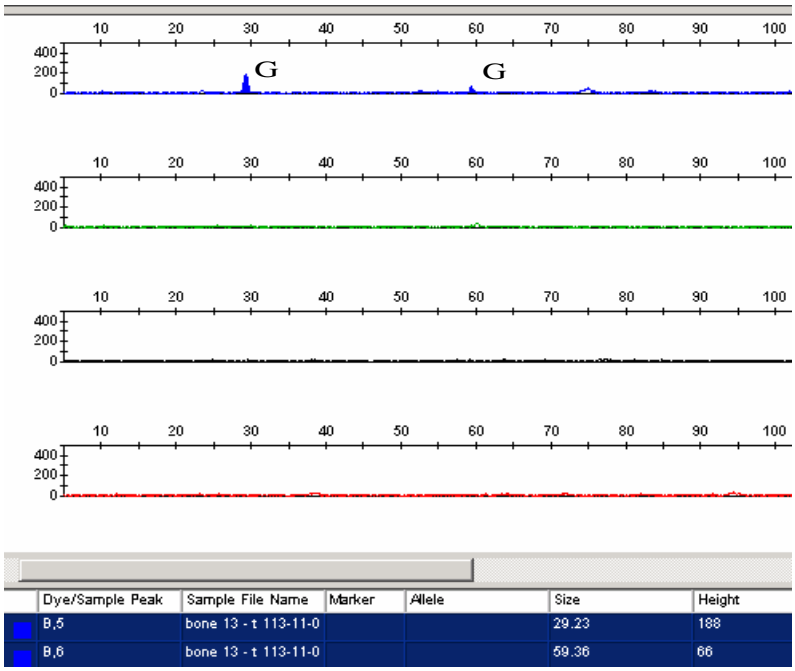


Figure 6.19. Shown above are Triplex 1 and 2 electropherograms for tooth sample 13 at 100 RFUs. Arrows represent alleles below 100 RFU.

Triplex 1



Triplex 2

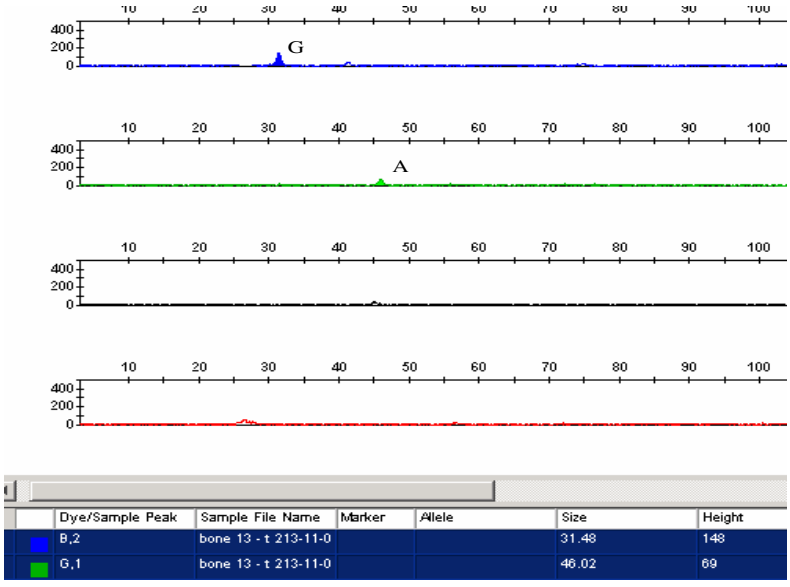


Figure 6.20. Shown above are electropherograms for Triplex 1 and 2 for tooth sample 13 with 50 RFUs defined as the cut off point. The additional G and A allele detected at height 66 and 69 RFUs respectively, increased the total profile 50%.

Due to the unavailability of STR reference profiles for the teeth samples, the calculation of the percentage of the allele profiles was based on the observation of the peak heights only. From these observations (based on tooth 14) the reference profiles had the following genotypes: D3S1358 is heterozygote; D16S539 is homozygote; D2S1338 is heterozygote; D8S1179 is heterozygote; D18S51 is homozygote; D19S433 is homozygote and THO1 is heterozygote.

The STR typing for sample 13 did not show any alleles, indicated a complete loss of loci (Figure 6.21). Twelve out of 20 alleles were partially profiled for sample 14, producing 60% of the total allele profile at 100 RFU threshold (Figure 6.22). Samples 15 and 16 both gave 0% profile. STR profiles for sample 11 and 12 were not available for the comparison.

Tooth 13

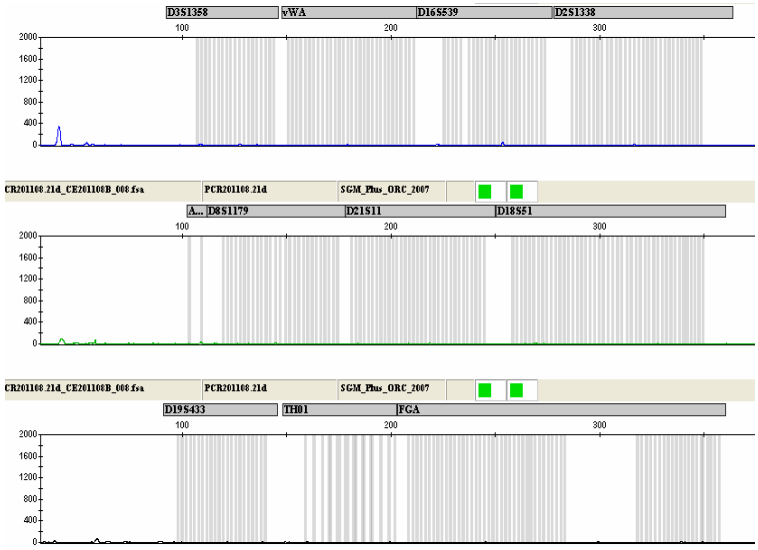


Figure 6.21. SGM plus® electropherogram for tooth sample 13. No alleles were observed.

Tooth 14

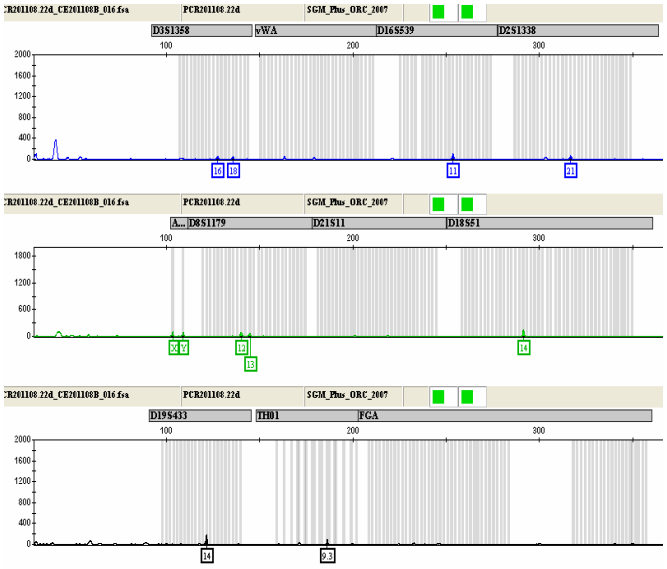


Figure 6.22. SGM plus® electropherogram for sample 14. There were 7 alleles (60%profiles).

6.5. Discussion

Saliva stains can be recovered from many objects left as evidence at scenes of crime including: cigarette butts, chewing gums, drinking containers and on a victims body as a result of rape cases (Bond et al., 2008). Alternatively, semen stain can be recovered from sexual assault scenes found on different items such as, clothes, bed sheets, body swabs and car seats. The successful profiling of such samples can be dependent upon the time taken to recover the stain coupled with the environmental temperatures. Therefore in order to obtain DNA genotyping from evidence, biological samples should be collected for analysis as quickly as possible.

Many factors influence the recovery of intact biological evidence from scenes of crime. Elements such as high temperature, humidity, and UV cause DNA degradation. Clearly, these elements are uncontrollable, if the evidence is found outdoors. This can lead to fragmentation of the DNA strands. The greater the exposure time to such insults, the more fragmentation is induced, and ultimately, the loss of genetic information that is useful for evidential purposes. However, the level of degradation also depends upon the type of the biological sample itself. Some samples tend to degraded faster than others. Saliva samples for example, because of the presence of other factors such as enzymes (amylase) and mouth microbial organisms, tend to enhance degradation more than blood and semen (Cotton et al., 2000).

Indoor Environmental studies

In this study, a comparison between SNP and STR genotyping was tested on artificially degraded semen and saliva samples. The ABI SNaPshot™ Triplex SNP set that was developed in this study was designed to amplify 90-147 bp of DNA template as part of a previous development. The STR genotyping was performed using the SGM plus® which generates amplicons ranging from approximately 100-360 bp. The performance

of SNP and STR analysis was greatly influenced by the degree of degradation. Semen samples were fully genotyped using SNPs and STRs: semen was less susceptible to degradation than saliva samples. Saliva DNA showed variation in degradation, producing both partial and a complete loss of loci.

Highly fragmented saliva DNA gave better results using SNP amplification because the small length of the SNP loci amplified more efficiently than the larger loci present in the STR system (Gill et al., 1998). Ultimately, a higher allele profile percentage was recovered in degraded samples using SNaPshot™ than SGM plus®, for example the saliva sample collected after 6 days incubation at 37 °C and 100% humidity gave a SNP profile of 72.2% and a STR profile of only 5%.

Outdoor Environmental Studies

Altough there have been many studies on environmental degradation of DNA samples, the study in this chapter focused upon the comparison of different climate conditions on saliva samples from different geographical places; the UAE and UK.

The DNA profiles obtained from the degradation in December/January at an average temperature of 22 °C (Met UAE) produced the most complete profiles in both systems (SNaPshot™ and SGM Plus®). The samples exposed to September with average temperature 34 °C (Met UAE), as expected, produced the lowest profiles in both systems. However, samples exposed to the UK climate of August with an average temperature of 16 °C resulted in fewer alleles profiling than the corresponding profiles obtained in December/ January (UAE). This clearly shows that lower temperature combined with high relative humidity such that observed in the UK in August are important.

Efficiency of obtaining DNA profiles

This chapter demonstrated that, the efficiency of obtaining DNA profiles did not only depend on the amount of starting template. Sufficient amounts of sample template can also result in low allele profiles if the samples are in a degraded state (Dixon et al., 2005b), such as the sample treated with DNase I for 10 minutes (8 pp). Although the amount of DNA was estimated to be more than 0.7 ng/ µl, only 70% of its profile was obtained with STRs profiling compared to 100% profile using SNP genotyping.

For a DNA template as little as 0.02 ng/µl, (bone sample number 13) 33.3% allele profiles were achieved at att loci, whilst genotyping with STRs failed to produce any profile for the same sample.

6.6. Conclusion

The SNP triplex set demonistred a higher level of sensitivity in obtaining genotypes from heavily degraded samples than SGM Plus®. This result, in addition to those of previous studies, represents the necessity to include SNPs as a method for genotyping for forensic samples. Also, from the observation of the performance of triplexes in this study, this indicates that the 66 singleplexes that were developed in Chapter 4 could be combinned into large multiplexes and used for the typing of degraded samples.

CHAPTER 7

GENERAL DISCUSSION and FUTURE WORK

7.1 General Discussion

The difficulty in analysing degraded samples has been the biggest challenge for obtaining DNA profiles using the STR method. An alternate method is therefore required to overcome the problem of typing such difficult samples. SNPs have shown promise and may become the future marker used for forensic applications (Esther et al., 2007). In this study, the results obtained from samples subjected to degradation and typed with the developed SNPs compare well to the results obtained using STRs, supporting the need for SNP typing of challenging samples.

The original goals of the Human Genome Project have been the construction of complete genetic and physical maps of the human genome (Sachidanandam et al., 2001). Since the completion of the human genome sequence, a comprehensive search for genetic influences in disease and individual genetic variation due to SNPs have been undertaken. According to the GenBank data base (db SNP) more than 14 million SNPs are submitted in the GenBank data up to date (06/10/2008).

The SNPs were primarily discovered by two projects: The SNP Consortium (TSC) and the International Human Genome Sequencing Consortium (HapMap), provides a public resource for defining haplotype variations across the genome, and help to identify biomedically important genes for diagnosis and therapy (Sachidanandam et al., 2001).

TSC contributed SNPs that were identified by shotgun sequencing of genomic fragments drawn from 24 ethnically diverse individuals, a representation of the human genome. This resulted in detecting more than one million SNPs with the sequence, physical and genetic maps of the human genome publicly available in GenBank (Sachidanandam et al., 2001).

HapMap project has looked at combinations of SNPs that are inherited together known as haplotypes to characterise linkage disequilibrium patterns across the genome to facilitate selection of most informative subsets of SNPs (Syvanen, 2005). These haplotypes enable geneticists to search for genes involved in diseases and for genome association studies. This required genotyping of 270 individuals from European, Sub-Saharan, Chinese and Japanese to generate allele frequencies. More than 4 million SNPs are validated by HapMap and made it publicly available in the GenBank data base.

The validated SNPs with allele frequencies and genotypic information that are presented in HapMap data base provide fundamental information for studying genetic variation in human population. However, the developed SNPs in this project were selected from Arab individuals rather than from HapMap data base. One important advantage of this selection was based on forensic application requirements to achieve high discrimination power and low match probability, therefore; SNPs with allele frequencies between 0.45 and 0.55 were selected and in turn high heterozygosity of 0.47 were achieved. Also, the SNPs with minor allele frequency provide little information for association and linkage study: minor alleles frequencies that are observed in one population can disappear in other populations (Goddard et al., 2000).

The recent developments in microarray technologies for SNP screening provide speed, efficiency and throughput. The benefit of using the Affymetrix® microarray method for screening the SNPs from the whole genome was achieved (Chapter 2). It allowed the identification of SNPs from autosomal chromosomes from United Arab Emirates and Kuwait Arab samples. The method requires high amount of starting samples for the screening (Matsuzaki et al., 2004), and is therefore of little value when typing forensic samples. It has proven to be successful for our needs in selecting polymorphic SNPs from this particular population.

The main objective of this study was to develop SNPs that can be useful for increasing the allele profiles for the identification of degraded DNA in forensic samples. In this project 66 SNPs were developed in order to meet the requirement of forensic applications (Chapter 5). SNaPshot™ is a simple convenient method that uses an instrument, of which there are several possible models, and which is available in most forensic laboratories: the ABI Prism® Genetic Analyzer (Applied Biosystems). SNP genotyping using this method provided valuable information that enabled samples to be analysed quickly.

All the 66 SNPs conformed to Hardy-Weinberg expectation, did not show any linkage disequilibrium and had high heterozygosity levels when compared with the existing 52 SNPs developed by Sanchez et al. (2006). The sensitivity study showed profiles were possible from as little as 100 pg DNA template with the optimum amount of 300 pg giving accurate results.

The triplexes developed as representative of the 66 SNPs were shown to be useful when analysing degraded samples. Artificially degraded samples under different environmental conditions showed fuller profiles when typed with SNPs compared to STRs. The amplicon of the SNPs, between 90 and 147 bp, showed more resistance to degradation than the larger STRs length (100-360 bp). The SNP genotypes were reproducible among different sample types and samples degraded over different time periods and conditions.

In addition to the usefulness of SNPs in typing artificially degraded samples, these SNPs were also tested in samples obtained from different scenarios. It was demonstrated in this project that these SNPs will be useful for the analysis of human remains such as teeth, common evidence found in mass disasters. Also, the small size of these SNPs gives them greater potential in producing allele profiles from enzymatically

degraded samples which produced partial profiles by STRs such as samples treated by DNase I.

In conclusion, the 66 candidate SNPs developed in this study were shown to be a new tool for Arab populations, recovering useful genetic information for forensic identification on degraded samples. This project supports the use of SNPs as forensic markers for degraded samples.

7.2 Future Work

The developed 50 autosomal SNPs have met the expectation of the project aim which was to introduce new forensic markers capable of increasing the power of identification for degraded samples. However, the strength of genotyping degraded samples can be improved markedly by using larger SNP multiplexes. Profiling of the degraded samples by the triplex was very promising, and by increasing the combination of both PCR and SBE primers will increase the number of loci to be profiled which in turn will increase the power of identification of samples. Moreover, the amount of starting sample will be reduced. Rather than needing the samples for two separate triplexes, a larger multiplex will only require one DNA template. This is advantageous for most forensic samples. No doubt in the future, technology will improve allowing more SNPs to be multiplexed in one tube. The existing method developed by Sanchez et al. (2006) enabled a maximum of 29 autosomal SNPs to be multiplexed in a single tube.

The result of genotyping SNPs using SNaPshot™ method showed a feature that needs to be considered in the future. The dyes that are used in the SBE method have disadvantages in some loci, especially when genotyping highly degraded samples. The red and yellow dyes that are incorporated to ddTTP and ddCTP respectively show very low signal, about 1/3 the signal obtained from ddGTP and 1/2 the signal obtained from ddATP (Sanchez et al., 2006). This variation in signal affected the allele calls as the first

loci that were below the RFUs threshold were found to be those incorporated with the yellow and red dye whilst the blue and green loci exhibited relatively high signals. It will be very helpful if the SBE method used in the SNaPshot™ analysis could improve this signal imbalance in future. This will increase the rate of allele calls better than the existing SBE dyes.

To date, the SNP markers have only been tested in an Arabic population. Further population studies, on diverse population groups will enable an assessment to be made as to how versatile the SNPs will be: many are likely to show similar allele frequencies in different populations; however, some may prove to be highly polymorphic only in the Arabic population.

Finally, in the future it will be very useful for UAE forensic laboratories to use SNPs as forensic markers. The harsh weather conditions in the UAE are observed on the incomplete recovery of genetic information in most samples, especially when temperatures and humidity exceed 45 °C and 80% respectively in most summer seasons.

REFERENCES

AL-GHUNAIM, A. (2007) *Selected Research from Kuwait History* Centre for research and studies on Kuwait. CRSK Press pp10-20.

ALTUKHOV, Y. P. & SALMENKOVA, E. A. (2002) DNA polymorphism in population genetics. *Russian Journal of Genetics,* 38**,** 989-1008.

ANDREASSON, H., NILSSON, M., BUDOWLE, B., LUNDBERG, H. & ALLEN, M. (2006) Nuclear and mitochondrial DNA quantification of various forensic materials. *Forensic Science International***,** 1-9.

BALTIMORE, D. (2001) Our genome unveiled. *Nature,* 409**,** 814-816.

BECKMANN, J. S. & WEBER, J. L. (1992) Survey of human and rat microsatellites. *Genomics,* 12**,** 627-631.

BENDER, K., FARFAN, M. J. & SCHNEIDER, P. M. (2004) Preparation of degraded human DNA under controlled conditions. *Forensic Science International,* 139**,** 135-140.

BIOSYSTEMS, A. (2000) ABI PRISM® SNaPshot™ multiplex kit protocol.

BOND, J. W. & HAMMOND, C. (2008) The value of DNA materials recovered from crime scenes. *Journal of Forensic Sciences,* 53**,** 797-801.

BROOKES, A. J. (1999) The essence of SNPs. *Gene,* 234**,** 177-186.

BUDIMLIJA, Z. M., PRINZ, M. K., MUNDORFF, A. Z., WIERSEMA, J., BARTELINK, E., MACKINNON, G., NAZZARUOLO, B. L., ESTACIO, S. M., HENNESSEY, M. J. & SHALER, R. C. (2003) World trade center human identification project: experiences with individual body identification cases. *Croatian Medical Journal,* 44**,** 259- 263.

BUDOWEL, B. (2004) SNP typing strategies. *Forensic Science International,* 146S**,** S139-S142.

BUDOWELE, B., BIEBER, F. R. & EISENBERG, A. J. (2005) Forensic aspects of mass disaster: strategic considerations for DNA based- human identification. *Legal Medicine,* 7**,** 230- 243.

BUDOWLE, B., HOBSON, D. L., SMERICK, J. B. & SMITH, J. A. L. (2001) Low copy number - consideration and caution. *laboratory Division of the Federal Bureau of Investigation,* 01-26.

BUTLER, J. M. (2006) Genetics and genomics of core short tandem repeat loci used in human identity testing. *Journal of Forensic Science,* 51**,** 253-265.

BUTLER, J. M. (2007) Short tandem repeat typing technologies used in human identity testig. *BioTechniques,* 43**,** Sii-Sv.

BUTLER, J. M., BUEL, E., CRIVELLENTE, F. & M$_C$CORD, B. R. (2004) Forensic DNA typing by capillary electrophoresis using the ABI prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis,* 25**,** 1397-1412.

BUTLER, J. M., COBLE, M. D. & VALLONE, P. M. (2007) STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic Science, Medicine, and Pathology,* 3**,** 200-205.

BUTLER, J. M., SHEN, Y. & MCCORD, B. R. (2003) The developement of reduced size STR amplifications as tools for analysis of degraded DNA. *Journal of Forensic Science,* 48**,** 1054-1064.

CHEN, X., LIVAK, K. J. & KWOK, P.-Y. (1998) A homogeneous, ligase- mediated DNA diagnostic test. *Genome Research,* 8**,** 549- 556.

CLAYTON, T. M., WHITAKER, J. P., FISHER, D. L., LEE, D. A., HOLLAND, M. M., WEEDN, V. W., MAGUIRE, C. N., DIZINNO, J. A., KIMPTON, C. P. & GILL, P. (1995) Further validation of a quadruplex STR DNA typing system: a collaborative effort to identify victims of a mass disaster. *Forensic Science International,* 76**,** 17-25.

CLAYTON, T. M., WHITAKER, J. P., SPARKES, R. & GILL, P. (1998) Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Science International,* 91**,** 55-70.

COBLE, M. D. & BUTLER, J. M. (2005) Characterization of new MiniSTR loci to aid analysis of degraded DNA. *Forensic Science,* 50**,** 1-11.

COLLINS, F. S., LANDER, E. S., ROGERS, J. & WATERSTON, R. H. (2004) Finishing the euchromatic sequence of the human genome. *Nature,* 431**,** 931-938.

COOPER, D. N., SMITH, B. A., COOKE, H. J., NIEMANN, S. & SCHMIDTKE, J. (1985) An estimate of unique DNA sequence hetrozygosity in the human genome. *Human Genetics,* 69**,** 201- 205.

COTTON, E. A., ALLSOP, R. F., GUEST, J. L., FRAZIER, R. R. E., KOUMI, P., CALLOW, I. P., SEAGER, A. & SPARKES, R. L. (2000) Validation of the AMPFlSTR® SGM Plus(TM) system for use in forensic casework. *Forensic Science International,* 112**,** 151-161.

DIEFFENBACH, C. W. & DVEKSLER, G. S. (2003) *PCR Primer: A Laboratory Manual* New York, Spring Harbor Laboratory Press.

DIVNE, A. M. & ALLEN, M. (2005) A DNA microarray system for forensic SNP analysis. *Forensic Science International,* 154**,** 111-121.

DIXON, L. A., DOBBINS, A. E., PULKER, H. K., BUTLER, J. M., VALLONE, P. M., COBLE, M. D., PARSON, W., BERGER, B., GURBWIESER, P., MOGENSEN, H. S., MORLING, N., NIELSEN, K., SANCHEZ, J. J., PETKOVSKI, E., CARRACEDO, A., SANCHEZ-DIZ, P., RAMOS-LUIS, E., BRION, M., IRWIN, J. A., JUST, R. S., LOREILLE, O., PARSONS, T. J., SYNDERCOMBE-COURT, D., SCHMITTER, H., STRADMANN-BELLINGHAUSEN, B., BENDER, K. & GILL, P. (2005a) Analysis of arificially degraded DNA using STRs and SNPs- results of a collaborative European (EDNAP) exercise. *Forensic Science International,* 164**,** 33-44.

DIXON, L. A., MURRAY, C. M., ARCHER, E. J., DOBBINS, A. E., KOUMI, P. & GILL, P. (2005b) Validation of a 21- locus autosomal SNP multiplex for forensic identification purposes. *Forensic Science International,* 154**,** 62-77.

FONDEVILA, M., PHILLIPS, C., NAVERAN, N., FERNANDEZ, L., CEREZO, M., SALAS, A., CARRACEDO, A. & LAREU, M. V. (2008) Case report: Identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extraced from a decomposed and charred femur. *Forensic Science International: Genetics,* 2**,** 212-218.

FORAN, D. R. (2006) Relative degradation of nuclear and Mitochondrial DNA: an experimental approach. *Journal Forensic Science,* 51**,** 766-770.

GIBSON, N. J. (2006) The use of real-time PCR methods in DNA sequence variation analysis. *Clinica Chemica Acta,* 363**,** 32-47.

GILL, P. (2001) Application of low copy number DNA profiling. *Croatian Medical Journal,* 42**,** 229-232.

GILL, P. (2002) Role of short tandem repeat DNA in forensic casework in the UK-past, present, and future prespectives. *BioTechniques,* 32**,** 366-385.

GILL, P., A, C. B., BRINKMANNC, B., BUDOWLED, B., CARRACEDOE, A., JOBLINGF, M. A., KNIJFFG, P. D., KAYSERH, M., KRAWCZAKI, M., MAYRJ, W. R., MORLINGK, N., OLAISENL, B., PASCALIM, V., PRINZN, M., ROEWERO, L., SCHNEIDERP, P. M., SAJANTILAQ, A. & TYLER-SMITHR, C. (2001) DNA Commission of the international society of forensic genetics: recommendations on forensic analysis using Y- chromosome STRs. *forensic science international,* 124**,** 5-10.

GILL, P., FOREMANB, L., BUCKLETONC, J. S., TRIGGSD, C. M. & ALLENA, H. (2003) A comparison of adjustment methods to test the robustness of an STR DNA database comprised of 24 European populations. *Forensic Science International,* 131**,** 184-196.

GILL, P., SPARKES, R., PINCHIN, R., CLAYTON, T., WHITAKER, J. & BUCKLETON, J. (1998) Interpreting simple STR mixtures using allele peak areas. *Forensic Science International,* 91**,** 41-53.

GOTO, S., TAKAHASHI, A., KAMISANGO, K. & MATSUBARA, K. (2002) Single nucleotide polymorphism analysis by hybridization protection assay on solid support. *Analytical Biochemistry,* 307**,** 25-32.

GRAY, I. C., CAMPBELL, D. A. & SPURR, N. K. (2000) Single nucleotide polymorphisms as tools in human genetics. *Human Molecular Genetics,* 9**,** 2403-2408.

HAFF, L. A. & SMIRNOV, I. P. (1997) Single- nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. *Genome Research,* 7**,** 378-388.

HALIM, N. S. & ALTSBULER, D. (2001) SNP maps and the promis of pharmacogenomics. *New England Biolabs,* 11**,** 1-16.

HALL, A. & BALLATYNE, J. (2004) Characterization of UVC-induced DNA damage in blood stains: forensic implications. *Analytical and Bioanalytical Chemistry,* 380**,** 72-83.

HOLLAND, M. & PARSONS, T. (1999) Mitochondrial DNA sequence analysis-validation and use for forensic casework. *Forensic Science Review,* 11**,** 22-50.

INAGAKI, S., YAMAMOTO, Y., DIO, Y., TAKATA, T., ISHIKAWA, T., IMABAYASHI, K., YOSHITOME, K., MIYAISHI, S. & ISHIZU, H. (2004) A New 39 plex analysis method for SNPs including 15 blood group loci. *Forensic Science International,* 144**,** 45-57.

INAGAKI, S., YAMAMOTOA, Y., DOIA, Y., TAKATAA, T., ISHIKAWAA, T., YOSHITOMEA, K., MIYAISHIA, S. & ISHIZUA, H. (2002) Typing of Y chromosome single nucleotide polymorphisms in a Japanese population by a multiplexed single nucleotide primer extension reaction *Legal Medicine,* 4**,** 202-206.

JEFFREYS, A. J., MACLEOD, A., TAMAKI, K., NEIL, D. L. & MONCKTON, D. G. (1991) Minisatellite repeat coding as a digital approach to DNA typing. *Nature,* 354**,** 204-209.

JENKINS, S. & GIBSON, N. (2002) High - throughput SNP genotyping. *Comparative and Functional Genomics,* 3**,** 57-66.

JOBLING, M. A. (2001) Y-chromosomal SNP haplotype diversity in forensic analysis. *Forensic Science International,* 118**,** 158-162.

JOBLING, M. A. & GILL, P. (2004) Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics,* 5**,** 739-751.

KADYROVA, F. A., GENSCHELA, J., FANGA, Y., PENLANDB, E., EDELMANNC, W. & MODRICH, P. (2009) A possible mechanism for exonuclease 1-independent eukaryotic mismatch repair. *PNAS(Proceeding of the National Academy of Science of the United States of America),* 106**,** 8495-8500.

KASHAYAB, V. K., SITALAXIMI, T., CHATTOPADHYAY, P. & TRIVEDI, R. (2004) DNA profiling technologies in forensic analysis. *International Journal of Human Genetic,* 4**,** 11-30.

KAYSER, M. (2007) Uni-parental markers in human identity testing including forensic DNA analysis. *BioTechniques,* 43**,** Sxv-Sxxi.

KIDD, K. K., PAKSTIS, A. J., SPEED, W. C., GRIGORENKO, E. L., KAJUNA, S. L. B., KAROMA, N. J., KUNGULILO, S., KIM, J. J., LU, R.-B., ODUNSI, A., OKONOFUA, F., PARNAS, J., SCHULZ, L. O., ZHUKOVA, O. V. & KIDD, J. R. (2006) Developing a SNP Panel for Forensic Identification of Individuals. *Forensic Science International,* 164**,** 20-32.

KLINE, M. C., BUEWER, D. L., REDMAN, J. W. & BUTLER, J. M. (2005) Results from the NIST 2004 DNA quantitation study. *Journal of Forensic Sciences,* 50**,** 571-578.

KLOOSTERMAN, A. D. & KERSBERGEN, P. (2003) Efficacy and limits of genotyping low copy number DNA samples by multiplex PCR of STR loci *International Congress Series,* 1239**,** 795-798.

KRAWCZAK, M. & SCHMIDTKE, J. (1994) *DNA Fingerprinting,* Oxford, Bios Scientific Publishers Ltd.

KRENKE, B. E., TEREBA, A., ANDERSON, S. J., BUEL, E., CULHANE, S., FINIS, C. J., TOMSEY, C. S., ZACHETTI, J. M. & SPRECHER, C. J. (2002) Validation of a 16-locus fluorescent multiplex system. *Journal of Forensic Sciences,* 47**,** 1-13.

LADD, C., LEE, H. C., YANG, N. & BIEBER, F. R. (2001) Interpretation of complex forensic DNA mixtures. *Croatian Medical Journal,* 42**,** 244-246.

LANDEGREN, U., KAISER, R., SANDERS, J. & HOOD, L. (1988) A ligase-mediated gene detection technique. *Science,* 241**,** 1077-1080.

LANDEGREN, U., NILSSON, M. & KWOK, P. Y. (1998) Reading bits of genetic information: methods for single nucleotide polymorphism analysis. *Genomic Research,* 8**,** 769-776.

LEWIN, B. (Ed.) (2004) *GENES VIII,* Pearson Prentice Hall.

LI, S., MA, L., LI, H., VANG, S., HU, Y., BOLUND, L. & WANG, J. (2006) Snap: an integrated SNP annotation platform *Nucleic Acids Research,* 00**,** D1-D4.

LINDBLAD-TOH, K., WINCHESTER, E., DALY, M. J., WANG, D. G., HIRSCHHORN, J. N., LAVIOLETTE, J.-P., ARDLIE, K., REICH, D. E., ROBINSON, E., SKLAR, P., SHAH, N., THOMAS, D., FAN, J.-B., GINGERAS, T., WARRINGTON, J., PATIL, N., HUDSON, T. J. & LANDER, E. S. (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nature Genetics,* 24**,** 381-386.

LIU, G., LORAINE, A. E., SHIGETA, R., CLINE, M., CHENG, J., VALMEEKAM, V., SUN, S., KULP, D. & SIANI-ROSE, M. A. (2003) NetAffx: Affymetrix probesets and annotations. *Nucleic Acids Research,* 31**,** 82-86.

LIVAK, K. J. (1999) Allelic discrimination using fluorogenic probs and the 5` nuclease assay. *Genetic Analysis: Biomolecular Engineering,* 14**,** 143- 149.

LOREILLE, O. M., DIEGOLI, T. M., IRWIN, J. A., COBLE, M. D. & PARSONS, T. J. (2007) High efficiency DNA extraction from bone by total demineralization. *Forensic Science International: Genetics,* 1**,** 191-195.

LU, M., KNICKERBOCKER, T., CAI, W., YANG, W., HAMERS, R. J. & SMITH, L. M. (2004) Invasive cleavage reactions on DNA-modified diamond surfaces. *Biopolymers,* 73**,** 606-613.

MATSUZAKI, H., LOI, H., DONG, S., TSAI, Y.-Y., FANG, J., LAW, J., XIAOJUN, D., LIU, W.-M., YANG, G., LIU, G., HUANG, J., KENNEDY, G. C., RYDER, T. B., MARCUS, G. A., WALSH, P. S., SHRIVER, M. D., PUCK, J. M., JONES, K. W. & MEI, R. (2004) Parallel genotyping of Over 10,1000 SNPs using a one -primer assay on a high-density oligonucleotide array. *Genome Research,* 14**,** 414-425.

MCGUIGAN, F. E. A. & RALSTON, S. H. (2002) Single nucleotide polymorphism detection: allelic discrimination using TagMan. *Psychiatric Genetics,* 12**,** 133-136.

METZKER, M. L. (2005) Emerging technologies in DNA sequencing. *Genome Research,* 15**,** 1767-1776.

MULERO, J. J., CHANG, C. W., LAGACE, R. E., WANG, D. Y., BAS, J. L., MCMAHON, T. P. & HENNESSY, L. K. (2008) Development and validaton of the AmpF*l*STR MiniFiler PCR amplification kit: A MiniSTR multiplex for the analysis of degraded and/or PCR inhibited DNA. *Journal Forensic Science,* 53**,** 838-852.

MULLIS, K., FALOONA, F., SCHARF, S., SAIKI, R., HORN, G. & ERLICH, H. (1986) Specific enzymatic amplification of DNA in Vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology,* 51**,** 263-273.

MUSGRAVE-BROWN, E., BALLARD, D., ÁLVAREZ, M. F., FANG, R., HARRISON, C., PHILLIPS, C., PRASAD, Y., REY, B. S., THACKER, C., WILUHN, J., CARRACEDO, A., SCHNEIDER, P. M., COURT, D. S. & CONSORTIUM, T. S. (2008) Forensic validation of the Genplex SNP typing system—Results of an inter-laboratory study *Forensic Science International: Genetics,* 1**,** 389-393.

NEAVES, K. J., COOPER, L. P., WHITE, J. H., CARNALLY, S. M., DRYDEN, D. T. F., EDWARDSON, J. M. & HENDERSON, R. M. (2009) Atomic force microscopy of the EcoKI Type I DNA restriction enzyme bound to DNA shows enzyme dimerization and DNA looping. *Nucleic Acids Research,* 37**,** 2053-2063.

NIEDERSTÄTTER, H., COBLE, M. D., GRUBWIESER, P., PARSONS, T. J. & PARSON, W. (2006) Characterization of mtDNA SNP typing and mixture ratio assessment with simultaneous real-time PCR quantification of both allelic states. *International Journal of Legal Medicine,* 120**,** 18-23.

OLIVER, D. H., THOMPSON, R. E., GRIFFIN, C. A. & ESHLEMAN, J. R. (2000) Use of single nucleotide polymorphisms (SNP) and real time polymeraze chain reaction for bone marrow engraftment analysis. *Journal of Molecular Diagnostics,* 2**,** 202-208.

OLIVIER, M., CHUANG, L. M., CHANG, M. S., CHEN, Y. T., PEI, D., RANADE, K., WITTE, A. D., ALLEN, J., TRAN, N., CURB, D., PRATT, R., NEEFS, H., INDIG, M. D. A., LAW, S., NERI, B., WANG, L. & COX, D. R. (2002) High-throughput genotyping of single nucleotide polymorphisms using new biplex invader technology. *Nucleic Acids Research,* 30**,** 1-8.

PÄÄBO, S., POINAR, H., SERRE, D., JAENICKE-DESPRÉS, V., HEBLER, J., ROHLAND, N., KUCH, M., KRAUSE, J., VIGILANT, L. & HOFREITER, M. (2004) Genetic analyses from ancient DNA. *Annual Review of Genetics,* 38**,** 645-679.

PANG, B. C. M. & CHEUNG, B. K. K. (2007) One-step generation of degraded DNA by UV irradiation. *Analytical Biochemistry,* 360**,** 163-165.

PATZELT, D. (2004) History of forensic serology and molecular genetics in the sphere of activity of the German Society for Forensic Medicine. *Forensic Science International,* 144**,** 185-191.

PEREZ-ARNAIZ, P., LAZARO, J. M., SALAS, M. & VEGA, M. D. (2006) Involvement of φ29 DNA polymeraze thumb subdomain in the proper coordination of synthesis and degradation during DNA replication. *Nucleic Acids Research,* 34**,** 3107-3115.

PHILLIPS, C., FANG, R., BALLARD, D., FONDEVILA, M., HARRISON, C., HYLAND, F., MUSGRAVE-BROWN, E., PROFF, C., RAMOS-LUIS, E., SOBRINO, B., CARRACEDO, A., FURTADO, M. R., COURT, D. S., SCHNEIDER, P. M. & CONSORTIUM, T. S. (2007) Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel *Forensic Science International: Genetics,* 1**,** 180-185.

PHILLIPS, C., LAREU, M., SANCHEZ, J., BRION, M., SOBRINO, B., MORLING, N., SCHNEIDER, P., SYNDERCOMBE, D. & CARRACEDO, A. (2004) Selecting single nucleotide polymorphisms for forensic applications. *International Congress Series,* 1261**,** 18-20.

POGOZELSKI, W. K. & TULLIUS, T. D. (1998) Oxidative strand scission of nucleic acids: routes initiated by hydrogen abstraction from the sugar moiety. *Chemical Reviews,* 98**,** 1089-1107.

QIAGEN® (2005) REPLI-g handbook. Qiagen.

QIAGEN® (2006) DNeasy® Blood & Tissue Handbook.

QIAGEN® (2007) QIAamp® DNA Investigator Handbook

RAO, K. V. N., STEVENS, P. W., HALL, J. G., LYAMICHEV, V., NERI, B. P. & KELSO, D. M. (2003) Genotyping single nucleotide polymorphisms directly from genomic DNA by invesive cleavage reaction on microspheres. *Nucleic Acids Research,* 32**,** 1-8.

REICH, D. E., CARGILL, M., BOLK, S., IRELAND, J., SABETI, P. C., RICHTER1, D. J., LAVERY, T., KOUYOUMJIAN, R., FARHADIAN, S. F., WARD, R. LANDER, E. S. (2001) Linkage disequilibrium in the human genome. *Nature,* 411**,** 199-204.

RICE, W. R. (1989) Analyzing Tables of Statistical Tests. *Evolution,* 43**,** 223-225.

RONAGHI, M. (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Research,* 11**,** 3-11.

SACHIDANANDAM, R., WEISSMAN, D., SCHMIDT, S. C., KAKOL, J. M., STEIN, L. D., MARTH, G., SHERRY, S., C.MULLIKIN, J., MORTIMORE, B. R. J., WILLEY, D. L., HUNT, S. E., COLE, C. G., COGGILL, P. C., RICE, C. M., NING, Z., ROGERS, J., BENTLEY, D. R., KWOK, P.-Y., MARDIS, E. R., YEH, R. T., SCHULTZ, B., COOK, L., DAVENPORT, R., DANTE, M., FULTON, L. & HILLIER, L. (2001) A Map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature,* 409**,** 928-933.

SAIKI, R. K., SCHARF, S., FALOONA, F., MULLIS, K. B., HORN, G. T., ERLICH, H. A. & ARNHEIM, N. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science,* 230**,** 1350-1354.

SALAS, A., BANDELT, H.-J., MACAULAY, V. & RICHARDS, M. B. (2007) Phylogeographic investigations: The role of trees in forensic genetics. *Forensic Science International,* 168**,** 1-13.

SANCHEZ, J. J., BORSTING, C., HALLENBERG, C., BUCHARD, A., HERNANDEZ, A. & MORLING, N. (2003) Multiplex PCR and minisequencing of SNPs- a model with 35 Y chromosome SNPs. *Forensic Science International,* 137**,** 74-84.

SANCHEZ, J. J. & ENDICOTT, P. (2006) Developing multiplexed SNP assays with special reference to degraded templates. *Nature Protocols,* 1**,** 1370-1378.

SANCHEZ, J. J., PHILLIPS, C., BØRSTING, C., BALOGH, K., BOGUS, M., FONDEVILA, M., HARRISON, C. D., MUSGRAVE-BROWN, E., SALAS, A., SYNDERCOMBE-COURT, D., SCHNEIDER, P. M., CARRACEDO, A. & MORLING, N. (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis,* 27**,** 1713-1724.

SCHNEIDER, P. M., BALOGH, K., NAVERAN, N., BOGUS, M., BENDER, K., LAREU, M. & CALLEGARO, A. (2004) Whole genome amplification- the solution for a common problem in forensic casework? *International Congress Series,* 1216**,** 24-26.

SCHOSKE, R., VALLONE, P. M., RUITBERG, C. M. & BUTLER, J. M. (2003) Multiplex PCR design strategy used for the simultaneous amplification of 10 Y chromosome short tandem repeat (STR) loci. *Analytical & Bioanalytical Chemistry,* 375**,** 333-343.

SOBRINO, B., BRION, M. & CARRACEDO, A. (2005) SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Science International,* 154**,** 181-194.

STAYNOV, D. Z. (2000) DNase I digestion reveals alternating asymmetrical protection of the nucleosome by the higher order chromatin structure. *Nucleic Acids Research,* 28**,** 3092-3099.

SYVANEN, A. C. (1999) From gels to chips: “minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. *Human Mutation,* 13**,** 1-10.

THOMPSON, M. D., BOWEN, R. A. R., WONG, B. Y. L., ANTAL, J., LIU, Z., YU, H., SIMINOVITCH, K., KREIGER, N., ROHAN, T. E. & COLE, D. E. C. (2005) Whole genome amplification of buccal cell DNA: genotyping concordance before and after multiple displacement amplification. *Clinical Chemistry and Laboratory Medicine,* 43**,** 157-162.

THORISSON, G. A. & STEIN, L. D. (2003) The SNP consortium wbsite: past, present and future. *Nucleic Acids Research,* 31**,** 124-127.

TSUKADA, K., TAKAYANAGI, K., ASAMURA, H., OTA, M. & FUKUSHIMA, H. (2002) Multiplex short tandem repeat typing in degraded samples using newly designed primers for the TH01, TPOX, CSF1PO, and vWA loci. *Legal Medicine,* 4**,** 239-245.

VAARNO, J., YLIKOSKI, E., MELTOLA, N. J., SOINI, J. T., HANNINEN, P., LAHESMAA, R. & SOINI, A. E. (2004) New separation free assay technique for SNPs using two -photon excitation fluorometry. *Nucleic Acids Research,* 32**,** 1-9.

VACCA, D. J., BLEAM, W. F. & HICKEY, W. J. (2005) Isolation of soil bacteria adapted to degrade humic acid-sorbed phenantherene. *Applied and Environmental Microbiology,* 71**,** 3797-3805.

VALLONE, P. M. & BUTLER, J. M. (2004) Autodimer:A screening tool for primer-dimer and hairpin structures. *BioTechniques,* 37**,** 226-231.

VALLONE, P. M., DECKER, A. E. & BUTLER, J. M. (2005) Allele frequencies for 70 autosomal SNP loci with U.S. Cuacasian, African-American, and Hispanic samples. *Forensic Science International,* 149**,** 279-286.

VALLONE, P. M., JUST, R. S., COBLE, M. D., BUTLER, J. M. & PARSONS, T. J. (2004) A multiplex allele specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. *International Journal of Legal Medicine,* 118**,** 147-157.

VEGA, F. M. D. L., LAZARUK, K. D., RHODES, M. D. & WENZ, M. H. (2005) Assessment of two flexible and compatible SNP genotyping platforms: TaqMan® SNP genotyping assays and the SNPlex™ genotyping system. *Mutation Research,* 573**,** 111-135.

VENTER, J. C., ADAMS, M. D., MYERS, E. W., LI, P. W. & ETAL (2001) The sequence of the human genome. *Science,* 291**,** 1304-1351.

WALLACE, R. B., SHAFFER, J., MURPHY, R. F., BONNER, J., HIROSE, T. & ITAKURA, K. (1979) Hybridization of synthetic oligodeoxyribonucleotides to x 174 DNA: the effect of single base pair msismatch. *Nucleic Acid Research,* 6**,** 3543-3557.

WOLFF, J. N. & GEMMELL, N. J. (2008) Combining allele - specific fluorescent probes and restriction assay in real - time PCR to achieve SNP scoring beyond allele ratios of 1:1000. *BioTechniques,* 44**,** 193-199.

YANG, I., KIM, Y.-H., BYUN, J.-Y. & PARK, S.-R. (2005) Use of multiplex polymerase chain reactions to indicate the accuracy of the annealing temperature of thermal cycling. *Analytical Biochemistry,* 338**,** 192-200.

ZAHRA, N. (2009) The development of PCR internal controls (PICs) for forensic DNA analysis. *School of Forensic and Investigative Sciences.* Preston, University of Central Lancashire.

Appendix A

A1. Indicated below are quantification results for DNA concentration obtained from 100 individuals from UAE. The highlighted 25 samples are selected for SNPs profiling.

Samples number	Concentrations (ng/µ)	Samples number	Concentrations (ng/µl)	Samples number	Concentrations (ng/µl)
1	1.26	44	0.61	87	7.55
2	2.74	45	2.18	88	5.88
3	5.58	46	1.55	89	1.23
4	8.16	47	2.22	90	3.95
5	0.95	48	1.10	91	0.25
6	1.54	49	2.54	92	3.69
7	6.41	50	2.17	93	3.69
8	4.59	51	5.33	94	3.45
9	4.98	52	26.76	95	3.45
10	4.95	53	11.23	96	0.7
11	0.86	54	0.38	97	1.72
12	17.8	55	2.25	98	1.00
13	3.02	56	1.64	99	3.74
14	7.71	57	2.58	100	3.07
15	3.02	58	1.11		
16	1.66	59	4.13		
17	5.46	60	3.64		
18	0.39	61	2.47		
19	1.22	62	8.76		
20	2.88	63	3.25		
21	4.55	64	8.57		
22	3.82	65	4.72		
23	0.54	66	1.74		
24	2.46	67	1.07		
25	3.93	68	0.92		
26	8.03	69	1.09		
27	6.86	70	1.66		
28	2.72	71	5.05		
29	1.39	72	2.69		
30	3.71	73	2.65		
31	0.67	74	5.64		
32	7.84	75	2.36		
33	26.32	76	5.23		
34	1.07	77	6.91		
35	0.95	78	2.45		
36	1.08	79	4.88		
37	14.28	80	1.31		
38	1.26	81	0.58		
39	18.46	82	0.98		
40	1.59	83	2.01		
41	1.18	84	0.36		
42	0.81	85	7.76		
43	0.50	86	5.65		

A2 A. Showing below are SNP RFUs obtained from artificially degraded DNA from saliva samples under 100% humidity and 37 °C. The results are for individual 1. [0A] represents the reference sample and numbers 3 to 18 are the durations of incubation. [np] indicates no profile and [pp] partial profile.

Triplex 1		Repeat 1			Repeat 2			Repeat 3	
SNP type amplicon size	AG/ 90 4-4	AG/ 110 19-2	CT/ 142 13-4	AG/ 90 4-4	AG/ 110 19-2	CT/ 142 13-4	AG/ 90 4-4	AG/ 110 19-2	CT/ 142 13-4
In house code									
0A	1115/2137	1220/1332	511 TT	1217/2203	1157/938	464TT	2957/7164	1660/2708	1404TT
3	1985/4925	1980/1127	388 TT	2261/5099	1425/1520	489TT	2136/4613	1758/1088	374 TT
6	310/919	205/100	np	374/485	118/pp	np	302/1403	407/503	147
9	165G/pp	np	np	np	np	np	207G/pp	102A/pp	np
12	np	np	np	np	np	np	np	np	np
15	np	np	np	np	np	np	np	np	np
18	np	np	np	np	np	np	np	np	np
Triplex 2	AG/ 92 21	CT/ 119 18-3	AC/147 17-3	AG/ 92 21	CT/ 119 18-3	AC/ 147 17-3	AG/ 92 21	CT/ 119 18-3	AC/ 147 21
In house code									
0A	4182 AA	219/619	4791AA	5845 AA	287/723	5244A	5613AA	295/755	5204AA
3	3113AA	210/450	3039AA	3758AA	190/502	3872AA	3984AA	214/482	3476AA
6	607AA	np	810 AA	425AA	np	312AA	785AA	128T/pp	718AA
9	150AA	np	np	124AA	np	313AA	274AA	np	np
12	260 AA	np	np	206	np	np	379AA	np	np
15	189AA	np	np	254	np	np	152 AA	np	np
18	134AA	np	np	126	np	np	145 AA	np	np

A2 B. Showing below are SNP RFUs obtained from artificially degraded DNA from saliva samples under 100% humidity and 37 °C. The results are for individual 2. 0B represents the reference sample and numbers 3 to 18 are the duration of incubation. [np] indicated no profile,[pp] partial profile and [np] sample was not tested because the template was estimated to be 0.00 ng/ µl.

Triplex 1	Repeat 1			Repeat 2			Repeat 3		
SNP type/ amplicon size	AG/ 90 4-4	AG/ 110 19-2	CT/ 142 13-4	AG/ 90 4-4	AG/ 110 19-2	CT/ 142 13-4	AG/ 90 4-4	AG/ 110 19-2	CT/ 142 13-4
In house code									
0B	6281AA	1687/2074	2166TT	6209AA	1770/3241	2244TT	5732AA	1686/2555	2242TT
3	1112AA	354/280	150TT	899AA	266/356	153TT	1161AA	348/508	np
6	2005	485/794	375TT	2786AA	619/845	819TT	2687	780/1207	525TT
9	nt	nt	nt	nt	nt	nt	nt	nt	nt
12	np	np	np	170AA	118 G/pp	np	155AA	np	np
15	np	np	np	np	np	np	np	np	np
18	nt	nt	nt	nt	nt	nt	nt	nt	nt
Triplex 2	AG/ 92 21	CT/ 119 18-3	AC/ 147 17-3	AG/ 92 21	CT/ 119 18-3	AC/ 147 17-3	AG/ 92 21	CT/ 119 18-3	AC/ 147 21
In house code									
0B	5747AA	244/570	2520/1298	5709AA	245/624	3050/1198	5504AA	234/607	2606/1275
3	781AA	NP	491/155	728AA	114T/pp	482/195	1464AA	188T/pp	864/358
6	2052AA	NP	551/286	2244AA	119/222	999/329	2183AA	106/208	1014/455
9	nt	nt	nt	nt	nt	nt	nt	nt	nt
12	176AA	np	141A/pp	103AA	np	119A/pp	121AA	np	np
15	114AA	np	np	114AA	np	np	128AA	np	np
18	nt	nt	nt	nt	nt	nt	nt	nt	nt

A3. Indicated below are SGM plus® DNA RFUs obtained from artificially degraded DNA from saliva samples (humidity/temperature). The percentage results were based on the number of loci successfully typed (excluding amelogenin). [0A] and [0B] represent reference samples and numbers 3 to 18 are the durations of incubations. [np] indicates no profile, [pp] partial profile and nt; sample was not tested because the template was estimated to be 0.00 ng/ µl.

Sample (days)	SGM plus® loci										Successful Results of SGM plus [a]
	D3S1358	vWA	D16S539	D2S1338	D8S1179	D21S11	D18S51	D19S433	THO1	FGA	
Ind 1											
0A	2826/2087	2273/23399	1726/1961	1595/1416	2459/1753	1804/1957	1108/1062	1530/1213	1226/1161	804/762	100
3	880/569	631/713	549/414	338/302	1068/706	419/348	135/145	447/504	388/271	122/132	100
6	np	110/pp	np	np	np	np	np	np	np	np	5
9	np	np	np	np	np	np	np	np	np	np	0
12	np	np	np	np	np	np	np	np	np	np	0
15	np	np	np	np	np	np	np	np	np	np	0
18	np	np	np	np	np	np	np	np	np	np	0
Ind 2											
0B	2888/2747	2507/2445	1883/1498	2504	1887/1797	2142/1982	1690/1231	1793/1544	2612	1164/911	100
3	np	181/pp	np	109	255/186	np	np	147/136	107	np	55
6	np	208/pp	121/pp	117	254/145	np	np	115/pp	np	np	40
9	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	0
12	np	np	np	np	np	np	np	np	np	np	0
15	np	np	np	np	np	np	np	np	np	np	0
18	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	0

[a] number of alleles observed in each period.

A4 A. Showing below are SNP RFUs obtained from artificially degraded DNA from semen samples under 100% humidity and 37 °C for individual 1. [0A] represents the reference sample and 3 to 18 are the durations of incubation.

Triplex 1	Repeat 1			Repeat 2			Repeat 3		
SNP type/ amplicon size	AG/ 90 4-4	AG/ 110 19-2	CT/ 142 13-4	AG/ 90 4-4	AG/ 110 19-2	CT/ 142 13-4	AG/ 90 4-4	AG/ 110 19-2	CT /142 13-4
In house code									
0A	3189AA	1165/1396	702 TT	4667AA	1607/1636	822 TT	4726AA	826/1305	699 TT
3	5583AA	1541/2439	850 TT	4496AA	1046/1536	906 TT	6216AA	2046/2596	1169 TT
6	1481AA	1343/2449	637 TT	1559AA	1904/1659	666 TT	2270AA	2036/2886	754 TT
9	904 AA	1029/1422	476 TT	904 AA	1149/1182	489 TT	1285AA	1487/1625	671 TT
12	2344AA	1386/2210	758 TT	2877AA	1482/1929	932 TT	3301AA	2055/3203	1120 TT
15	2758AA	1462/1976	919 TT	2479AA	1700/1811	755 TT	2553AA	1589/2162	791 TT
18	2099AA	1573/2634	821 TT	2175AA	1686/2539	913 TT	1941AA	2023/1952	885 TT
Triplex 2	AG/ 92 21	CT/ 119 18-3	AC/ 147 17-3	AG/ 92 21	CT/ 119 18-3	AC /147 17-3	AG/ 92 21	CT/ 119 18-3	AC/ 147 17-3
In house code									
0A	3304AA	157/392	1405/515	3603AA	197/497	1779/748	2950AA	170/489	1469/671
3	1592AA	137/333	982/421	3524AA	234/624	1460/662	3246AA	228/507	1579/689
6	2445AA	134/422	1871/575	1931AA	137/324	1912/745	2376AA	109/455	2009/708
9	2748AA	108/388	2297/935	2330AA	156/285	198/684	2079AA	136/287	1968/695
12	3927AA	189/555	2210/884	2996AA	165/476	2131/653	1870AA	128/307	1250/606
15	2975AA	148/343	1851/755	1792AA	122/443	1457/627	3258AA	156/424	1996/837
18	3146AA	149/367	2302/1133	2400AA	168/380	2260/1075	2648AA	186/389	2682/723

A4B. Showing below are SNP RFUs obtained from artificially degraded DNA from semen under 100% humidity and 37 °C for individual 2. [0B] represents the reference sample and 3 to 18 are the durations of incubation. na; indicates sample not available.

Triplex 1	Repeat 1			Repeat 2			Repeat 3		
SNP type/ amplicon size	AG/ 90 4-4	AG/ 110 19-2	CT/ 142 13-4	AG/ 90 4-4	AG/110 19-2	CT/ 142 13-4	AG/ 90 4-4	AG/ 110 19-2	CT/142 13-4
In house code									
0B	2208/546	2430AA	806 TT	2790/5397	2552AA	906 TT	1956/4284	2218AA	409 TT
3	703/1843	1244AA	256 TT	908/1896	946 AA	293 TT	901/2957	1794 AA	585 TT
6	1346/2807	1722AA	490 TT	2161/4114	2577AA	712 TT	1563/3074	2289AA	507 TT
9	1385/2787	1698AA	663 TT	1542/3210	1858AA	534 TT	1689/2615	1737 AA	614 TT
12	na	na	na	na	na	na	na	na	na
15	1622/3888	2246AA	713 TT	2023/4427	2963AA	758 TT	1660/3431	2271 AA	625 TT
18	1803/2635	2378AA	735 TT	1756/5241	2831AA	957 TT	1648/3194	2716AA	775 TT
Triplex 2	AG/ 92 21	CT/ 119 18-3	AC/ 147 17-3	AG/ 92 21	CT/ 119 18-3	AC/147 17-3	AG/ 92 21	CT/ 119 18-3	AC/ 147 21
In house code									
0B	3727AA	250/640	2012/1107	4207AA	623/515	2345/823	3653AA	241/570	1762/1030
3	1980AA	120/267	789/240	1853AA	121/291	1131/408	2126AA	122/190	1183/479
6	2187AA	146/288	1268/384	2447 AA	133/358	147/509	2610 AA	205/304	1668/600
9	2691 AA	186/428	1272/367	2154 AA	168/314	1283/553	2738 AA	176/523	1117/787
12	na	na	na	na	na	na	na	na	na
15	4010 AA	196/504	2140/996	3753 AA	194/396	2042/817	2906 AA	157/491	1578/711
18	3812 AA	205/487	2506/1049	3142 AA	196/425	2074/837	2238 AA	153/357	1241/770

A4C. Indicated below SGM plus® DNA RFUs obtained from artificially degraded DNA from semen samples (humidity/temperature). The percentage results were based on the number of loci successfully typed (excluding amelogenin). [0A] and [0B] represent reference samples and numbers 3 to 18 are the durations of incubations. na; indicates sample was not available.

Sample (days)	SGM plus® loci										Successful[a] Results of SGM plus®
	D3S1358	vWA	D16S539	D2S1338	D8S1179	D21S11	D18S51	D19S433	THO1	FGA	
Ind 1											
0A	1216/1974	752/1092	862/832	1355	1583/1004	971/1117	842/1179	861/809	1339	956/510	100
3	2823/2871	2727/1877	1978/2059	2702	3468/3006	2727/2249	1622/1607	1452/1495	2101	1381/1383	100
6	1187/1529	1297/834	914/894	1543	868/648	510/497	524/370	817/831	1197	289/314	100
9	869/1273	832/689	627/613	811	600/660	149/190	305/306	341/440	770	220/151	100
12	2072/2738	1884/1566	1727/1297	1857	1621/1421	1289/1141	1324/1176	970/1122	1836	994/783	100
15	1209/1542	1019/531	783/592	1135	1063/693	241/604	739/561	443/420	1082	376/401	100
18	1130/1563	1196/1006	681/763	864	701/870	387/384	405/368	652/510	1273	295/345	100
Ind 2											
0B	1707/2426	3101	1017/1145	1077/728	2302	1588/1390	1087/930	712/774	863/564	633/744	100
3	419/996	59	277/337	310/221	576	454/346	477/343	175/214	484/268	240/215	100
6	747/1380	1383	510/768	369/256	1289	1012/420	437/478	334/375	509/364	191/291	100
9	684/1004	1045	610/509	321/192	988	308/344	390/352	365/356	516/381	161/236	100
12	na	na	na	na	na	na	na	na	na	na	na
15	857/1005	1334	508/604	386/324	1117	513/400	521/465	399/389	538/477	457/298	100
18	983/1139	1958	693/600	291/312	1438	671/350	515/567	575/512	670/416	244/346	100

[a] number of alleles observed in each period.

A5A. Showing below are SNP RFUs obtained from artificially degraded DNA from saliva samples under room temperature (22 °C), the results are for individual 1.

Triplex 1	Repeat 1			Repeat 2			Repeat 3		
SNP type/ amplicon size	AG/ 90 4-4	AG/ 110 19-2	C/ 142 13-4	AG/ 90 4-4	AG/ 110 19-2	CT/ 142 13-4	AG/ 90 4-4	AG/ 110 19-2	CT/ 142 13-4
In house code									
0A	1115/2137	1220/1332	511 TT	1217/2203	1157/938	464 TT	2957/7164	1660/2708	1404 TT
3	1431/4091	921/1003	1026 TT	2050/3780	1172/1062	918 TT	1486/2783	702/1026	920 TT
6	1993/2858	1069/735	959 TT	1491/2372	886/841	938 TT	2327/3097	992/888	1287 TT
9	2538/3609	1409/1258	1182 TT	1713/3905	766/941	1266 TT	2179/4102	999/289	1442 TT
12	2046/5609	1051/1294	1379 TT	1691/3369	791/999	1034 TT	2089/5277	1382/1461	1707 TT
15	2002/3642	910/847	1155 TT	1239/4080	796/661	844 TT	1539/2743	642/943	816 TT
18	2047/3495	685/808	779 TT	1554/3087	622/626	793 TT	1404/3024	678/722	912 TT
Triplex 2	AG/ 92 21	CT/ 119 18-3	AC/ 147 17-3	AG/ 92 21	CT/ 119 18-3	AC/147 17-3	AG/ 92 21	CT/ 119 18-3	AC/ 147 21
In house code									
0A	4182 AA	219/619	4791AA	5845 AA	287/723	5244 AA	5613 AA	295/755	5204 AA
3	3261AA	148/352	2995AA	2721 AA	122/401	3011 AA	2792 AA	112/367	2551 AA
6	1455AA	105/250	1741AA	2612 AA	122/340	2522 AA	2591 AA	100/252	2059 AA
9	2865AA	139/233	2163AA	3012 AA	141/296	2555 AA	3091 AA	121/277	2697 AA
12	3573AA	151/411	3595AA	3552 AA	191/386	3198 AA	2776 AA	139/343	2655 AA
15	2368AA	115/276	2064AA	2634 AA	148/324	2390 AA	2447 AA	113/259	1978 AA
18	1723AA	102/286	2238AA	2204AA	105/256	2593 AA	2493 AA	288 T/PP	2221 AA

A5B. Showing below are SNP RFUS obtained from artificially degraded DNA from saliva samples under room temperature (22 °C), the results are for individual 2

Triplex 1		Repeat 1			Repeat 2			Repeat 3	
SNP type/ amplicon size	AG/ 90 4-4	AG/ 110 19-2	CT/ 142 13-4	AG/ 90 4-4	AG/ 110 19-2	CT/ 142 13-4	AG/ 90 4-4	AG/ 110 19-2	CT/ 142 13-4
In house code									
0B	6281AA	1687/2074	2166TT	6209AA	1770/3241	2244 TT	5732AA	1686/2555	2242TT
3	2459AA	1012/1379	1189TT	3081AA	1007/1145	1146TT	3819AA	1241/1474	1325TT
6	5878AA	2281/3987	2728TT	5874AA	2320/2795	2383TT	5896AA	2391/3153	2172TT
9	3884AA	1190/1545	1510TT	3983AA	1346/1564	1500TT	4461AA	1354/1684	1506TT
12	1976AA	653/648	677 TT	2147AA	732/743	832 TT	2898AA	817/705	887TT
15	4251AA	1001/1533	1329TT	4105AA	922/1463	1316TT	2385AA	641/834	781TT
18	5941AA	1403/2292	1581TT	5083AA	1444/1777	1619TT	4514AA	1386/1538	1537TT
Triplex 2	AG/ 92 21	CT/ 119 18-3	AC/ 147 17-3	AG/ 92 21	CT/ 119 18-3	AC/ 147 17-3	AG/ 92 21	CT/ 119 18-3	AC/ 147 21
In house code									
0B	5747AA	244/570	2520/1298	5709AA	245/624	3050/1198	5504AA	234/607	2606/1275
3	3245AA	122/326	1455/588	3041AA	132/341	1121/694	2709AA	168/257	1639/578
6	5663AA	226/633	2671/1188	6088AA	234/608	2760/1471	5927AA	204/602	2549/1397
9	3134AA	137/293	1259/530	3999AA	158/402	1317/633	2533AA	125/349	1356/715
12	2363AA	114/284	1157/530	2300AA	138/263	1070/589	2434AA	101/350	1227/455
15	4539AA	186/428	1719/1087	5023AA	198/409	2026/753	4836AA	134/481	2241/777
18	5030AA	195/533	2542/101	5126AA	225/594	2545/1196	4373AA	184/445	1192/510

A5C. Indicating below are SGM plus® DNA profiles obtained from artificially degraded DNA from saliva samples (room temperature). The percentage results were based on a number of loci types successfully except for amelogenin.0A and 0B represent reference samples and numbers 3 to 18 are the durations of incubations.

Sample (days)	SGM plus® loci										Successful Results of SGM plus®
	D3S1358	vWA	D16S539	D2S1338	D8S1179	D21S11	D18S51	D19S433	THO1	FGA	
Ind 1											
0A	2826/2087	2273/23399	1726/1961	1595/1416	2459/1753	1804/1957	1108/1062	1530/1213	1226/1161	804/762	100
3	918/1207	768/666	730/932	564/377	856/862	813/661	506/719	428/346	488/535	501/469	100
6	1106/632	794/833	596/741	630/521	1034/568	822/754	507/614	447/447	492/473	404/448	100
9	1002/790	998/852	706/525	456/237	638/679	503/645	511/596	356/481	537/394	437/395	100
12	1276/987	772/996	890/503	651/558	863/693	780/904	539/704	456/408	479/272	380/438	100
15	947/813	530/835	655/599	545/358	917/700	745/567	402/323	512/427	141/333	354/313	100
18	794/579	494/508	536/361	565/307	849/452	690/653	314/389	290/255	316/238	295/153	100
Ind 2											
0B	2888/2747	2507/2445	1883/1498	2504	1887/1797	2142/1982	1690/1231	1793/1544	2612	1164/911	100
3	1409/1365	1009/992	1093/836	1090	822/848	868/1026	926/761	545/629	1076	513/295	100
6	2165/2871	2519/2201	2119/1455	2544	2090/1721	2008/1196	1517/1402	1413/1271	2507	1014/957	100
9	1042/1181	1075/1023	930/725	909	693/670	677/876	682/622	682/449	1056	562/455	100
12	687/606	547/426	608/414	682	376/421	427/357	452/584	387/310	662	242/247	100
15	1593/1481	1460/1141	880/793	1365	1139/1149	1025/1137	891/945	624/754	1239	633/606	100
18	1487/1738	1800/1586	1476/1222	1564	1721/1607	1910/1272	1082/1402	1000/845	1680	819/637	100

[a] number of alleles observed in each period

A6. Showing below are SNP RFUs obtained from artificially degraded DNA from saliva samples under UAE natural conditions for December 2007/January 2008. 0A; represents reference sample and numbers 3 to 12 are the durations of incubation. [np] indicates no profile, [pp] indicates partial profile.

Triplex 1		Repeat 1			Repeat 2			Repeat 3	
SNP type/ amplicon size	AG 90 4-4	AG 110 19-2	CT 147 13-4	AG 90 4-4	AG 110 19-2	CT 114 13-4	AG 90 4-4	AG 110 119-2	CT 147 13-4
In house code									
0A	2638/7491	3005/4446	3007TT	2159/6762	2144/3034	2189 TT	1918/5349	2814/4043	2790 TT
3	955/2221	794/482	311TT	656/1608	530/314	301TT	511/2122	516/301	299 TT
6	2254/7556	3081/3194	1509 TT	298/1051	409/296	213 TT	516/1313	489/311	229 TT
12	801/1382	1644/580	467 TT	319/868	244/164	Np	353/573	263/146	131 TT
Triplex 2	AG 92 21	CT 119 18-3	AC 147 17-3	AG 92 21	CT 119 18-3	AC 147 17-3	AG 92 21	CT 119 18-3	AC 147 17-3
In house code									
0A	7605 AA	498/1935	7458 AA	7287 AA	456/1781	3218 AA	2831 AA	249/377	5503 AA
3	3173 AA	101/293	2108 AA	2580 AA	250 T/PP	1738 AA	2108 AA	104/194	1424 AA
6	3655 AA	262/726	2473 AA	2708 AA	104/194	1424 AA	1146 AA	157T/pp	1103 AA
12	2519 AA	265/403	1802 AA	2519 AA	265/403	1802 AA	1425 AA	151 T/pp	931 AA

[a] number of alleles observed in each period

A7. Indicated below is SGM plus® DNA RFUs obtained from artificially degraded DNA from saliva samples under UAE natural conditions for December 2007/January 2008. [0A] represents reference sample and numbers 3 to 12 are the durations of incubation. [np] indicates no profile observed, [pp] partial profile.

Sample (days)	SGM plus® loci										Successful Results of SGM plus®
	D3S1358	vWA	D16S539	D2S1338	D8S1179	D21S11	D18S51	D19S433	THO1	FGA	
0A	981/1207	768/668	730/932	564/377	856/862	813/661	506/719	428/346	488/535	501/549	100
3	330/331	281/246	254/159	124/118	465/441	255/216	104/PP	436/387	179/142	np	85
6	355/320	243/173	139/pp	np	356/247	102/119	np	231/253	137/pp	np	60
12	np	np	np	np	np	np	np	np	np	np	0

A8. Showing below are SNP RFUs obtained from artificially degraded DNA from saliva samples under UAE natural conditions for September 2008. [0A] represents reference sample and numbers 3 to 18 are the durations of incubation. [np] indicates no profile observed, [pp] indicates partial profile.

Triplex 1	Repeat 1			Repeat 2			Repeat 3		
SNP type/ amplicon size	AG 90 4-4	AG 110 19-2	CT 147 13-4	AG 90 4-4	AG 110 19-2	CT 114 13-4	AG 90 4-4	AG 110 119-2	CT 147 13-4
In house code									
0A	2010/5213	1480/893	1144 TT	2128/5274	1585/938	1422 TT	2430/6055	1845/1105	1590TT
3	1754/3272	697/554	552 TT	1878/3257	742/557	687 TT	1524/2768	631/489	508 TT
6	1191/2684	824/675	382 TT	1296/2795	924/717	438 TT	1044/2214	777/601	363 TT
12	np	np	np	np	np	np	np	np	np
18	np	np	np	np	np	np	np	np	np
Triplex 2	AG 92 21	CT 119 18-3	AC 147 17-3	AG 92 21	CT 119 18-3	AC 147 17-3	AG 92 21	CT 119 18-3	AC 147 17-3
In house code									
0A	2673 AA	148/296	2656 AA	3159 AA	179/349	3159 AA	3424 AA	189/381	3286 AA
3	1747 AA	238T/pp	1045 AA	1777 AA	187/106	1252 AA	2242 AA	215/102	1851 AA
6	1997 AA	104/145	761 AA	1898 AA	170 T/pp	830 AA	1705 AA	179 T/pp	1018 AA
12	np	np	np	np	np	np	np	np	np
18	np	np	np	np	np	np	np	np	np

[a] number of alleles observed in each period

A9. Indicated below is SGM plus® DNA RFUs obtained from artificially degraded DNA from saliva samples under UAE natural conditions for September 2008. [A]. represents reference sample and numbers 3 to 18 are the durations of incubation. [np] indicates no profile observed, [pp] indicates partial profile.

Sample (days)	SGM plus® loci										Successful[a] Results of SGM plus®
	D3S1358	vWA	D16S539	D2S1338	D8S1179	D21S11	D18S51	D19S433	THO1	FGA	
0A	1272/1553	1306/1203	905/788	871/622	1270/1075	776/88	1131/647	1007/765	549/458	688/402	100
3	618/532	318/202	221/172	NP	445/418	199/165	NP	341/236	156/128	NP	14
6	244/PP	110/PP	NP	NP	190/PP	NP	NP	169/176	139/PP	NP	6
12	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	0
18	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	0

A10. Showing below are SNP RFUs obtained from artificially degraded DNA from saliva samples under UK natural conditions for August 2008. [0A] represents reference sample and numbers 3 to 18 are the durations of incubation. [np] indicates no profile observed, [pp] indicates partial profile.

Triplex 1		Repeat 1			Repeat 2			Repeat 3	
SNP type/ amplicon size	AG 90 4-4	AG110 19-2	CT142 13-4	AG 90 4-4	AG110 19-2	CT142 13-4	AG 90 4-4	AG110 19-2	CT142 13-4
In house code									
0A	1115/2137	1220/1332	511 TT	1217/2203	1157/938	464 TT	2957/7164	1660/2708	1404TT
3	822/1591	420/426	479 TT	902/1929	521/533	526 TT	1778/3401	957/984	801TT
6	2333/5777	1433/1304	1588TT	1322/4087	959/942	1026 TT	700/2785	758/832	672TT
9	2856/5590	731/686	855TT	2609/6784	718/727	1052TT	2350/6399	917/554	1055TT
12	1730/2293	403/331	690 TT	845/2107	229/204	226 TT	1904/3748	354/405	478 TT
15	288/901	114A/pp	111 TT	208/819	112/177	186 TT	193/780	265/108	251 TT
18	np	np	np	np	np	np	np	np	np
Triplex 2	AG92 21	CT119 18-3	AC147 17-3	AG 92 21	CT119 18-3	AC147 17-3	AG 92 21	CT119 18-3	AC147 21
In house code									
0A	4182 AA	219/619	4791	5845AA	287/723	5244AA	5613AA	295/755	5204AA
3	4169 AA	151/428	3043AA	4755AA	230/298	2840AA	4733AA	158/381	3237AA
6	4332AA	152/331	3060AA	5658AA	157/385	3051AA	5214AA	159/466	3072AA
9	3716AA	246T/pp	1408AA	2992AA	102/223	1058AA	2969AA	165T/pp	1429AA
12	3111AA	138T/pp	465AA	2765AA	176T/pp	808AA	2320AA	122T/pp	648AA
15	1209AA	np	364AA	1929AA	115T/pp	431AA	1906AA	125T/pp	421AA
18	np	np	np	np	np	np	np	np	np

A11. Indicated below is SGM plus® DNA RFUs obtained from artificially degraded DNA from saliva samples under UK weather conditions UK for August 2008. [0A] represents reference sample and numbers 3 to 18 are the durations of incubation. [np] indicates no profile observed, [pp] indicates partial profile.

Sample (days)	SGM plus® loci										Successful[a] Results of SGM plus®
	D3S1358	vWA	D16S539	D2S1338	D8S1179	D21S11	D18S51	D19S433	THO1	FGA	
0A	2826/2087	2273/23399	1726/1961	1595/1416	2459/1753	1804/1957	1108/1062	1530/1213	1226/1161	804/762	100
3	615/825	802/743	612/501	428/202	904/447	491/336	384/4116	594/513	359/275	280/253	100
6	656/1052	533/401	286/433	255/214	725/683	315/358	341/225	589/404	238/155	182/140	100
9	345/221	175/PP	np	np	209/pp	np	np	400/292	np	np	35
12	144/pp	np	np	np	159/pp	np	np	120/pp	np	np	15
15	np	np	np	np	122/pp	np	np	np	np	np	5
18	np	np	np	np	np	np	np	np	np	np	0

[a] number of alleles observed in each period

APPENDIX B

A. Courses Attended

1- Reference Manager Introduction

2- Technical Writing

3- Communication and Presentation skills workshop

4- Microsoft Excel

- Teambuilding, Networking and Leadership Skills
- Research Skills Workshop
- Word for Researchers
- A guide to the Examination Process: Writing and Oral
- NVivo for Research Students
- Research Skills Conflict Management
- Career Skills Workshop
- Adobe Photoshop Element
- SPSS1 and SPSS2
- PowerPoint for Researchers
- www. for Researchers
- Central Postgraduate Research Student Induction Day

B. Conference Proceedings

- Annual Faculty Research Day, June 2006- Poster presentation
- Annual Research Conference, June 2007- Poster Presentation
- 2nd National Forrest Conference 2006 – Poster Presentation

National Conferences

- The Forensic Science Society and Centre for Forensic Investigative,
- University of Teesside, September 2006 - Poster Presentation
- Lancaster University April, 2008
- University of Sheffield, July 2009

International Conferences

- ESWG 2006 Conference in Tuusula, Finland
- ISFG Congress 2007 in Copenhagen, Denmark- Poster presentation
- Applied Biosystems Seminar , May 2008 in Dubai

C. Publication

S.H. Sanqoor, S. Hadi, W. Gooodwin (2008) the study of single nucleotide polymorphisms (SNP) in Arab population – A tool for the analysis of degraded DNA. *Forensic Science International*: *Genetics* (in press).

Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International: Genetics Supplement Series 1 (2008) 484–486
www.elsevier.com/locate/FSIGSS

Research article

The study of single nucleotide polymorphisms (SNPs) in Arab populations—A tool for the analysis of degraded DNA

S.H. Sanqoor, S. Hadi, W. Goodwin *

School of Forensic and Investigative Sciences, University of Central Lancashire, Preston PR1 2HE, UK

Received 24 August 2007; accepted 10 October 2007

Abstract

DNA samples from Arab individuals from the United Arab Emirates and the State of Kuwait have been analysed using the Affymetrix GeneChip® Mapping 250 K STY Array to identify SNPs that are polymorphic within Arab populations. The screening process has allowed hundreds of SNPs that are polymorphic to be identified. SNaPshot™ assays have been developed for selected SNPs to assess the allele frequencies in a larger population sample and ultimately to create a SNP multiplex.

Keywords: Single nucleotide polymorphisms; Arab populations; Autosomal

1. Introduction

During the last few years, there has been an increase in the use of SNPs for the analysis of degraded DNA. SNPs are suitable for profiling degraded forensic samples due to the small PCR amplicons that are required [1].

Biological samples exposed to the harsh environmental conditions found in the Arabian Gulf region (such as high levels of UV light, high temperatures and high humidity) are prone to rapid degradation and therefore a large amount of forensic evidence can be lost using conventional STR technology.

Previous studies have identified SNPs that are suitable for forensic analysis [2]. However, the forensic utility of a large number of characterised SNPs has not been assessed to date. The initial aim of this study is to identify SNPs that will be informative in a forensic context within the Arab population.

2. Materials and methods

Ten blood samples were collected as dried stains; five on fabric swatches from unrelated individuals from the United Arab Emirates and five on FTA® cards from unrelated Kuwait individuals. DNA extraction was carried out using phenol/chloroform extraction.

An Affymetrix GeneChip® Mapping 250 K STY Array (Geneservice Ltd.) was used to profile SNPs. Based on the following criteria SNPs were selected for further analysis: allele frequencies of 0.45–0.55; a minimum distance of 100 kb between any two SNPs; flanking sequence matched with the reported NCBI Fasta sequence; at least 100 kb between the selected SNP and any STR locus commonly used in forensic analysis.

3. Results and discussion

One hundred SNPs have been selected in proportion to each of the 22 autosomal chromosomes (Fig. 1). As a

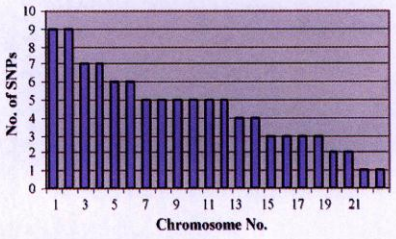


Fig. 1. Number of SNPs selected from each of the 22 autosomal chromosomes in proportion to chromosome length.

* Corresponding author. Tel.: +44 1772 894254; fax: +44 1772 894981.
E-mail address: whgoodwin@uclan.ac.uk (W. Goodwin).

1875-1768/$ – see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigss.2007.10.138