CHARACTERISATION OF EIGHT NON-CODIS MINISTRS IN FOUR SOUTH AFRICAN POPULATIONS TO AID THE ANALYSIS OF DEGRADED DNA

By

Aneesah Ismail

A thesis submitted in partial fulfillment of the requirements
for the degree of

MAGISTER SCIENTIAE

In the Department of Biotechnology
The University of the Western Cape

15 May 2009

Supervisors:

Prof. Sean Davison

Dr. Maria Eugenia d'Amato

Keywords

Short Tandem Repeat (STR)

Microsatellites

Loci

Combined DNA Index System (CODIS)

Degraded DNA

Mitochondrial DNA (mtDNA)

MiniSTR

Polymerase Chain Reaction (PCR)

Electrophoresis

Genotyping

Polymorphic Information Content (PIC)

Fixation Indexes

THE UNIVERSITY OF THE WESTERN CAPE

ABSTRACT

Characterization of eight non-CODIS miniSTRs in four South African populations to aid the analysis of degraded DNA

By Aneesah Ismail

In many forensic cases, such as mass disasters reconstruction cases, the recovered DNA is highly degraded. In such incidences, typing of STR loci has become one of the most powerful tools for retrieving information from the degraded DNA. However, as DNA degradation proceeds, three phenomena occur consecutively: loci imbalance, allele dropout and no amplification. To solve the problem of degraded DNA, redesigned primer sets have been developed in which the primers were positioned as close as possible to the STR repeat region. These reduced primer sets were called Miniplexes. Unfortunately, a few of the CODIS STR loci cannot be made into smaller amplicons. For this reason non-CODIS miniSTRs have been developed.

The present study was undertaken for the population genetic analysis of microsatellite variation in four South African populations; Afrikaner, Xhosa, Mixed Ancestry and Asian Indian using eight non-CODIS miniSTR loci. These miniSTRs loci were characterized within the populations by estimating the levels of diversity of the markers, estimating the population genetic parameters, and studying the inter-population relationships. All of the miniSTRs were amplified successfully and the genetic variability parameters across all loci in Afrikaner, Mixed Ancestry, Asian Indian and Xhosa were estimated to be in the range of 3 (D4S2364) to 12 (D9S2157) alleles, the total number of alleles over all loci ranged from 100 to 204, the allelic richness ranged from 3.612 to 10.307 and the heterozygosity ranged from 0.4360 to 0.8073. Genetic distance was least between Afrikaner and Asian Indian and highest between Xhosa and Mixed Ancestry. Deviations from Hardy-Weinberg equilibrium were not observed for most of the loci. The low mean FIS (-0.027) and FIT (-0.010) and FST (0.017) values across the populations indicated low level of inbreeding within (FIS) and among (FST) the populations. The Asian Indian population showed higher levels of the inbreeding coefficient, indicating less gene exchange between it and other populations.

These 8 markers can be used for genetic investigations and assessing population structure. The study contributed to the knowledge and genetic characterization of four South African populations. In addition, these MiniSTRs prove to be useful in cases where more genetic information is needed.

DECLARATION

I declare that *Characterization of eight non-CODIS miniSTRs in four South African populations to aid in the analysis of degraded DNA* is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Aneesah Ismail 15 May 2009

Signed.....

LIST OF ABBREVIATIONS

DNA	Deoxyribonucleic Acid
STR	Short Tandem Repeat
CODIS	Combined DNA Index System
MtDNA	Mitochondrial DNA
SNP	Single nucleotide polymorphism
EDNAP	European DNA Profiling Group
ENFSI	European Network of Forensic Science Institutes
PCR	Polymerase Chain Reaction
$MgCl_2$	Magnesium Chloride
dNTPs	deoxi Nucleotide TriPhosphates
BSA	Bovine Serum Albumin
EDTA	Ethylene Diamine Tetra Acetic Di-Sodium Salt
RFU	Relative Fluorescent Units
TEMED	N, N, N', N' Tetramethyl-EthyleneDiamine
AMPS	Ammonium Persulfate
PIC	Polymorphic Information Content

TABLE OF CONTENTS

LIST OF FIGURES

Figure:

LIST OF TABLES

Table:

ACKNOWLEDGMENTS

I would like to express my deepest gratitude and appreciation to my supervisors Prof. Sean Davison and Dr. Maria Eugenia d'Amato for their guidance, encouragement and support in my dissertation research. I would also like to thank the rest of the Forensic DNA Laboratory team for their valuable help and support. Thank you to the NRF for providing me with the finance to complete my research.

Finally and foremost, I would like to thank my family and friends. Without their love and support, I would not have accomplished this thesis.

Chapter 1

1.1 Introduction to DNA Typing

Overview of Forensic DNA Typing

Until recently, progress in forensic analysis was quite slow. The first genetic markers used for human identification purposes were the ABO blood groups in 1900. Nineteenth century scientists mixed the blood from different individuals in the laboratory while investigating the causes of blood-transfusion reactions and discovered the four blood groups designated A, B, AB and O. Different human populations were found to differ in the frequencies of the four types (Dean 2005).

Several other red-blood cell systems were discovered afterwards. By 1960, 17 systems had been discovered, but not all of them were useful for identification. The most useful was the Human Leukocyte Antigen (HLA) typing system because it was highly polymorphic. HLA are proteins or markers found on most cells in your body. The HLA system consisted of serological tests and some laboratories included a few serum proteins and enzymes. The frequency of a combination of such tests was typically one in a few hundred or less (NIJ Report 2000). However, criminal cases require a higher standard of evidence. Although a combination of blood groups and serum proteins gave small probabilities for a match between two unrelated individuals, more powerful methods were needed.

The year 1985 brought the major discovery of VNTRs (Variable Number of Tandem Repeats). VNTRs are DNA regions in which a short sequence, of ~8-35 bases in length, are tandemly

repeated 100 times or more (Jeffreys *et al.*, 1985). The amounts of repeats differ greatly between individuals, so VNTRs are highly variable. Currently the methods used by forensic scientists include RFLP (Restriction Fragment Length polymorphism) typing of VNTR loci and the amplification of specified loci by PCR and subsequent genotyping of these specific markers. Any material that contains nucleated cells such as blood, semen, saliva, hair, bones and teeth can potentially be analysed for DNA polymorphisms (Gill 2001).

Typing of VNTR loci by RFLP analysis is found to be the most discriminating technology for forensic DNA samples. Although this approach is found to be very reliable for both forensic and paternity testing, it has certain limitations:

For RFLP analysis at least 50ng of high molecular weight DNA is required (Budowle and Baechtel, 1990). Therefore samples that have been substantially degraded cannot be analysed by RFLP typing. RFLP analysis is also laborious and time consuming, and the number of validated VNTR loci are limited (Budowle *et al.*, 1998, NIJ Report 2000).

Alternatively, PCR-based assays have been used for DNA typing. Compared to the RFLP approach, PCR-based assays are more sensitive and specific and require less time and labour. Also, many degraded DNA samples can be amplified by PCR and subsequently typed because the amplified alleles are usually much smaller in size compared with those detected by RFLP analysis (Budowle *et al.*, 1998).

Introduction to Short Tandem Repeats

The technology forensic DNA laboratories use today includes numerous genetic markers. These markers differ in their power of discrimination and the speed at which the analysis can be performed. Eukaryotic genomes contain numerous repeated DNA sequences. The repeat units vary in size and are classified according to the length of the core repeat unit and the number of contiguous repeat units, or the overall length of the repeated region (Butler 2000). Certain long repeat units may contain core repeats of several hundred to several thousand bases and these regions are referred to as satellite DNA (Butler 2000).

Microsatellites, or Short Tandem Repeats (STRs), are repeated DNA units that are 2-6bp in length (Tautz 1993). For human identification purposes it is sensible to use markers that are highly polymorphic and exhibit the highest possible variation or a number of less polymorphic markers that can be combined to increase the power of discrimination between samples. STRs are therefore ideal for human individualisation, as these markers are highly polymorphic, which makes them highly variable between individuals (Bennet 2000).

STRs have a number of advantages compared to VNTRs. STRs have small alleles that make them easily amplifiable; their narrow size range permits multiplexing which allows multiple loci to be amplified simultaneously; their narrow size range also reduces allelic dropout from preferential amplification of smaller alleles and they have reduced stutter product formation. And lastly, because STRs generate small PCR products, more information may be retrieved from degraded DNA samples (Butler 2000). Typing of STR loci has become one of the most powerful tools for

individual discrimination because these loci are even found to be stable in decomposed tissue (Hoff-Olsen *et al*., 2001).

Another advantage of using STR technology is that polyacrylamide gels can resolve DNA fragments differing by as little as one nucleotide in length, allowing for precise allele designation (Wallin *et al*., 2002). And because electrophoretic separation is combined with automated fragment analysis and sizing software, such as those routinely used by commercially available DNA sequencing machines, the technique is quite rapid. The combination of these factors means that STR systems are therefore more likely to be successful on challenging forensic evidence such as mixtures and highly degraded DNA.

Allelic Ladders and Size Markers

Allelic ladders are used to ensure accurate designation of alleles and preliminary identification of rare alleles (Watson *et al.*, 2001). The fragment lengths of the DNA samples are calculated by directly comparing the electrophoretic migration of the unknown samples to that of a constant set of fragments of known size. Because the same ladders are run in each gel in the same amounts, the reproducibility of the detection method can be checked by the intensity of the bands or its corresponding peak height.

The way DNA migrates in gel electrophoresis is not only dependent on the size of the fragment, but depends on the nucleotide sequence as well (Frank and Körster, 1979). Thus only when the DNA length and sequence of the sample and size marker is the same can readings truly be accurate. For this purpose, allelic ladders are used in STR systems. The components of the allelic ladder and the sample fragments have the same length as well as the same sequence. As a result these migrate the same distance regardless of environmental changes such as percentage acrylamide, amount of salt in the buffer or electrophoresis voltage (Schumm, 1997). Allelic ladders are therefore run on every gel to verify the alleles.

CODIS – Combined DNA Index System

In 1989 the Technical Working Group of DNA Analysis Methods (TWGDAM) suggested that DNA technology be combined with computer science capabilities to help reduce the rate of violent crimes (Butler 2000). Consequently, in 1997 the FBI officially launched its national DNA database. While creating this database, named the Combined DNA Index System (CODIS), 17

candidate STR loci were evaluated. Only 13 of these were chosen. The 13 CODIS core STR loci are: CSF1PO, FGA, THO1, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S359, D18S51 and D21S11 (Butler 2000). The three most polymorphic markers are FGA, D18S51 and D21S11 with mutation rates of 0.28%, 0.22% and 0.19% respectively. While TPOX and THO1 shows the least variation amongst individuals with both having a mutation rate of 0.01% (Coble and Butler, 2005).

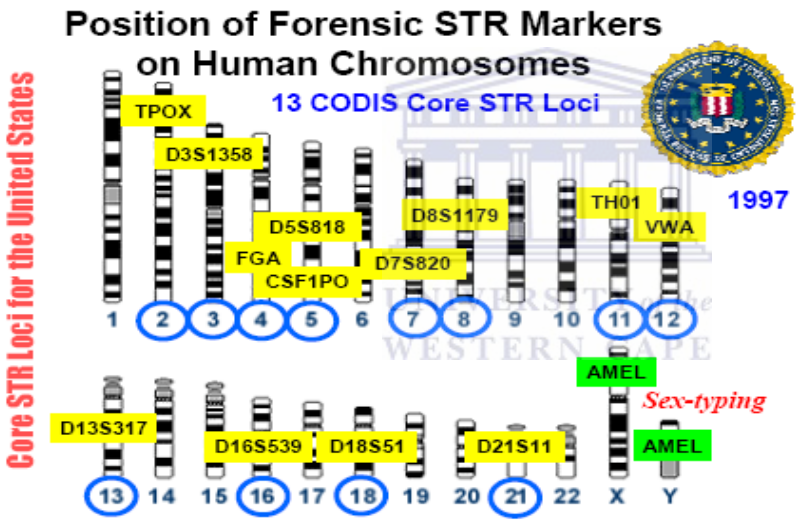


Fig. 1.1. Core CODIS loci and the sex determining marker Amelogenin (http://www.cstl.nist.gov/biotech/strbase/training.htm)

Commercial STR kits

The factors initially considered as disadvantageous to typing all 13 STR were an increased cost and labour, the increased time required to type all loci and the lack of availability of all 13 STR loci in commercial typing kits from a single manufacturer (Budowle *et al.*, 1998). However, the ability to amplify several markers simultaneously is very useful as it conserves the DNA sample, increases sample throughput and conserves reagents and consumables. Also, due to the development of the multi-coloured fluorescent systems, STR markers with overlapping size ranges can be amplified simultaneously (Wallin *et al.*, 2002). Therefore, due to the adoption of the 13 CODIS STR loci, two commercially available STR multiplex kits are available: the PowerPlex® 16 from Promega and the AmpF*l*STR® Identifiler™ kit from Applied Biosystems. These kits contain all 13 CODIS STR loci as well as additional loci to increase the power of discrimination. Table 1.1 provides information on the STR loci included in these kits.

Table 1.1: Summary information on the PowerPlex® 16 system and the AmpFlSTR® Identifiler™ kit. The STR loci included for each kit are indicated.

STR loci	PowerPlex® 16 (Promega)	AmpFlSTR® Identifiler™ (Applied Biosystems)
Amelogenin	•	•
THO1	•	•
CSF1PO	•	•
TPOX	•	•
FGA	•	•
D21S11	•	•
D7S820	•	•
D5S818	•	•
D8S1179	•	•
D16S539	•	•
vWA	•	•
D18S51	•	•
D13S317	•	•
D3S1358	•	•
D2S1338		•
D19S433		•
Penta D	•	
Penta E	•	

1.2 DNA Degradation

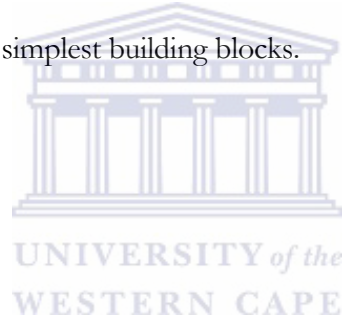
The process of Decomposition

The body starts decomposing approximately 4 minutes after death (Dent *et al.*, 2004; Thomas *et al.*, 2003; Vass 2001). The first process is autolysis or self-digestion. Because there is no more oxygen intake, the levels of carbon dioxide increases and the pH levels decrease. This causes waste to accumulate and poison the cells. At the same time, cellular enzymes such as lipases and proteases start digesting the cells from inside out, causing the cells to rupture (Pääbo *et al.*, 2004). Tissues that have a high enzyme or water content, such as the liver or brain, will undergo autolysis more rapidly. But eventually all the cells in the body will be affected. When enough cells have ruptured, the process of putrefaction will occur. Putrefaction is the process whereby microbes (bacteria, fungi and protozoa) cause soft tissues to deteriorate (Dent *et al.*, 2004; Hofreiter *et al.*, 2001). These microbes break down the tissues into gases, liquids and simple molecules. The gases formed are carbon dioxide, ammonia, hydrogen sulphide, methane, sulphur dioxide and hydrogen. These gases cause the body to swell. The accumulation of gas and fluids purge from the orifices (liquefaction) and shortly afterwards active decay begins (Vass 2001). Putrefaction is generally not observed until at least 48-72 hours after death (Dent *et al.*, 2004).

Fat and proteins are decomposed and produce phenolic compounds and glycerols. Amino acids are broken down to produce volatile fatty acids. In warm, moist environments adipocere formation occurs after putrefaction. Adipocere develops as the result of fat hydrolysis with the release of fatty acids, and is seen as deposits of yellowish-white, greasy, wax-like substance. The

tissues that survive the active decay process are usually mummified. The remaining skin is converted to leathery-like sheets that cling to the bone (Dent *et al.*, 2004).

The bone now undergoes a process called diagenesis. Diagenesis is a natural process that alters the proportions of collagen (organic) and calcium, magnesium and hydroxy-apatite (inorganic) of bone that is exposed to the environment. This is achieved by the exchange of natural bone constituents, filling in voids, adsorption onto the bone surface and leaching from the bone (Collins *et al.*, 2002). Human decomposition has been described in four stages: fresh (autolysis); bloat (putrefaction); decay (putrefaction, liquefaction and carnivores) and dry (diagenesis). All these processes combined result in complex structures like proteins, carbohydrates, sugars, collagen and lipids retuning to their simplest building blocks.



The process of DNA degradation

The rate of DNA degradation is dependent on light, humidity and the temperature of the environment (Bender *et al.*, 2004; Burger *et al.*, 1999, Hofreiter *et al.*, 2001). As the DNA decomposes it results in a permanent loss of nucleotide sequence information. However, PCR technology has made it possible to occasionally retrieve information from samples in which DNA breakdown is not yet complete. The first successful retrieval of ancient DNA sequences were the molecular cloning of DNA from a quagga (Higuchi *et al.*, 1984) and an Egyptian mummy (Pääbo 1985). The two main types of damage that will affect DNA are hydrolytic damage and oxidative damage (Hofreiter *et al.*, 2001; Höss *et al.*, 1996, Rogan and Salvo, 1990). Hydrolytic damage results in deamination of bases and in depurination and depyrimidation, which lead to destabilization and breaks in DNA molecules. Oxidative damage is caused by the direct interaction of ionising radiation with the DNA, as well as mediated by free-radicals created from water molecules by ionising radiation, which results in modified bases (Höss *et al.*, 1996). Fig 1.2 shows the types of damage that will most likely affect ancient DNA.

Normal cellular metabolism, carcinogenic compounds and ionising radiation produce free-radicals. One of these free-radicals is the highly reactive hydroxyl radical (OH), which causes damage to DNA and other biological molecules. This is perceived to be the major form of DNA damage and is referred to as oxidative damage (Dizdaroglu *et al.*, 2002; Rogan and Salvo, 1990; Thomas *et al.*, 2003). There are several mechanisms of oxidative damage to DNA:

1. The hydroxyl radical adds to double bonds of heterocyclic DNA bases and abstracts an H-atom from the methyl group of thymine and each of the five carbon atoms of 2-

deoxyribose (Rogan and Salvo, 1990). Further reactions of base and sugar radicals generate a variety of modified bases and sugars, base-free sites, strand breaks and DNA-protein cross-links (Dizdaroglu *et al.*, 2002).

2. Hydroxyl radicals add to the C5- and C6- positions of thymine and cytosine, generating C5-OH- and C6-OH- adduct radicals respectively. Oxidation reactions of the C5-OH-adduct radicals of cytosine and thymine followed by the addition of OH^- lead to the formation of cytosine glycol and thymine glycol, respectively. Oxidation products of the pyrimidine bases (cytosine and thymine) are called Hydantions. Hydantoins block DNA polymerises and thus PCR (Dizdaroglu *et al.*, 2002; Hofreiter *et al.*, 2001, Thomas *et al.*, 2003).

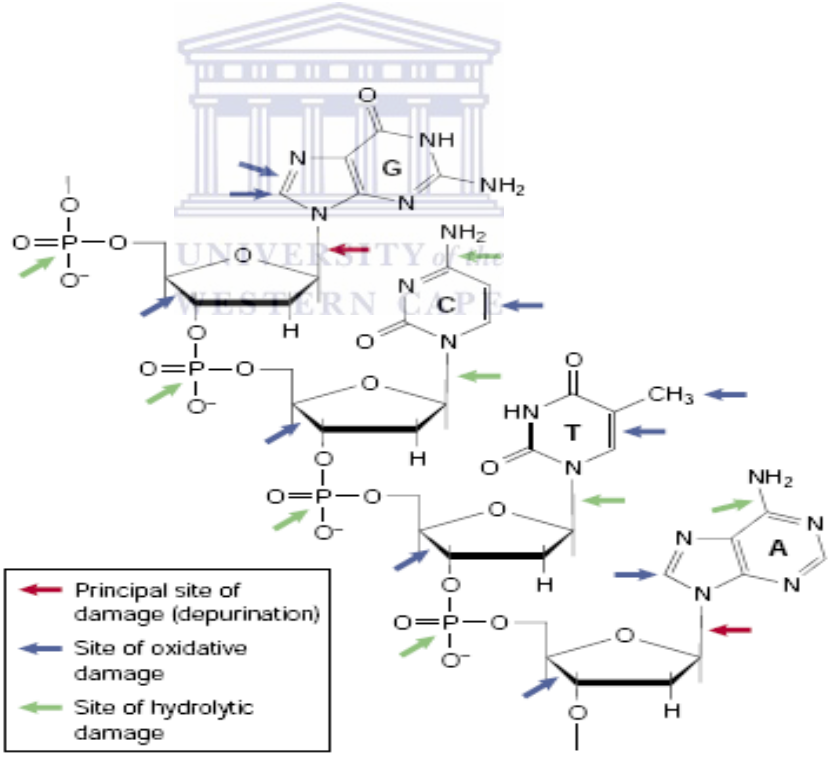


Fig. 1.2: Types of damage that affects ancient DNA. A short fragment of one strand of the DNA double helix containing the four common bases. The red arrows indicate principal sites of

damage, the blue arrows shows sites susceptible to oxidative damage and the green arrows show the sites prone to hydrolytic damage. Taken from Hofreiter, 2001.

3. Hydroxyl radicals add to purines giving rise to C4-OH-, C5-OH- and C8-OH- adduct radicals (Eglinton *et al*., 1991). C4-OH- and C5-OH-adduct radicals undergo dehydration and produces oxidising purine ($-H^+$) radicals, which reconstitute purines upon reduction (Mitchell *et al*., 2005).

4. The hydroxyl radical reacts with the sugar moiety of DNA by abstracting an H-atom from each of the carbon atoms. Further reactions of the formed carbon-centered sugar radicals generate various sugar products, strand breaks and base-free sites by various mechanisms (Höss *et al*., 1996). The modified sugars are either released from, or bound to the DNA with one or both phosphate linkages **(**Golenberg, 1996**)**.

The degradation of template DNA affects the success of genotyping that DNA (Bender *et al*, 2004; Graw *et al*., 2000; Rogan and Salvo, 1990; Utsuno and Minaguchi, 2004). Therefore one should understand the influence of template degradation on PCR. It has been found that DNA extracted from ancient remains is degraded to small fragment sizes, generally between 100-400bp. In rare cases ~500bp fragments have been retrieved (Pääbo *et al*., 2004). The reduction in size is due to the autolysis process, as well as non-enzymatic hydrolytic cleavage of phosphodiester bonds in the phosphate-bonds backbone that generate single strand nicks (Bender *et al*, 2004). In addition to these fragmentations and DNA modifications that block the elongation of DNA strands by Taq polymerase (Pääbo *et al*, 2004; Friedberg *et al*, 1995; Rogan and Salvo, 1990) other known and unknown types of damage are common when amplifying degraded DNA. Some of

these DNA modifications cause problems because even though they allow these template molecules to be amplified, they cause the incorporation of incorrect bases during PCR. The most common form of this incorrect base incorporation is the hydrolytic loss of amino groups from the bases adenine, cytosine, 5-methylcytosine and guanine, resulting in hypoxanthine, uracil, thymine and xanthine respectively (Friedberg *et al*., 1995). The deamination products of cytosine (uracil), 5-methylcytosine (thymine) and adenine (hypoxanthine) are particularly relevant to the amplification of degraded DNA because these cause the insertion of incorrect bases (A instead of G and C instead of T) when DNA polymerase synthesises new DNA strands (Friedberg *et al*, 1995).

These DNA modifications and fragmentations coupled with the absence of specific repair enzymes in postmortem cells may result in the production of STR artifacts or even the complete loss of the fragment (Bender *et al*, 2004; Höss *et al*., 1996; Takahashi *et al*., 1997).

The problem with Degradation

Utsuno and Minaguchi (2004) found that samples, which are stored correctly in dry conditions after collection, are well amplified and typing from these samples is reliable. But in many cases DNA is highly degraded either due to the quality of the sample itself or the surrounding environmental conditions (Hellman *et al*, 2001). These scenarios may include skeletal remains buried underground (Holland *et al*, 1993), decomposed bodies (Hoff-Olsen *et al*, 1999) or shed telogen hairs (Hellman *et al*, 2001). As DNA degradation proceeds, three phenomena occur consecutively loci imbalance, allele dropout and no amplification.

Allelic imbalance is often the first sign of the progression of DNA degradation. When allele imbalance of STR typing occurs, it is normally observed that the larger allele is fainter than the smaller allele (Utsuno and Minaguchi, 2004). So this may suggest that smaller loci have higher amplification efficiency than that of larger ones. This theory is further supported by Hummels *et al* (1999) who found that in samples where only some of the STR loci could be amplified, it were usually the larger loci that failed to amplify. Bender *et al* (2004) also reported incomplete or no results for STR loci larger than 250bp amplified from degraded DNA samples. This is to be expected due to the degradation of ancient DNA. Therefore, the poor amplification of larger size loci (300-500bp) in standard multiplex typing kits is common, and the yield of complete target fragments is significantly reduced. Thus, in multiplex kits with a wide variety of amplicon sizes, a decay curve is seen (Fig. 1.3), in which the peak height is inversely proportional to the amplicons length. In this case, the larger amplicons often have lower sensitivity and fall below the detection threshold. The occurrence of allele dropout is also observed; where one allele of a heterozygote fails to PCR amplify. This results in a partial genetic profile.

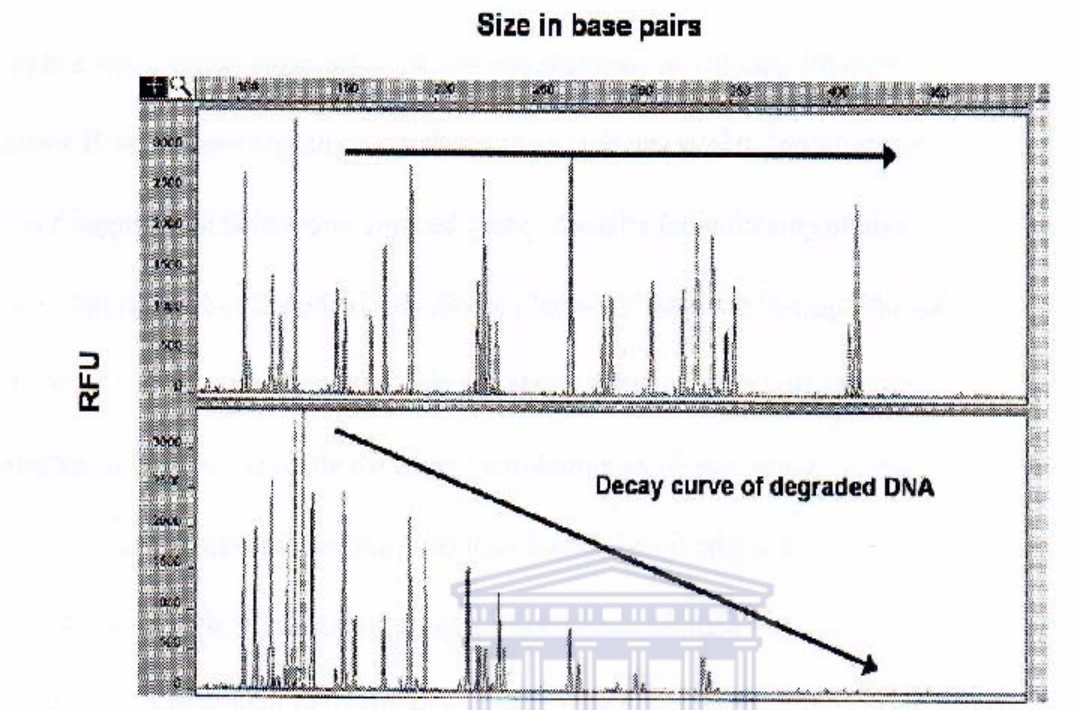


Fig. 1.3: Decay Curve of Degraded DNA. Top panel: Positive control sample amplified with the Powerplex® 16 kit. Lower panel: Bone sample that has been exposed to the environment for 3 years, also amplified with the Powerplex® 16 system. Note a clearly decreasing peak height with increasing fragment length. Taken from Chung *et al.*, 2004

Due to this limitation of commercial STR kits when typing highly degraded DNA samples, sequence analysis of hypervariable regions of mitochondrial DNA (mtDNA) is frequently used to obtain information for genotyping (Bender *et al.*, 2004). MtDNA is found in the mitochondrion (an organelle located in the cytoplasm of the cell) and is not associated with nuclear chromosomes. Mitochondria are transmitted by the eggs but not by the sperm. Each cell has approximately 500 copies of mtDNA, compared to two copies of nuclear DNA (Satoh and Kuroiwa, 1991). However, only 25% of the total DNA in the cell is mtDNA. This is due to the considerably small size of the mtDNA genome, consisting of 16569bp (Fig. 1.4) compared to the 3 billion bp of DNA in the nucleas (Butler 2000).

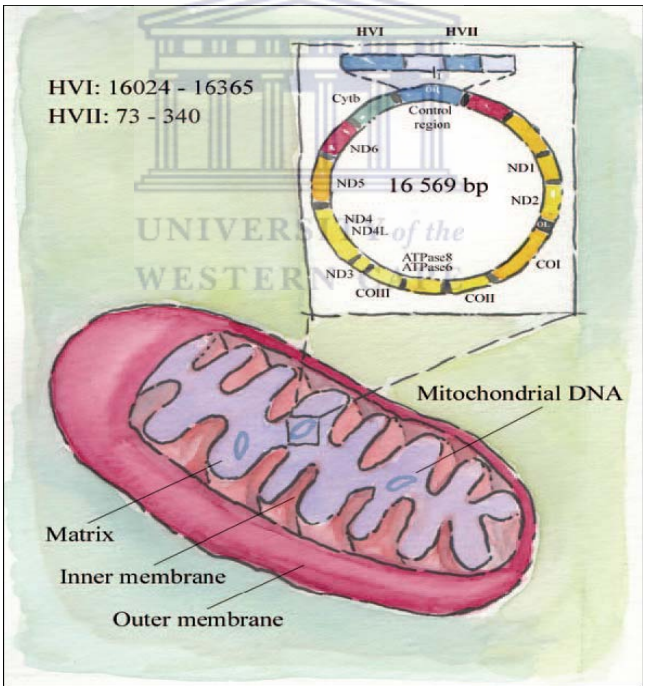


Fig. 1.4: Overview of the mitochondrion. Illustration taken from Nilsson (2007). Usually approximately 610bp are sequenced in routine forensic analysis (HVI:~341bp and HVII:~267)

Using mtDNA has numerous benefits over nuclear DNA. The higher copy number of mtDNA ensures a better PCR amplification success rate over nuclear loci. In addition, the mtDNA hypervariable region is small in size (~610bp), which makes it easily amplifiable (Forster *et al.*, 2002). These characteristics of mtDNA make it an attractive option for typing degraded DNA.

However, mtDNA testing may not be as effective for identification purposes. Because mtDNA is maternally inherited, it might be difficult to discriminate between maternally related individuals and mtDNA sequencing is a time consuming process. Furthermore, different mtDNA types may be present within the same individual (Holland and Parsons, 1999). This phenomenon, known as heteroplasmy, might result from point mutations in the germline (during meiosis) or somatic mutations (during replication in mitosis). Subsequent to the occurrence of a mutation, more than one type of mtDNA can be present in a cell or tissue because of the multicopy feature of mtDNA (Nilsson 2007). These features can make mtDNA analyses highly challenging.

Alternatively, a better approach has been developed. To solve the problem of degraded DNA, redesigned primer sets have been developed in which the primers were positioned as close as possible to the STR repeat region (Asamura *et al.*, 2006; Butler *et al.*, 2003; Grubweiser *et al.*, 2006; Hellman *et al.*, 2001; Opel *et al.*, 2006), thereby greatly reducing the PCR product size. More information can therefore be retrieved from degraded DNA samples. These reduced primer sets were called Miniplexes (Butler *et al.*, 2003).

1.3 MiniSTRs

MiniSTRs for degraded DNA

Many papers have demonstrated that a reduction of target size significantly increases amplification efficiency from highly degraded DNA samples. Hellman *et al.* (2001) used primer pairs producing PCR fragments less than 110bp for three STR loci; FES, THO1 and TPOX for typing DNA extracted from human telogen hairs. In a study done by Ricci *et al.* (1999), the primer pair for D12S391 STR loci were decreased from 205-253bp to 125-173 bp, and successfully applied to the analyses of degraded DNA samples. Reductions in primer pairs for THO1, D10S2325, DYS319 and DYS19 have also been reported (Wiegand and Kleiber, 2001). However, Butler *et al.* (2003) have produced a set of miniSTRs that complete the full set of CODIS loci for genotyping of forensic samples (Fig. 1.5). The reduction in size can be seen in Table 1.2.

In studying the minimal amount of template DNA required for determination of STR polymorphism, Bender *et al.*, (2000) determined that the allele amplification of loci with large allele sizes (>200bp) required a greater amount of template DNA with a higher molecular size than that of loci less than 200bp. Thus, because the miniplexes only require short fragments of DNA, the possibility of obtaining full profiles with degraded DNA or samples in low amounts are increased.

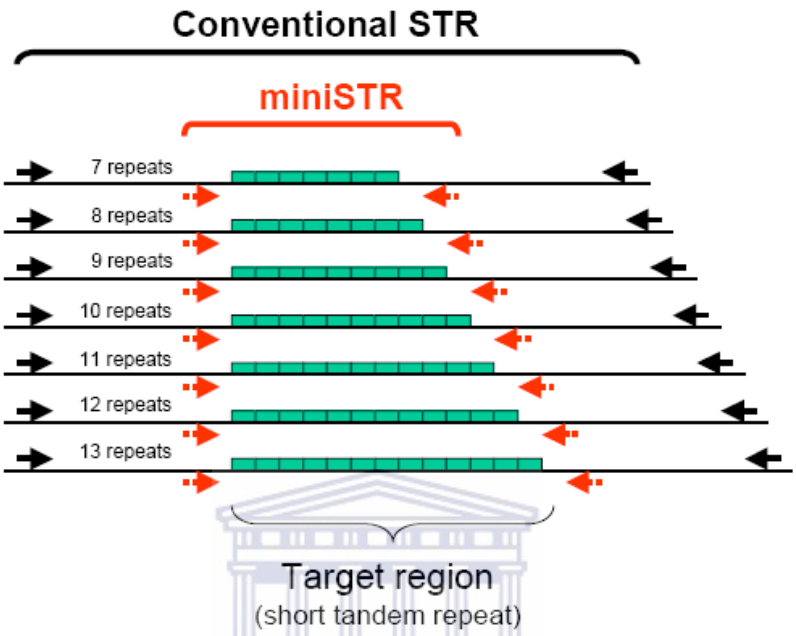


Fig. 1.5. Diagram comparing conventional STRs and MiniSTRs. The black arrows represent the conventional STR primers sets and the red arrows the reduced sized MiniSTRs (http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm).

CODIS MiniSTRs

The idea behind the development of the miniSTRs was to construct a set of primers, which produced amplified PCR products as short as possible. In commercial kits this cannot be done as it is necessary to analyse several loci in each dye lane. To avoid these loci from overlapping, commercial amplicons sizes are expanded to 350bp or more (Butler *et al.*, 2003, McCord 2005). This problem is eliminated in the miniplexes by placing only one STR amplicon in each dye lane. Although this makes it necessary to prepare 3 – 4 amplifications to cover all 13 CODIS loci, it may yield results for samples that previously did not amplify. Furthermore, Miniplex 1 and Miniplex 3 differ in size range and can be combined together to create a six-loci set called the "Big-mini" (Chung *et al.*, 2004).

New primers were not designed for the D3S1358, D19S433 and the amelogenin (sex-typing) markers, as these are already small in commercial kits. There is also a very small size reduction with the miniSTR primers for D5S818, D8S1179 and D21S11 (McCord 2005). But these redesigned primers might still be useful in checking for potential allele dropout caused by primer-binding site mutations. Table 1.2 shows the 5 miniplexes sets that were designed, each consisting of 3 loci each labeled with a different dye colour. The fifth multiplex consists of 3 non-CODIS loci.

Table 1.2: The 5 Miniplex sets with the difference in size of the amplified miniSTR compared to the commercial multiplex kits.

Dye Label	**Blue**	**Green**	*Yellow*
Miniplex 1	**THO1 -105**	**CSF1PO -191**	**TPOX -148**
Miniplex 2	**D5S818 -53**	**D8S1179 -37**	**D16S539 -152**
Miniplex 3	**FGA -71**	**D21S11 -33**	**D7S820 -117**
Miniplex 4	**VWA -64**	**D18S51 -151**	**D13S317 -105**
Miniplex 5	**Penta D -282**	**Penta E -299**	**D2S1338 -198**

Non-CODIS MiniSTRs

The major advantage of the CODIS miniSTRs is that database compatibility can be kept. But a few of the CODIS loci cannot be made into smaller amplicons. The flanking sequences for some of these markers contain polymorphic nucleotides, partial repeats, mononucleotide repeat stretches or insertions/deletions that could prevent stable primer annealing (Butler *et al.*, 2003). As a result, not all primers were designed to bind a few nucleotides away from the core repeat region. For example, FGA has a partial repeat and a mononucleotide repeat stretch TTTC TTCC TTTC TTTTTT immediately downstream of the repeat region. This impacts on the position of the reverse primer, causing the 3' end of the FGA miniSTR reverse primer to be located 23 nucleotides away from the end of the core repeat unit (Butler *et al.*, 2003). Another example would be D7S820. The reverse primer of this locus is located furthest away from the marker's repeat region because of a poly-T stretch 13 nucleotides downstream of the repeat region. This poly-T stretch has been shown to contain eight, nine or ten Ts (Egyed *et al.*, 2000). Consequently, the reverse primer for D7S820 was designed to be downstream of the poly-T stretch.

For these reasons, Coble and Butler (2005) outlined the development of novel Non-CODIS miniSTR loci that can be useful for the analyses of degraded DNA. These loci are not meant to replace the existing CODIS markers, but instead complement them in cases where additional loci are needed to increase the possibility of identifying remains when only a partial profile is obtained with the current CODIS markers. The selected markers are therefore genetically unlinked to the CODIS loci. The selection process for these non-CODIS loci begin with the screening of 920 previously characterised loci.

Ideal miniSTR markers should have a narrow allele range, have heterozygosity values of 0.7 or greater, and have "clean" flanking sequences that can generate amplicon lengths of ~100bp. Fig. 1.6 shows a schematic diagram outlining the screening process for identifying these non-CODIS markers. Of the 920 loci, 107 met the criteria for creating miniSTRs. These loci had an allelic spread of <24bp, observed heterozygosities of >0.7, and contained tri- or tetranucleotide repeats. From the 107 loci, 61 had "clean" flanking regions allowing primers to be close to the repeat region. 47 of these markers were selected for having amplicon sizes less than 125bp. From these 47 loci, 18 candidates were selected for initial testing and 6 of these loci were focused on for further characterisation. These markers are D1S1677, D2S441, D4S2364, D10S1248, D14S1434 and D22S1045. The D22S1045 locus consists of a trinucleotide repeat whereas all the other loci consist of tetranucleotide repeats. Two of these loci are found on chromosomes that have CODIS loci. TPOX is on chromosome 2 and FGA on chromosome 4. However with the use of BLAT sequence alignment tool available at http://genome.ucsc.edu/cgi-bin/hgBlat it was found that D2S441 and D4S2364 were more than 50 centimorgans from TPOX and FGA respectively. Implying that these markers are probably unlinked. The six loci selected for characterisation were arranged into two miniplexes; NC01 consisted of D10S1248, D14S1434 and D22S1045, and NC02 consisted of D1S1677, D2S441 and D4S2364. The NC01 Miniplex performed well in the EDNAP/ENFSI collaboration study on degraded DNA analysis (Dixon *et al*, 2006) and has been recommended for inclusion in the next generation STR multiplex kit for Europe (Gill *et al*, 2006).

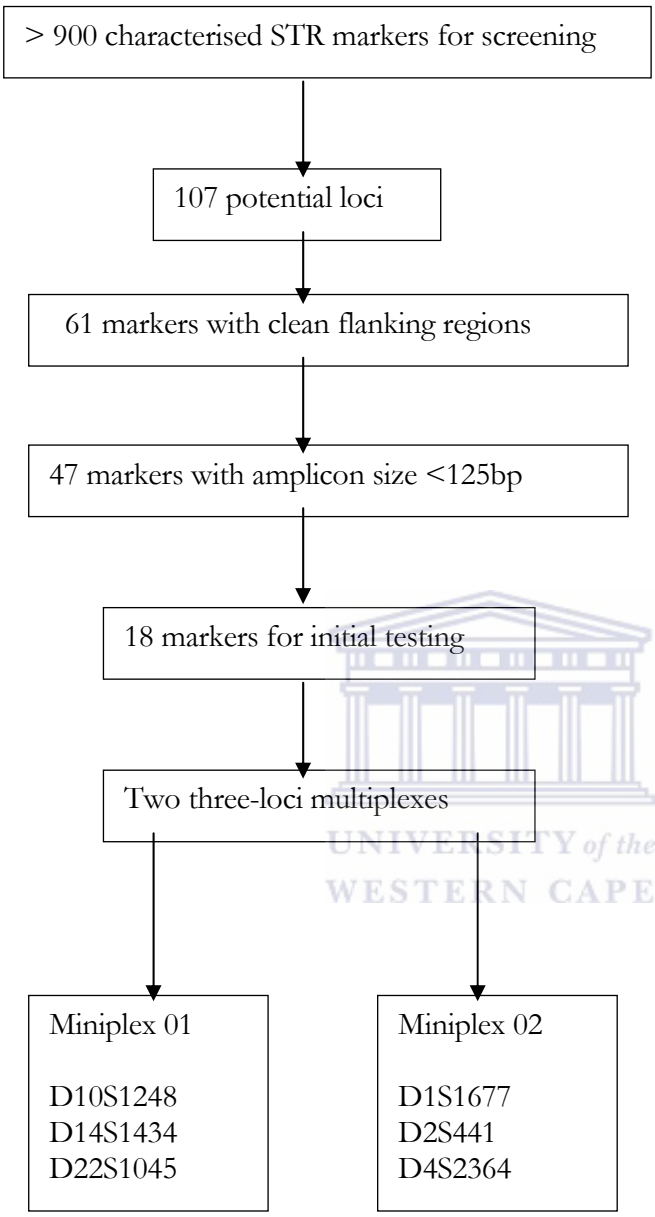


Fig. 1.6: shows a schematic diagram outlining the screening process for identifying new miniSTR markers. 920 previously characterised STRs were screened for ideal markers to make small amplicons. Adapted from Coble and Butler, 2005.

Objective of this study

Additional miniSTR loci will serve as useful complements to the CODIS loci and aid in the analysis of degraded DNA. These new loci will be valuable in a variety of scenarios, particularly for paternity cases, missing persons work, or mass fatality DNA identification testing.

However, before these markers can be used in casework we have to determine how well these markers work in South African populations. Therefore the objective of this study was to characterize eight NC-miniSTRs in South African populations as follows, in order to aid the analysis of degraded DNA:

1. To estimate the levels of diversity of these markers in South African populations
2. To estimate the population genetic parameters of the population groups
3. To study the inter-population relationships

Chapter 2

Materials and Methods

Loci Selection

The NC01 (Non-CODIS Set 01) consisting of the D10S1248, D14S1434, D22S1045 loci and NC02 (Non-CODIS Set 02) consisting of the D1S1677, D2S441, D4S2364 loci were selected for this study. In addition, D9S2157 and D12ATA63 loci were selected from the remaining 20 NC-miniSTRs. Primers were used as published in Hill *et al.*, 2006. The loci information and GenBank® (http://www.ncbi.nlm.nih.gov/) accession numbers for each marker are listed in Table 2.1.

PCR Primers and Other Reagents

Fluorescently labeled primers, Supertherm Gold® DNA polymerase and associated buffers were obtained from Applied Biosystems (Foster City, CA). All forward miniSTR primers were labeled with 6-FAM™, VIC™, NED™ and PET™ dyes (Applied Biosystems). These dyes allowed the use of available matrix standards and colour separations on the ABI377 Genetic Analyzer. Table 2.2 provides information on the miniSTR primers used in this study.

DNA Extraction

Chelex Extraction Procedure

The Chelex extraction procedure was used according to the protocol for DNA extractions from oral swabs (Appendix I). This is a fast, cheap and effective method of DNA extraction. Because DNA extraction is the first step towards amplification of the template, excellent sterile technique must be maintained. Also, since the Chelex beads are left in the extraction mixture; care should be

taken to only use the supernatant for PCR reactions, as the Chelex beads will inactivate the Taq enzyme.

Table 2.1: Information on eight miniSTRs used in this study; the chromosomal location of each marker was determined using BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat). The observed allele ranges and the repeat motifs are from this study. The observed size ranges (bp) and 9974A Standard DNA genotype is based on the miniSTR amplicon sizes that were observed in this study relative to the GS500 LIZ size standard. The GenBank® Accession numbers were obtained from nucleotide searches through NCBI (http://ncbi.nlm.nih.gov/)

Locus Name	Chromosomal Position	GenBank® Accession	Observed Allele Size (bp)	9974A Standard DNA genotype	Allele Range	Repeat Motif
D1S1677	Chr 1 160.747 Mb	AL513307	83-115	13,14	10-18	$(TTCC)_n$
D2S441	Chr 2 68.214 Mb	AC079112	80-108	10,14	9-16	$(TCTA)_n$
D4S2364	Chr 4 93.976 Mb	AC022317	74-86	9,10	8-11	$(GAAT)(GGAT)(GAAT)_n$
D9S2157	Chr 9 133.065 Mb	AL162417	76-106	7,13	7-17	$(ATA)_n$
D10S1248	Chr 10 130.567 Mb	AL391869	93-121	13,15	11-18	(GGAA)n
D14S1434	Chr 14 93.298 Mb	AL121612	76-108	11,13	10-18	$(GATA)_n(GACA)_n$
D22S1045	Chr 22 35.779 Mb	AL022314	84-108	11,14	10-18	$(ATT)_n$
D12ATA63	Chr 12 106.825 Mb	AC009771	84-111	13,13	11-20	$(TAA)(CAA)_n$

Table 2.2: MiniSTR primers used in this study. The oligo distance from repeat refers to the distance of the 3' end of the primer to the repeat region. A negative number indicates that the primer is in the repeat region (with the nucleotides underlined at the 3' end of the primer).

Marker Name	Primer Sequence (5'-3')	Primer Concentration (µM)	Oligo Distance from Repeat (bp)
D1S1677	for [NED] - TTCTGTTGGTATAGAGCAGTGTTT rev GTGACAGGAAGGACGGAATG	0.1	-2 2
D2S441	for [VIC] – CTGTGGCTCATCTATGAAAACTT rev GAAGTGGCTGTGGTGTTATGAT	0.1	-1 1
D4S2364	for [FAM] – CTAGGAGATCATGTGGGTATGATT rev GCAGTGAATAAATGAACGAATGGA	0.1	2 -7
D9S2157	for [PET] – CAAAGCGAGACTCTGTCTCAA	0.5	1

	rev GAAAATGCTATCCTCTTTGGTATAAAT		6
D10S1248	for [FAM] – TTAATGAATTGAACAAATGAGTGAG rev GCAACTCTGGTTGTATTGTCTTCAT	0.1	1 0
D14S1434	for [VIC] – TGTAATAACTCTACGACTGTCTGTCTG rev GAATAGGAGGTGGATGGATGG	0.05	-11 0
D22S1045	for [NED] – ATTTTCCCCGATGATAGTAGTCT rev GCGAATGTATGATTGGCAATATTTTT	0.1	1 5
D12ATA63	for [PET] – GAGCGAGACCCTGTCTCAAG rev GGAAAAGACATAGGATAGCAATTT	0.1	0 7

DNA Extraction using QIAamp® Blood kit

DNA samples from blood were extracted using the QIAamp® Blood kit (Qiagen Inc.). These kits use silica-gel membrane to bind the DNA template. DNA samples absorb to the silica membrane in the presence of high-salt concentrations and low pH. Elution of DNA from the membrane is achieved under basic conditions and low salt concentrations. The template is eluted with the elution buffer provided in the kit (QIAGEN Worldwide, 2001). The step-by-step protocol for this method can be found in Appendix I.

Source of DNA for Population Testing

Anonymous liquid blood samples with self-identified ethnicities were previously obtained and extracted as described in the section above. A subset of 358 samples representing four major South African populations consisting of 103 Afrikaner, 105 Mixed Ancestry, 100 Asian Indian and 60 Xhosa were used in this study.

DNA Quantification

Nanodrop® ND-1000 Spectrophotometer V3.3.0

The Nanodrop® ND-1000 is a full-spectrum (220-750nm) spectrophotometer that measures 2µl samples with high accuracy and reproducibility. It utilizes a patented sample retention technology that uses surface tension alone to hold the sample in place. This eliminates the need for cuvettes and other sample containment devices and allows for clean up in seconds. In addition, the ND-1000 has the ability to measure highly concentrated samples without dilution (75X higher concentration than the samples measured by a standard cuvette spectrophotometer) (Nanodop® Technologies, 2004). The Nanodop® spectrophotometers consist of numerous programs and modules for analysis. The Nucleic Acid Measurement Module was used to check the concentration and quality of the DNA samples. A screenshot is shown in Fig. 2.1.

The "Sample Type" tab used to select the type of nucleic acid being measured. The user can select "DNA-50" for dsDNA, "RNA-40" for RNA or "Other" for other nucleic acids. The default is DNA-50. λ and Abs refers to the user selected wavelength and corresponding absorbance. The "A260" tab is the absorbance of the sample at 260 nm represented as if measured with a 10 mm path and the "A280" tab is the sample absorbance at 280 nm represented as if measured with a 10 mm path. "260/280" refers to ratio of sample absorbance at 260 and 280 nm. The ratio of absorbance at 260 and 280 nm is used to assess the purity of DNA and RNA. A ratio of ~1.8 is generally accepted as "pure" for DNA; a ratio of ~2.0 is generally accepted as "pure" for RNA. If the ratio is appreciably lower in either case, it may indicate the presence of protein, phenol or other contaminants that absorb strongly at or near 280 nm.

"260/230" refers to the ratio of sample absorbance at 260 and 230 nm. This is a secondary measure of nucleic acid purity. The 260/230 values for "pure" nucleic acid are often higher than the respective 260/280 values. They are commonly in the range of 1.8-2.2. If the ratio is appreciably lower, this may indicate the presence of co-purified contaminants. Lastly, the "ng/ul" tab shows the sample concentration in ng/ul based on absorbance at 260 nm and the selected analysis constant.

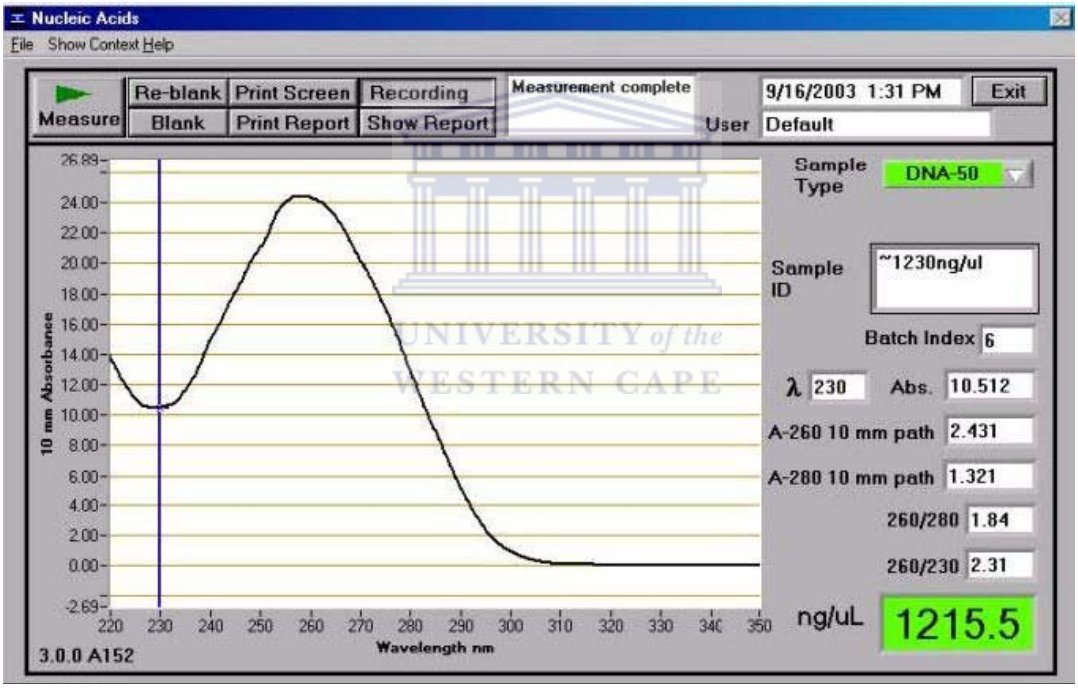


Fig. 2.1: Screenshot of Nucleic Acid Measurement Module of the The Nanodop® spectrophotometers (Nanodop® Technologies, 2004)

DNA Amplification

The standard conditions and reaction components used for amplification consist of: 10X relevant primer mix, 10X Supertherm PCR Buffer (With 15mM $MgCl_2$), 2mM dNTPs, 4mg/ml BSA, 0.5ng of template DNA at 1ng/µl and 5 units Supertherm Gold® DNA polymerase in a final reaction volume of 10 µl using individual 0.2ml MicroAmp Reaction tubes with caps (Applied Biosystems). The thermal cycling parameters are 95°C for 10 minutes (hot-start), 35 cycles at 94°C for 1 minute (denature), 55°C for 1 minute (anneal), and 72°C for 1 minute (extend), followed by a final incubation of 60°C for 45minutes and 4°C on hold.

Electrophoresis and Data Analysis

The Miniplex sets use a five-dye system of 6-FAM™ (blue), VIC™ (green), NED™ (yellow), PET™ (red) and LIZ™ (orange) (Applied Biosystems). Amplicons were analysed with the ABI PRISM® 377 Genetic Analyzer according to the manufacturer's instructions (Appendix I) using the filter set G5. Internal lane size standard was included with every sample to allow automatic sizing of alleles and to normalise differences in electrophoretic migration of samples between gel lanes. Samples, run on either 48 or 64 lane gels, were prepared by combining 1µl of PCR product to 1µl of load mix consisting of GeneScan 500™ [LIZ] Internal Lane Standard (Applied Biosystems), blue loading dye and deionised formamide at a ratio of 0.5:1:5. After denaturation for 3 min at 95^{0}C and snap cooling, 1µl of sample containing DNA fragments ranging from 74-121bp was loaded onto a 3% polyacrylamide gel. Gels were electrophoresed for 2.5hrs at 2400V. Data were analysed at a minimum threshold of 150RFU and a baseline of 21 using GeneScan® (V.3.7/NT) and Genotyper® (V.3.7/NT) from Applied Biosystems.

Allelic Ladders and Genotyper Macros

Allelic ladders were created by mixing and amplifying samples previously typed so as to include all observed alleles in the population samples. A 1:1000 dilution of amplified mixed samples was prepared and then 2µl of this dilution were amplified individually for each set of primers according to the PCR protocol outlined in the DNA amplification section. However, the final extension was increased to 90 min since there are more PCR products to generate. Genotyper macros were constructed for each of the miniSTR loci using fixed bin allele sizes to perform allele designations. Bin width ranged from 0.5 to 0.7bp for each marker. The number of repeats was calibrated to allele size by sequencing.

DNA Sequencing

At least two homozygote samples of each marker were sequenced to calibrate the number of repeats. Sequencing primers located at least 100bp away from the STR repeat region was designed using web-based Primer3 (Rozen and Skaletsky, 1998). Typical default parameters were used: primer Tm values from 57-63^{0}C with 60^{0}C as the optimum; minimum primer size of 18bp and the primer GC% range from 20-80. The resulting primers for each locus are provided in Table 2.3. There are numerous advantageous of using the sequencing primers instead of the reduced-sized amplification primers. These include the ability to fully sequence the repeat as well as sequence the primer binding region to detect nucleotide mutations that can create null alleles. PCR for sequencing was performed separately for each non-labelled primer set. The PCR conditions were as follows: 95ºC for 10 min; 35 cycles of 94ºC for 1 min; 57ºC for 30 sec; 72ºC for 30 sec; and a final incubation of 60ºC for 45 min. The BigDye v1.1 Ready Reaction Kit (Applied Biosystems) was used for sequencing reactions with identical primers after

purification (See Appendix I). Electrophoresis was carried out using an ABI PRISM® 377 Genetic Analyser according to the manufacturer's instructions (Appendix I). Sequencing was determined with Sequencing Analysis Software 5.1.1 (Applied Biosystems) and Bio-edit (Version 7.0.9).

Table 2.3: Unlabelled primers used for sequencing. These primers bind outside of the miniSTR primer region and will help detect any miniSTR primer binding site mutations which may cause allele dropout.

Locus Name	Sequence
D1S1677	for GTAGTGCTGGTGCAGCGTAA rev TGCAATAGCAAATATCAGAATGTGT
D2S441	for CTGTTCCTGAGCCCTAATGC rev CACCACACCCAGCCATAAAT
D4S2364	for TGTTGTCTGTAGGAGCTGAGAAA rev GGTGTTTGGAGATGGCTGTT
D10S1248	for AGCAAACCTGAGCATTAGCC rev AGTGCTTGGCAAAGAGCAGA
D14S1434	for TTCCCAGCCTCCATAATCAG rev TGCAAATGCACACAGATTTC
D22S1045	for CCCACTATGGGCAAACCTTA rev TGTGCTTCAGTCTCCTCAGC
D9S2157	for AAAATTAGCGTGTGTGCCTGTAA rev ACATCAGTCACTCATTCATGTGC
D12ATA63	for AAAAACTAACCTGGCATGCTG rev GGGCAGAAACATTGCTAAGAATAG

Statistical Analysis

Allele frequency is a concept used to quantify genetic variation. It is defined as a measure of the commonness of a given allele in a population, that is, the proportion of all alleles of that locus in the population that are specifically this type. To study population genetics, allele and genotype frequencies must be calculated, and how these frequencies change from one generation to the next. Populations are dynamic due to changes in birth and death rates, migration or interbreeding with other populations. This dynamic nature of populations has important consequences and can, over time, lead to changes in the population gene pool. The large variation in the frequency of specific alleles raises a number of questions. For example, can we expect the allele to increase in populations in which it is currently rare? Such questions can be explored using a mathematical model developed by the British mathematician Godfrey H Hardy and the German physician Wilhelm Weinberg. The model was accordingly called the Hardy-Weinberg Equilibrium and states that both allele and genotype frequencies in a population remain constant or are in equilibrium from generation to generation unless specific disturbing influences are introduced. Those disturbing influences include non-random mating, mutations, selection, limited population size, random genetic drift and gene flow. Hardy-Weinberg principle provides a measurement by which changes in allele frequency can be measured.

Hardy-Weinberg equilibrium was determined by an exact test using GENEPOP (Version 4.0) software package designed by Raymond and Rousset (1995). All other statistical values were calculated using the PowerStats V12 spreadsheet (Promega Corporation). Allele frequencies, heterozygosity values, p-values from the Hardy-Weinberg test, power of discrimination, polymorphic information content (PIC) and the probability of paternity exclusion for each locus are listed in Tables 3.1 – 3.17. The STR data from this study was also used to ascertain

phylogenetic relationships between the four populations. A phylogenetic analysis was executed based upon the STR allele frequencies obtained, using Populations V1.2.30 available at: http://bioinformatics.org/project/?group_id=84. Distance was estimated using the formula of Saitou and Nei (1987) and a phylogeny was inferred by the neighbor joining option. The phylogenetic tree was subsequently viewed using Tree View V32.

Results and Discussion

3.1.1 Electrophoresis

The objective of this study was to analyse human genetic variation within 362 South African samples comprised of Afrikaner (n=103), Mixed ancestry (n=105), Asian Indian (n=100) and Xhosa (n=60). Eight NC-miniSTRs were selected from the potential 26 suggested by Hill *et al.*, 2006. These loci were amplified in two multiplexes. The fluorescently labeled PCR products were detected by laser-induced fluorescent excitation gel electrophoresis. Allele sizes were calculated automatically by comparison to an in-lane size standard, LIZ™ (orange). Size (bp) versus relative fluorescent units (RFU) was plotted automatically and displayed as an electropherogram (Fig. 3.1). Size resolution to the nearest bp is usually possible under good conditions.

3.1.2 Stutter Artifacts

A "stutter peak" is a small non-allelic peak that appears on an electropherogram 1-3bp before an allelic peak (Balding, 2005). Stutter is the result of PCR products that have one, two or three repeat units less than the main allele. Fig. 3.2 shows an electropherogram of the D12ATA63 locus. The two alleles are shown, allele 13 (90.35bp in size) and allele 16 (99.22bp in size). Adjacent to each allele is a very small peak 3bp less. Since both alleles are detected at this locus, the possibility of allelic dropout does not exist. In addition, extreme imbalance such as this is very unlikely to occur under normal conditions. Therefore, the very small peaks are presumed to be stutter artifacts. Stutter is generally the result of PCR, caused by mispairing of strands during replication, although some stutter peaks might be the result of somatic mutations. Stutter usually

has less than 10% of the height of the main peak; however the average height can increase with allele length (Butler, 2000). Stutter peak height usually decreases with increasing length of the repeat unit. The D9S2157, D22S1045 and D12ATA63 loci have trinucleotide repeat structures, and all the other loci used in this study have tetranucleotide repeats. However, the stutter percentage for the trinucleotide loci was not observed to be more than that of the tetranucleotide loci.

3.1.3 Allele Designation

Allele typing results were obtained from 500pg of human genomic DNA as PCR template; some results were also obtained from templates containing less than 500pg of DNA. Allele designations were assigned by comparing sample allele sizes to the sizes obtained for alleles in the allelic ladder. Allelic ladders (Fig. 3.3) contained common alleles for each locus labeled with the same dye as the sample alleles, and were run alongside population samples on each gel. DNA sequencing determined the length and repeat structure of each allelic ladder allele.

DNA samples were thus screened for their allelic combination and allele sizes. The samples were then analysed for total number of alleles at a particular locus, their frequency, heterozygosity values, PIC, power of discrimination, probability of paternity exclusion, F-Stat and genetic distances. The data were also used to determine Hardy-Weinberg equilibrium (HWE).

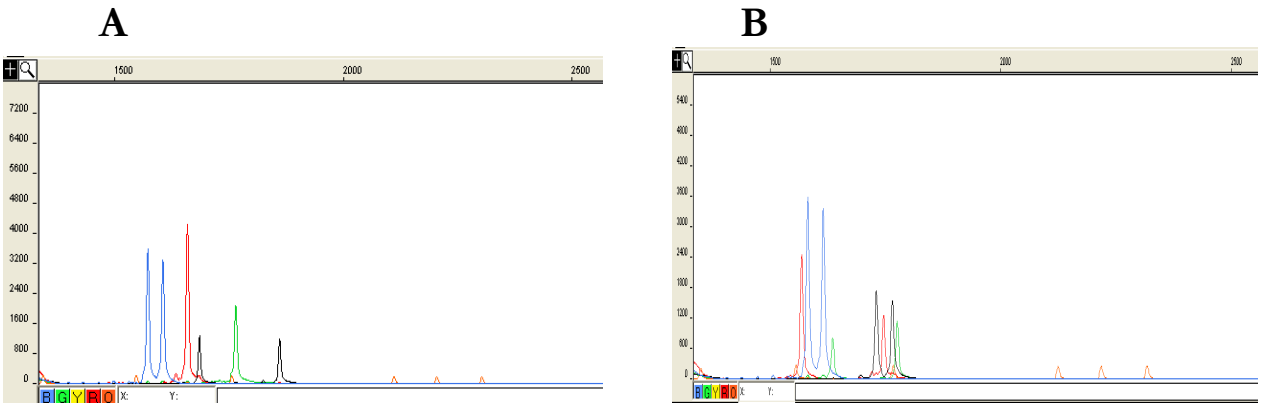
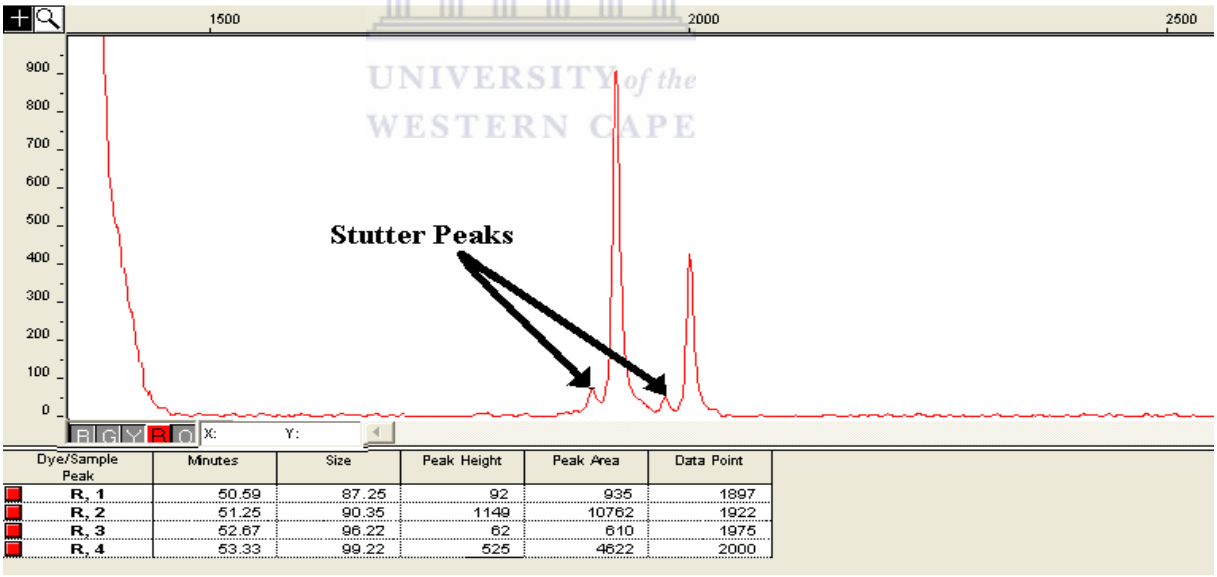
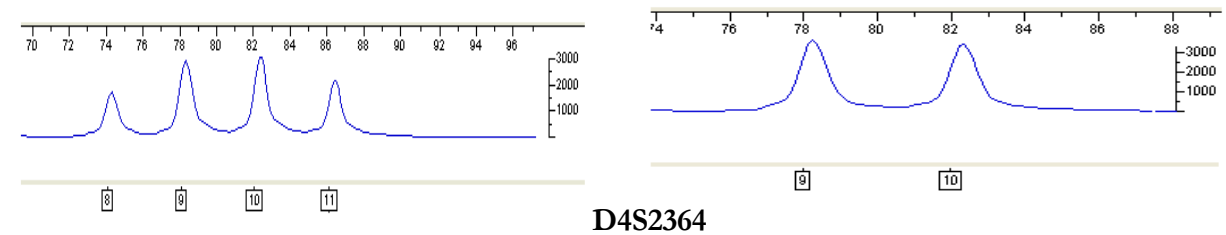


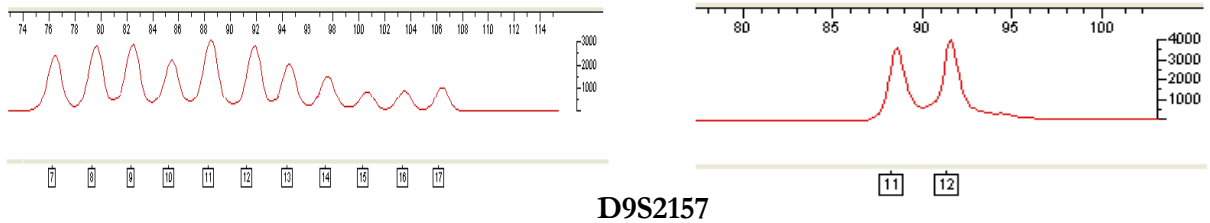
Fig. 3.1: STR profile of an individual amplified with A) Miniplex 01 consisting of the D10S1248, D14S1434, D22S1045 and D12ATA63 loci; and B) Miniplex 02 consisting of the D1S1677, D2S441, D4S2364 and D9S2157 loci. The amplified fragments were separated on the ABI 377 DNA Sequencer and analysed using GeneScan® and Genotyper® software.

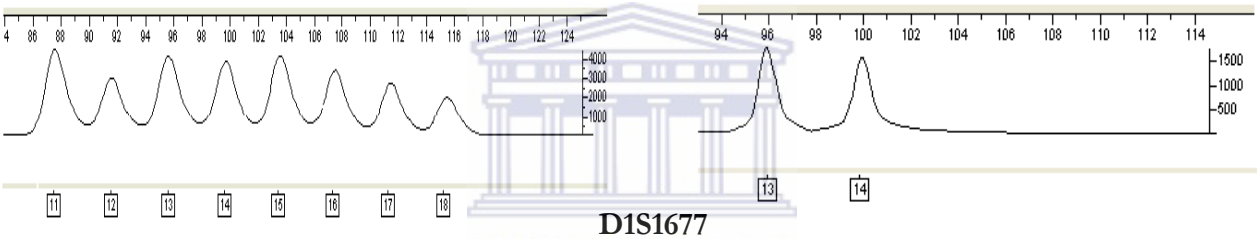


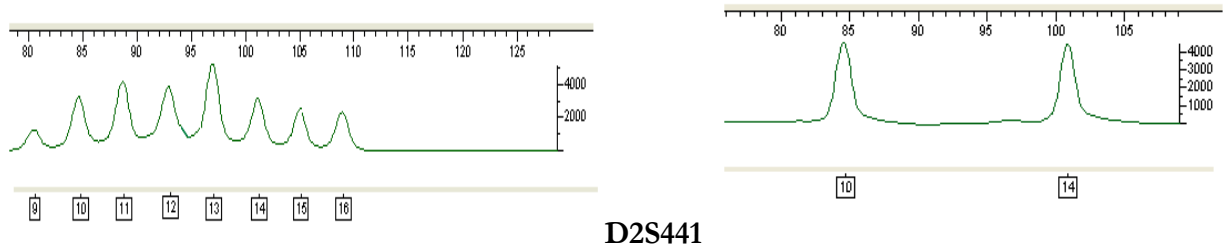
Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
R, 1	50.59	87.25	92	935	1897
R, 2	51.25	90.35	1149	10762	1922
R, 3	52.67	96.22	62	610	1975
R, 4	53.33	99.22	525	4622	2000

Fig. 3.2: Electropherogram showing two D12ATA63 alleles and adjacent to each a stutter artifact 3bp less. Stutter is generally the result of PCR, caused by mispairing of strands during replication, although some stutter peaks might be the result of somatic mutations.









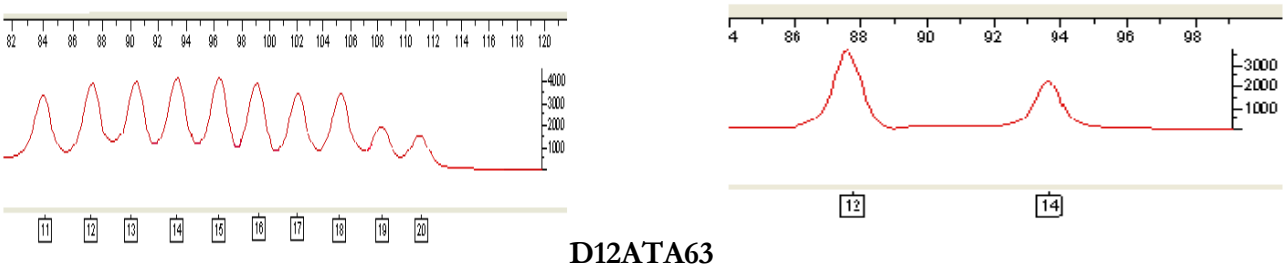


Fig. 3.3: Examples of Genotyper® Software plots of size (bp) versus relative fluorescence units (RFU); and assigned allele designations of the generated allelic ladders (panels on the left in each colour) and one corresponding population sample (right panel).

Microsatellite Characterisation

3.2.1 The D12ATA63 locus

Accession number: AC009771

Chromosomal position: Chr 12 106.825 Mb

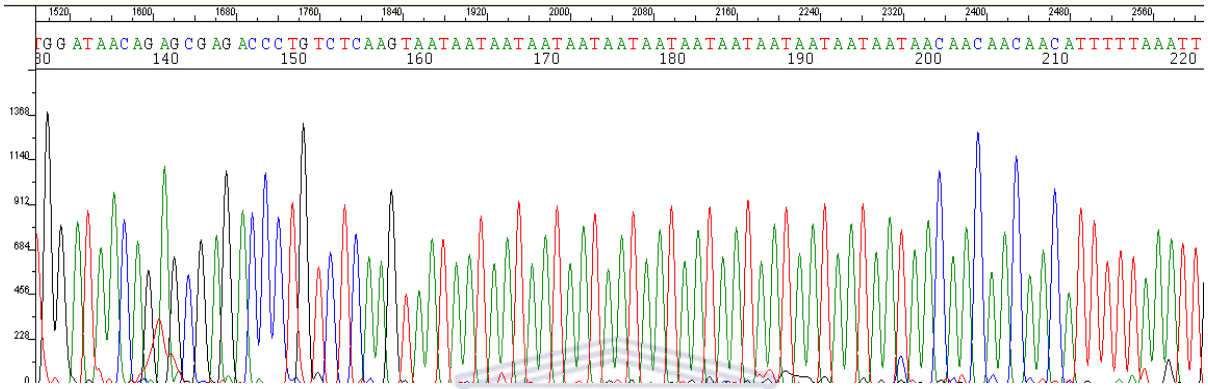


Fig. 3.4: D12ATA63 sequence; showing homozygous individual with 17 [(TAA) 14(CAA)3] repeats

The D12ATA63 locus has an observed allele size of 84 – 111bp and an allele range of 11-20 repeats. This locus is located on chromosome 12 and has the trinucleotide repeat motif (TAA)n(CAA)n (Fig. 3.4). D12ATA63 showed a total of 10 different alleles within the populations, with allele 17 being the most predominant in the Afrikaner population, allele 12 in both Mixed Ancestry and Asian Indian populations, and allele 15 in the Xhosa populations. Hill *et al.*, (2006) observed 11 alleles for this locus within the US populations, though allele 10 had a combined frequency of 0.0008. Afrikaner exhibited eight, Mixed Ancestry and Asian Indian exhibited nine, and Xhosa exhibited seven alleles. While overall form of the STR allele distributions between the four population samples was quite varied (Fig. 3.5), coincidence of the most common allele occurred at allele 12 with a frequency of 17.4%, 31.4%, 33.9% and 27% within Afrikaner, Mixed Ancestry, Asian Indian and Xhosa populations respectively. The most noticeable difference in allele frequencies between the population samples was at allele 17

showing the highest frequency of 26.8% and the lowest being 1%, within the Afrikaner and Xhosa populations, respectively. All populations were in HWE (Table 3 .1).

Table 3.1: Allele frequency distributions of D12ATA63 within four South African populations and HWE test. The highest frequency within each population is highlighted in bold.

Allele	Afrikaner (N=103)	Mixed Ancestry (N=105)	Asian Indian (N=96)	Xhosa (N=54)
11	----	0.015	----	0.010
12	0.174	**0.314**	**0.339**	0.270
13	0.205	0.074	0.056	0.220
14	0.037	0.113	0.067	0.100
15	0.205	0.157	0.133	**0.340**
16	0.053	0.118	0.100	0.050
17	**0.268**	0.113	0.206	0.010
18	0.053	0.093	0.083	----
19	0.005	0.005	0.011	----
20	----	----	0.006	----
HWE *p*-values	0.9968	0.8952	0.9475	0.4265

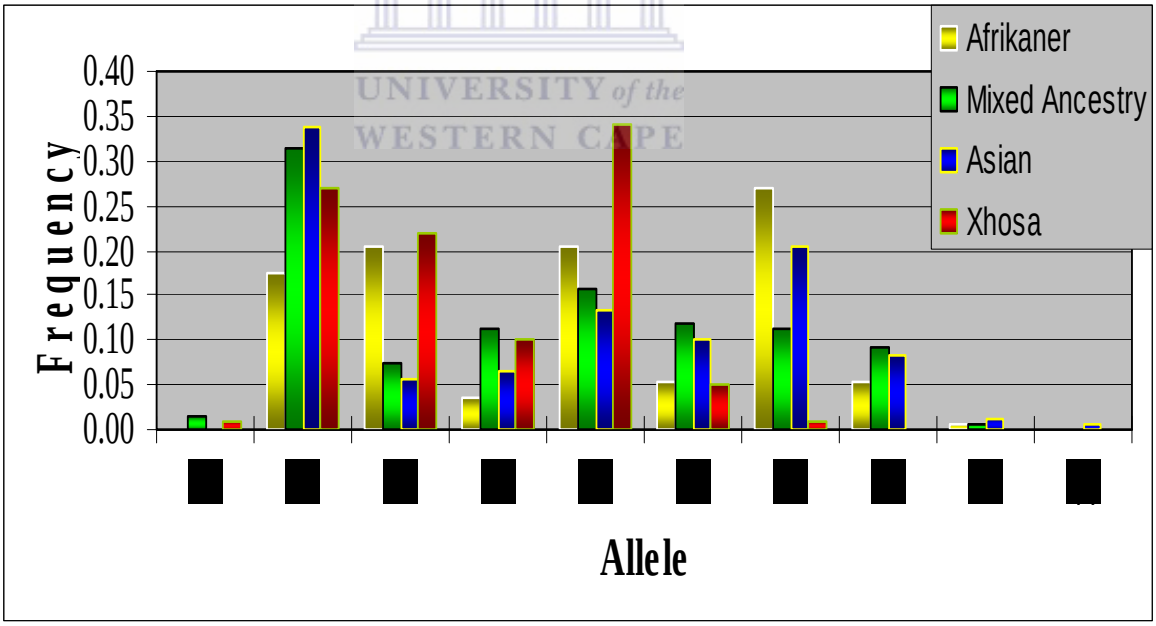


Fig. 3.5: Histogram of allele frequency distribution of the D12ATA63 locus within South African populations

3.2.2 The D22S1045 locus

Accession number: AL022314

Chromosomal position: Chr 22 35.779 Mb

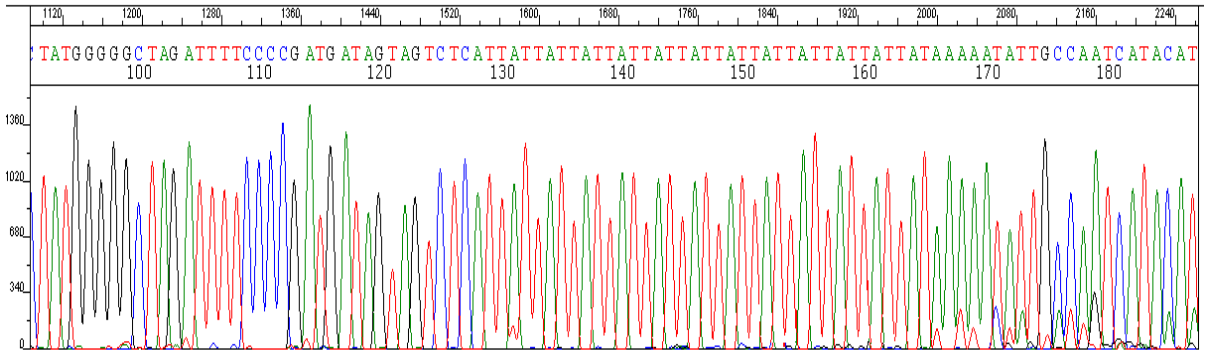


Fig. 3.6: D22S1045 sequence; showing homozygous individual with 12 ATT repeats

The D22S1045 locus has an observed allele size of 84–108bp and an allele range of 10 – 18 repeats. This locus is located on chromosome 22 and has the trinucleotide repeat motif (ATT)n (Fig. 3.6). D22S1045 showed a total of nine different alleles within the populations, with allele 16 being the most predominant in the Afrikaner population, allele 15 in both Mixed Ancestry and Asian Indian populations, and alleles 11 and 15 in the Xhosa populations (Table 3.2).

Coble and Butler (2005) observed 11 alleles within the US populations, with alleles 5, 7, 8 and 9 having combined frequencies of 0.0021, 0.0211, 0.1089 and 0.0317 respectfully.

Afrikaner, Mixed Ancestry and Xhosa exhibited eight alleles, and Asian Indian exhibited nine alleles. The overall distribution of alleles was similar (Fig. 3.7). Coincidence of the most common allele occurred at allele 15 with a frequency of 34.3%, 30.4%, 37.1% and 27.5% within Afrikaner, Mixed Ancestry, Asian Indian and Xhosa populations respectfully. The Afrikaner, Mixed ancestry and Xhosa populations were in HWE, whereas the Asian Indian population showed significant deviation from the equilibrium with a *p*-value of 0.0292.

Table 3.2: Allele frequency distributions of D22S1045 within four South African populations and HWE test. The highest frequency within each population is highlighted in bold.

Allele	Afrikaner (N=103)	Mixed Ancestry (N=105)	Asian Indian (N=96)	Xhosa (N=54)
10	----	0.054	0.005	0.059
11	0.146	0.167	0.323	**0.275**
12	0.005	0.005	0.005	0.020
13	0.010	0.015	0.005	0.059
14	0.066	0.181	0.059	0.039
15	0.343	**0.304**	**0.371**	**0.275**
16	**0.348**	0.196	0.145	0.196
17	0.071	0.078	0.065	0.059
18	0.010	----	0.022	0.020
HWE *p*-values	0.6159	0.8921	0.0292	0.9902

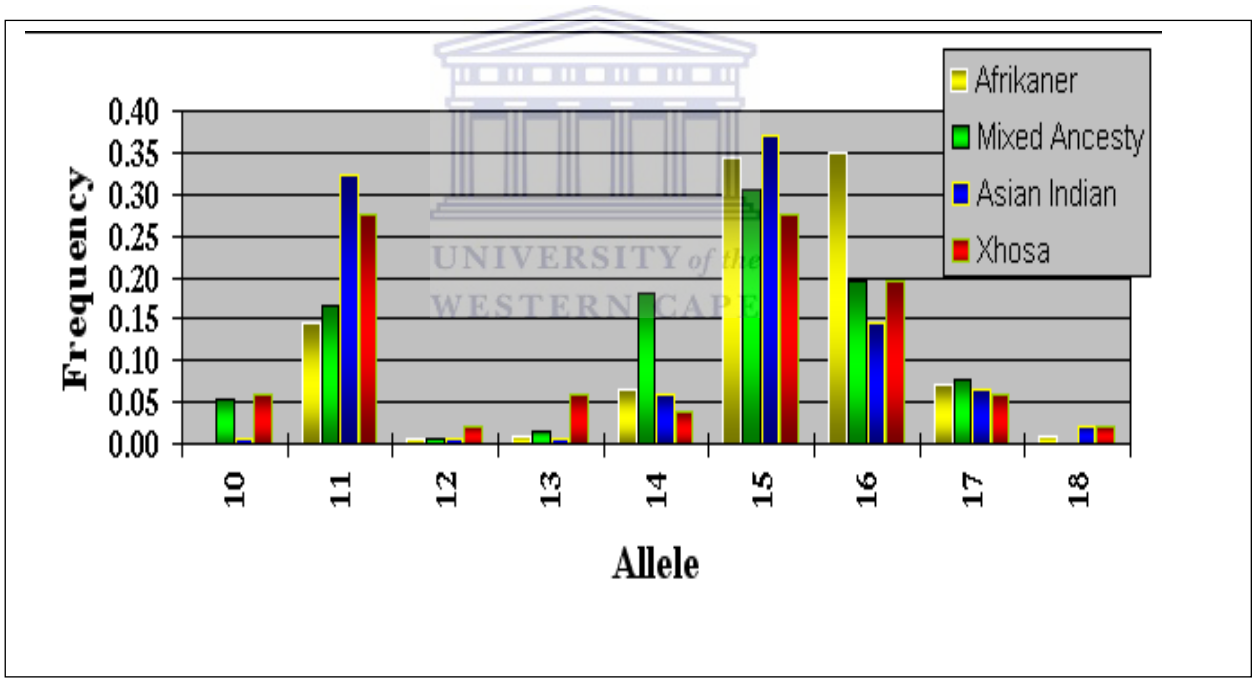


Fig. 3.7: Histogram of allele frequency distribution of the D22S1045 locus within South African populations

3.2.3 The D14S1434 locus

Accession number: AL121612

Chromosomal position: Chr 14 93.298 Mb

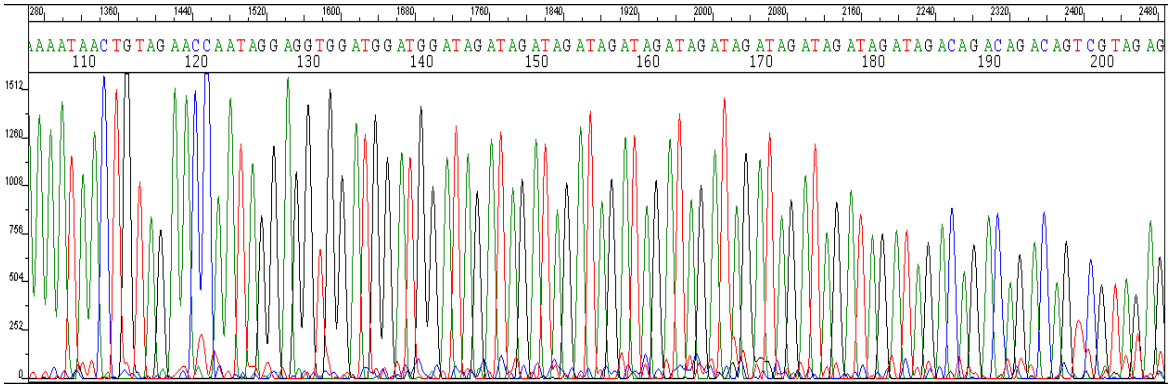


Fig. 3.8: D14S1434 sequence; showing homozygous individual with 14 [(GATA)11(GACA)3] repeats

The D14S1434 locus has an observed allele size of 76 – 108bp and an allele range of 10-18 repeats. This locus is located on chromosome 14 and has the tetranucleotide repeat motif (GATA)n(GACA)n (Fig. 3.8). A total of 10 alleles were observed for this locus across the South African populations. The major alleles for each population were 17 (26.8%) in the Afrikaner, 12 in both the Mixed Ancestry (31.4%) and Asian Indian (33.9%), and 15 (34%) in the Xhosa population. Allele 20 was only observed in the Asian Indian population with a frequency of 0.6%. Yong *et al.*, (2006) observed nine alleles within Singapore populations with the absence of alleles 11-13 and the addition of allele 21 only present within the Indian population. Martìn *et al.*, (2007) observed only seven alleles within the Spanish population and Coble and Butler (2005) observed eight alleles within the US populations. Mixed Ancestry, Asian Indian and Xhosa showed nine, and Afrikaner showed eight alleles. Allele distributions across the populations showed similarity (Fig. 3.9), with alleles 16, 17 and 18 present at the lowest frequencies within all populations.

Coincidence of the most common alleles in all populations occurred at allele 14. All populations showed to be in HWE (Table 3.3).

Table 3.3: Allele frequency distributions of D14S1434 within four South African populations and HWE test. The highest frequency within each population is highlighted in bold.

Allele	Afrikaner (N=103)	Mixed Ancestry (N=105)	Asian Indian (N=96)	Xhosa (N=54)
10	0.175	0.278	0.091	0.186
11	0.035	0.052	0.129	----
12	0.020	0.062	0.032	0.078
13	0.370	0.273	**0.366**	0.314
14	**0.375**	**0.309**	0.360	**0.402**
15	0.010	0.015	0.005	0.020
16	0.010	0.010	0.005	----
17	0.005	----	0.005	----
18	----	----	0.005	----
HWE *p*-values	0.9538	0.9925	0.2906	0.7139

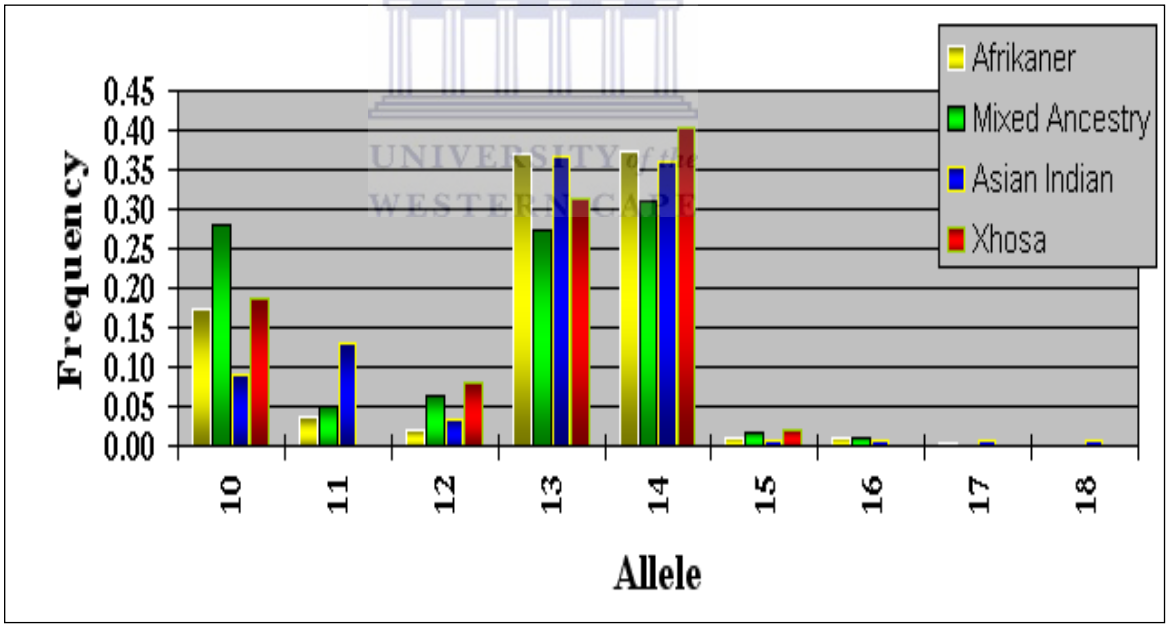


Fig. 3.9: Histogram of allele frequency distribution of the D14S1434 locus within South African populations

3.2.4 The D10S1248 Locus

Accession number: AL391869

Chromosomal position: Chr 10 130.567 Mb

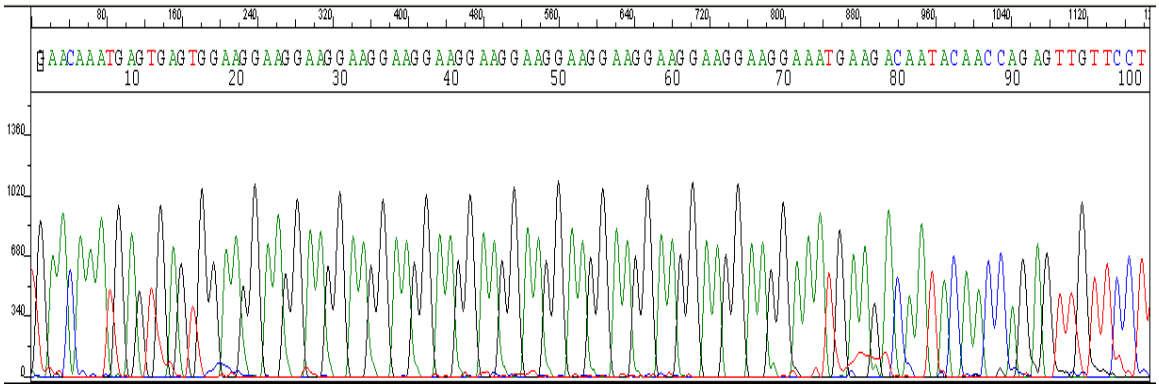


Fig. 3.10: D10S1248 sequence; showing homozygous individual with 14 GGAA repeats

The D10S1248 locus has an observed allele size of 93 – 121bp and an allele range of 11-18 repeats. This locus is located on chromosome 10 and has the tetranucleotide repeat motif (GGAA)n (Fig. 3.10). Eight alleles were present in the four populations. Afrikaner showed 7, Mixed Ancestry, Asian Indian and Xhosa showed 8 alleles. Allele 14 showed the highest frequency across all populations with 33% in Afrikaner, 27.5% in Mixed Ancestry, 30% in Asian Indian and 23.5% frequency in the Xhosa population. Yong *et al.*, (2006) and Martìn *et al.*, (2007) reported 10 alleles with the presence of alleles 10 and 19 in both Spanish and Singapore populations respectively. Coble and Butler (2005) observed 11 alleles within the US populations, and Chung *et al.*, (2007) observed only 7 within the Korean populations. The allelic frequency distributions among the populations (Fig. 3.11) showed great similarity with the lowest frequencies being observed at alleles 17 and 18. The most noticeable difference in allele frequencies between population samples was at alleles 11 and 12 where the Mixed Ancestry and Xhosa frequencies

were significantly greater than that of Afrikaner and Asian Indian samples. All populations were in HWE (Table 3.4).

Table 3.4: Allele frequency distributions of D10S1248 within four South African populations and HWE test. The highest frequency within each population is highlighted in bold.

Allele	Afrikaner (N=103)	Mixed Ancestry (N=105)	Asian Indian (N=96)	Xhosa (N=54)
10	0.175	0.278	0.091	0.186
11	0.035	0.052	0.129	----
12	0.020	0.062	0.032	0.078
13	0.370	0.273	**0.366**	0.314
14	**0.375**	**0.309**	0.360	**0.402**
15	0.010	0.015	0.005	0.020
16	0.010	0.010	0.005	----
17	0.005	----	0.005	----
18	----	----	0.005	----
HWE *p*-values	0.9538	0.9925	0.2906	0.7139

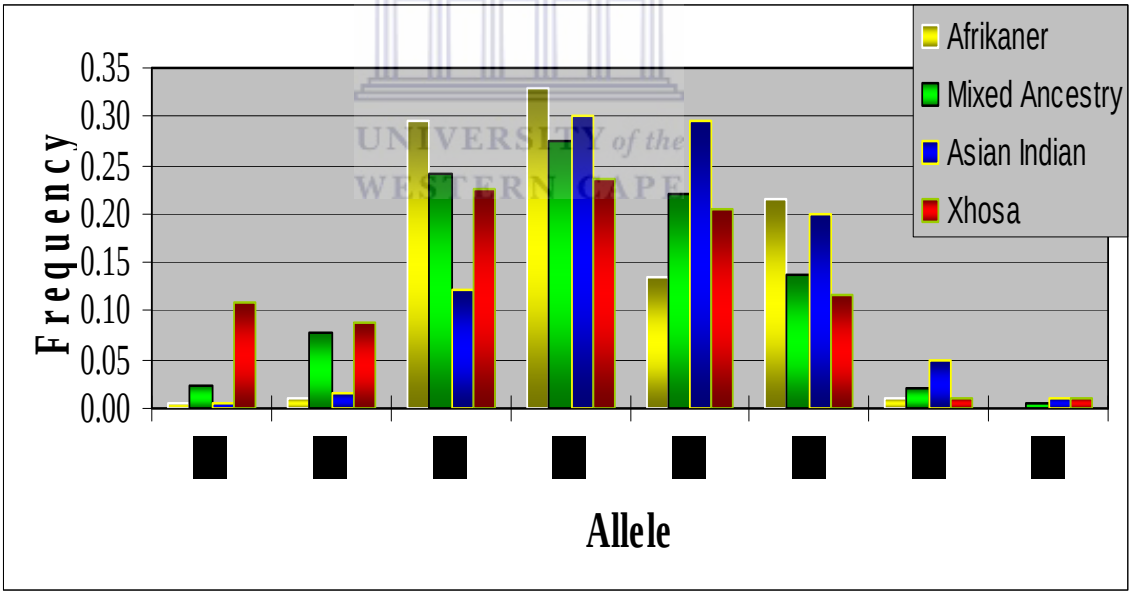


Fig. 3.11: Histogram of allele frequency distribution of the D10S1248 locus within South African populations

3.2.5 The D1S1677 Locus

Accession number: AL513307

Chromosomal position: Chr 1 160.747 Mb

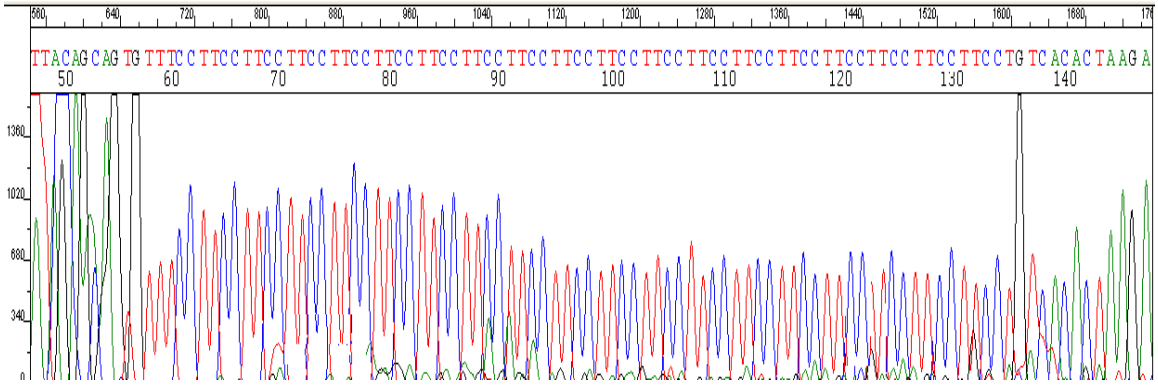


Fig. 3.12: D1S1677 sequence; showing homozygous individual with 19 TTCC repeats

D1S1677 is located on chromosome 1 and has the tetranucleotide repeat motif (TTCC)n (Fig. 3.12). The locus has an observed allele size of 83 – 115bp and an allele range of 10-18 repeats. Total numbers of alleles exhibited in four populations were nine. Allele 14 was the most predominant in Afrikaner (29.2%), Mixed Ancestry (38%) and Asian Indian (38.3%) populations; allele 15 was the major allele (39%) in the Xhosa population. Afrikaner and Xhosa displayed 7, Mixed Ancestry displayed 8, and Asian Indian displayed 9 alleles. Yong *et al.*, (2006) observed 10 alleles within Singapore populations with the presence of allele 9. Nine alleles were observed in US populations (Coble and Butler, 2005), 8 in the Korean populations (Chung *et al.*, 2007) and 7 within the Spanish populations. The allelic distributions within the populations were similar (Fig. 3.13), with alleles 13, 14 and 15 being the most frequent overall. Alleles 10, 11, 17 and 18 were the least frequent, having frequencies of 0.6 – 2.6% across all populations. No significant deviation from HWE was observed (Table 3.5).

Table 3.5: Allele frequency distributions of D1S1677 within four South African populations and HWE test. The highest frequency within each population is highlighted in bold.

Allele	Afrikaner (N=103)	Mixed Ancestry (N=105)	Asian Indian (N=96)	Xhosa (N=54)
10	----	----	0.006	----
11	0.010	0.016	0.006	0.020
12	0.115	0.031	0.056	0.100
13	0.255	0.135	0.189	0.140
14	**0.292**	**0.380**	**0.383**	0.240
15	0.255	0.339	0.283	**0.390**
16	0.047	0.068	0.056	0.100
17	0.026	0.021	0.017	0.010
18	----	0.010	0.006	----
HWE *p*-values	0.6904	0.5171	0.7634	0.1795

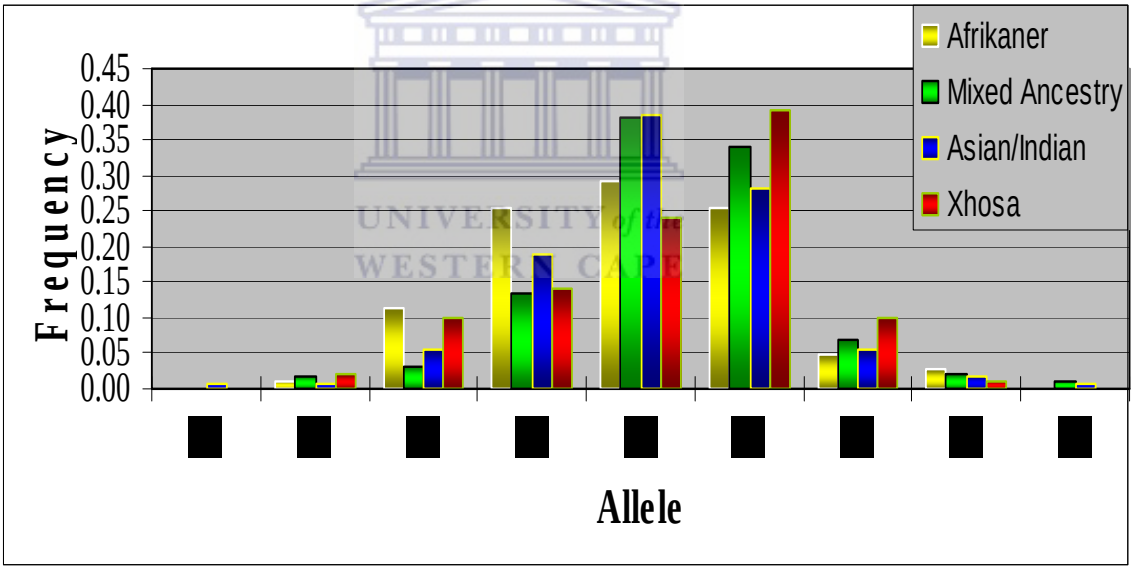


Fig. 3.13: Histogram of allele frequency distribution of the D1S1677 locus within South African populations

3.2.6 The D2S441 Locus

Accession number: AC079112

Chromosomal position: Chr 2 68.214 Mb

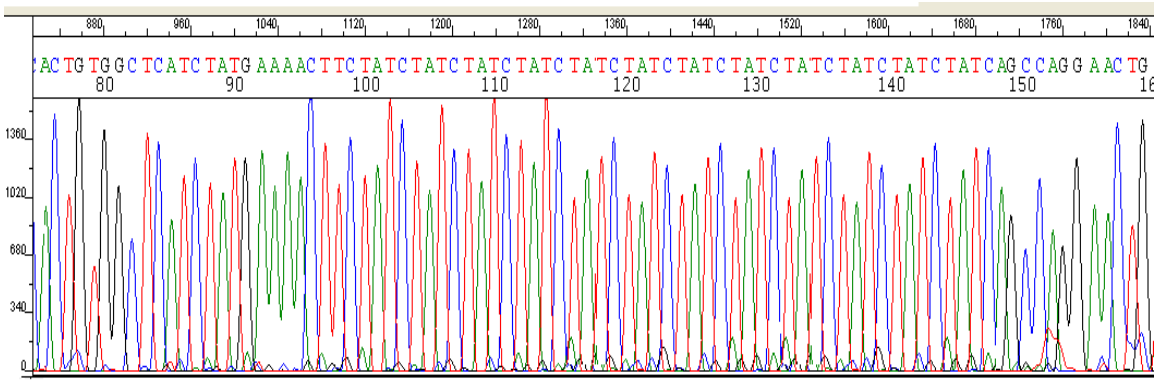


Fig. 3.14: D2S441 sequence; showing homozygous individual with 12 TCTA repeats

The D2S441 locus has an observed allele size of 80 – 108bp with an allele range of 9 – 16 repeats. This locus is located on chromosome 2 and has the tetranucleotide repeat motif (TCTA)n (Fig. 3.14). A total of 11 alleles were observed for this locus, showing maximum frequency for allele 11 across all populations. The frequencies for this allele were 32%, 35.7%, 40.3% and 26% for the Afrikaner, Mixed Ancestry, Asian Indian and Xhosa populations respectively. Afrikaner, Asian Indian and Xhosa displayed 7, and Mixed Ancestry displayed 9 alleles. The distribution of alleles across the populations for this locus was quite diverse (Fig. 3.15). Microvariant repeats were observed in the Mixed Ancestry (alleles 10.2 and 14.1) and Asian Indian (allele 10.1) populations. Allele 9 was only observed in the Afrikaner population with a frequency of 1.5%. Thirteen alleles were observed for this locus in the US populations with the addition of alleles 11.3, 12.3, 13.3, 14.3 and 17 (Coble and Butler, 2005). Allele 9 was present in US Caucasian samples only. All populations were in HWE, with both the Mixed Ancestry and Asian Indian populations having a HWE *p*-Value of 1.000 (Table 3.6).

Table 3.6: Allele frequency distributions of D2S441 within four South African populations and HWE test. The highest frequency within each population is highlighted in bold.

Allele	Afrikaner (N=103)	Mixed Ancestry (N=105)	Asian Indian (N=96)	Xhosa (N=54)
9	0.015	---	---	---
10	0.191	0.230	0.256	0.070
10.1	---	---	0.011	---
10.2	---	0.010	---	---
11	**0.320**	**0.357**	**0.403**	**0.260**
12	0.088	0.143	0.136	0.220
13	0.072	0.031	0.006	0.080
14	0.263	0.148	0.165	0.190
14.1	---	0.046	---	---
15	0.052	0.026	0.023	0.150
16	---	0.010	---	0.030
HWE *p*-values	0.3390	1.0000	1.0000	0.6077

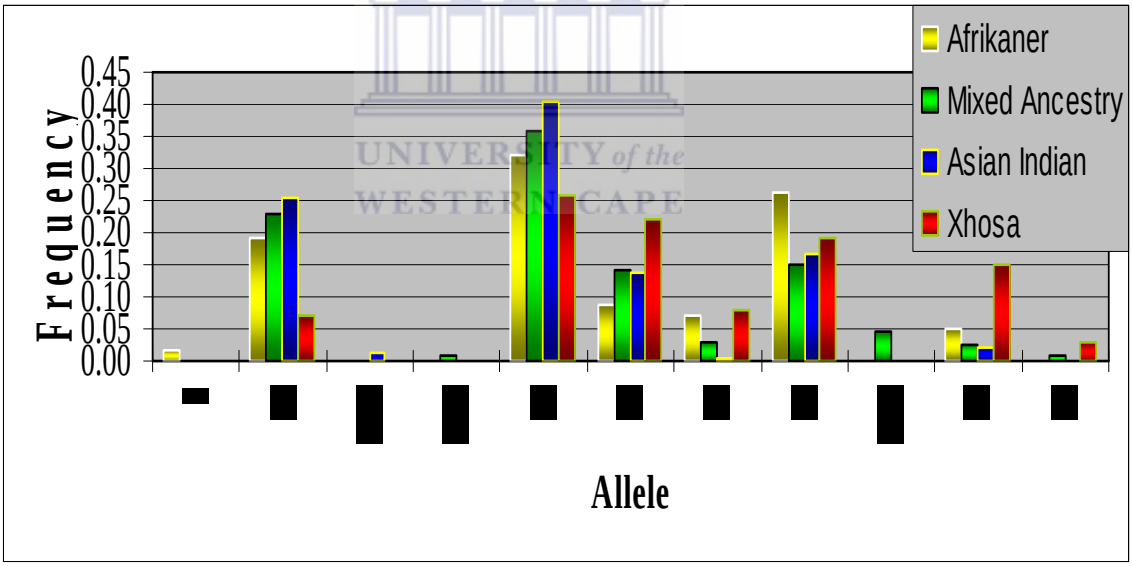


Fig. 3.15: Histogram of allele frequency distribution of the D2S441 locus within South African populations

3.2.7 The D4S2364 Locus

Accession number: AC022317

Chromosomal position: Chr 4 93.976 Mb

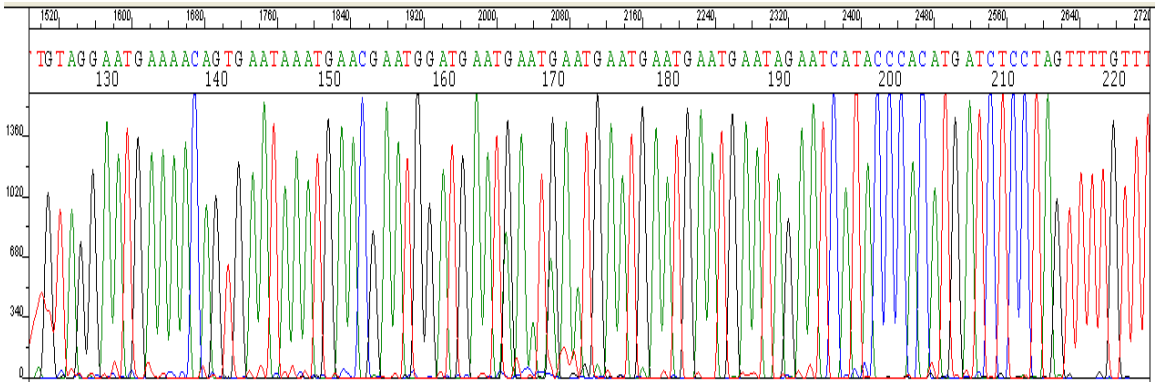


Fig. 3.16: D4S2364 sequence; showing homozygous individual with 9 [(GAAT)(GGAT)(GAAT)7] repeats

The D4S2364 locus has an observed allele size of 74 – 86bp and an allele range of 8 – 11 repeats. This locus is located on chromosome 4 and has the tetranucleotide repeat motif (GAAT)(GGAT)(GAAT)n (Fig. 3.16). The D4S2364 showed the least amount of variation, with only four alleles observed among the population samples. The Afrikaner and Xhosa populations exhibited 3 alleles; and the Mixed Ancestry and Asian Indian populations exhibited 4. Allele 9 was the most predominant in all populations, having frequencies of 68.8%, 60.3%, 58.8% and 86.3% in Afrikaner, Mixed Ancestry, Asian Indian and Xhosa populations respectively. The overall form of the allele distributions between the four samples were similar (Fig. 3.17), especially that of the Mixed Ancestry and Asian Indian populations. Five alleles were observed in the US (Coble and Butler, 2005); Korean (Chung *et al.*, 2007) and Singapore (Yong *et al.*, 2006) populations, though combined frequencies for the additional allele 12 ranged between 0.0011 – 0.008. All populations were in HWE (Table 3.7).

Table 3.7: Allele frequency distributions of D4S2364 within four South African populations and HWE test. The highest frequency within each population is highlighted in bold.

Allele	Afrikaner (N=103)	Mixed Ancestry (N=105)	Asian Indian (N=96)	Xhosa (N=54)
8	0.156	0.172	0.159	0.069
9	**0.688**	**0.603**	**0.588**	**0.863**
10	0.156	0.221	0.225	0.069
11	---	0.005	0.027	---
HWE *p*-values	0.5090	0.8383	0.9767	0.9822

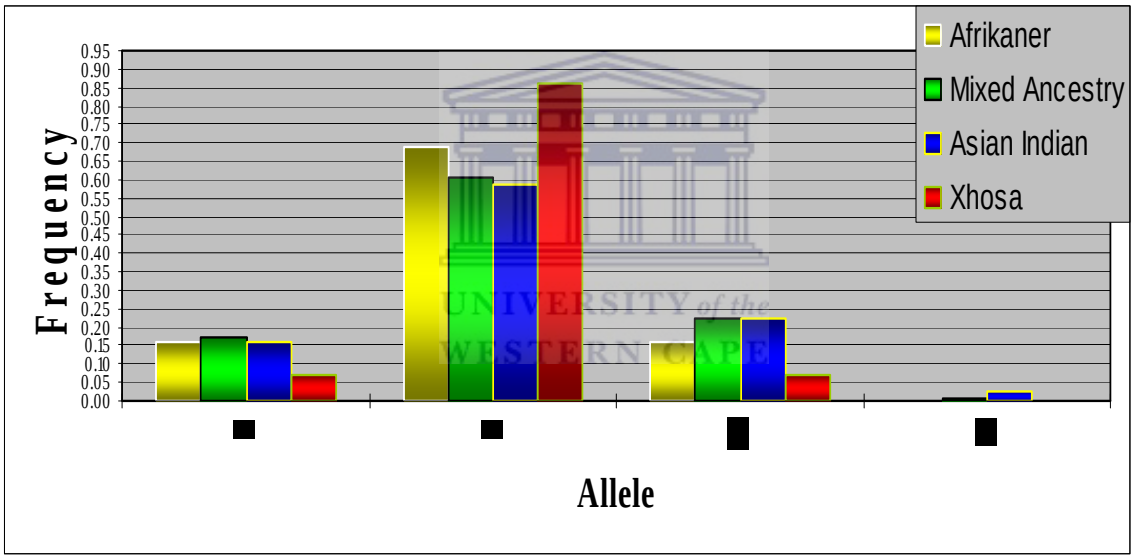


Fig. 3.17: Histogram of allele frequency distribution of the D4S2364 locus within South African populations

3.2.8 The D9S1677 Locus

Accession number: AL162417

Chromosomal position: Chr 9 133.065 Mb

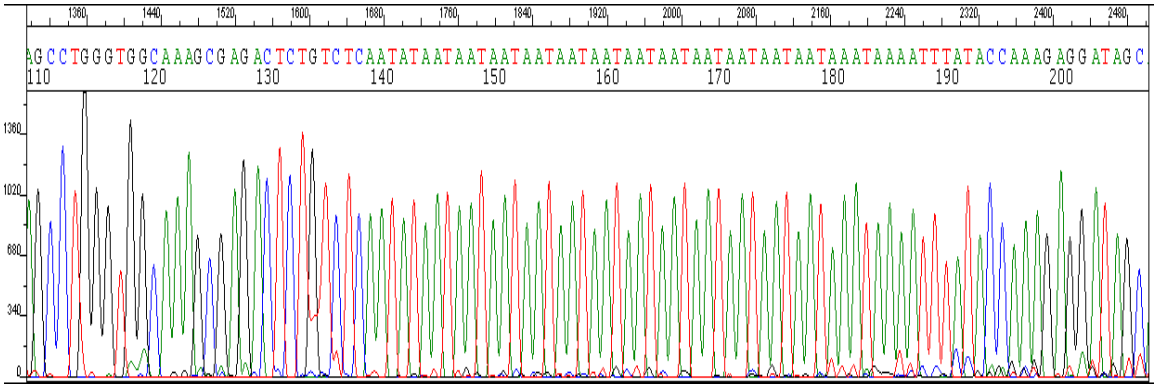


Fig. 3.18: D9S2157 sequence; showing homozygous individual with 13 ATA repeats

The D9S2157 locus has an observed allele size of 76 – 106bp and an allele range of 7 – 17 repeats. This locus is located on chromosome 9 and has the trinucleotide repeat motif (ATA)n (Fig. 3.18). The D9S2157 locus showed the most variation with 12 alleles observed among the populations. The major alleles for each population were 11 for the Afrikaner (32.7%), Mixed Ancestry (18.9%) and Xhosa (38.2%) populations. And allele 13 was the predominant allele in the Asian Indian population with a frequency of 32.6%. Hill *et al.*, (2006) observed 13 alleles for this locus in the US populations; the additional allele 19 was only observed in African American individuals. Afrikaner, Mixed Ancestry and Xhosa showed 11, and Asian Indian showed 9 alleles for this locus.

Allele frequency distributions across the populations were similar (Fig. 3.19). Allele 8 was only observed in the Afrikaner population with a frequency of 0.5%, this allele may be present in other populations at very low frequencies. Alleles 9.1, 10 and 17 displayed the lowest frequencies ranging between 0.5 – 2%. Coincidence of the most common allele occurred at allele 11, except

for the Asian Indian population where allele 13 was the most common. All populations showed to be in HWE (Table 3.8).

Table 3.8: Allele frequency distributions of D9S2157 within four South African populations and HWE test. The highest frequency within each population is highlighted in bold

Allele	Afrikaner (N=103)	Mixed Ancestry (N=105)	Asian Indian (N=96)	Xhosa (N=54)
7	0.138	0.097	0.087	0.088
8	0.005	---	---	---
9	0.122	0.122	0.065	0.196
9.1	---	0.020	---	0.020
10	0.005	0.010	---	0.098
11	**0.327**	**0.189**	0.174	**0.382**
12	0.087	0.071	0.027	0.029
13	0.158	0.184	**0.326**	0.029
14	0.077	0.128	0.109	0.029
15	0.061	0.097	0.136	0.069
16	0.015	0.066	0.071	0.039
17	0.005	0.015	0.005	0.020
HWE *p*-values	0.5733	0.5244	0.9429	0.9873

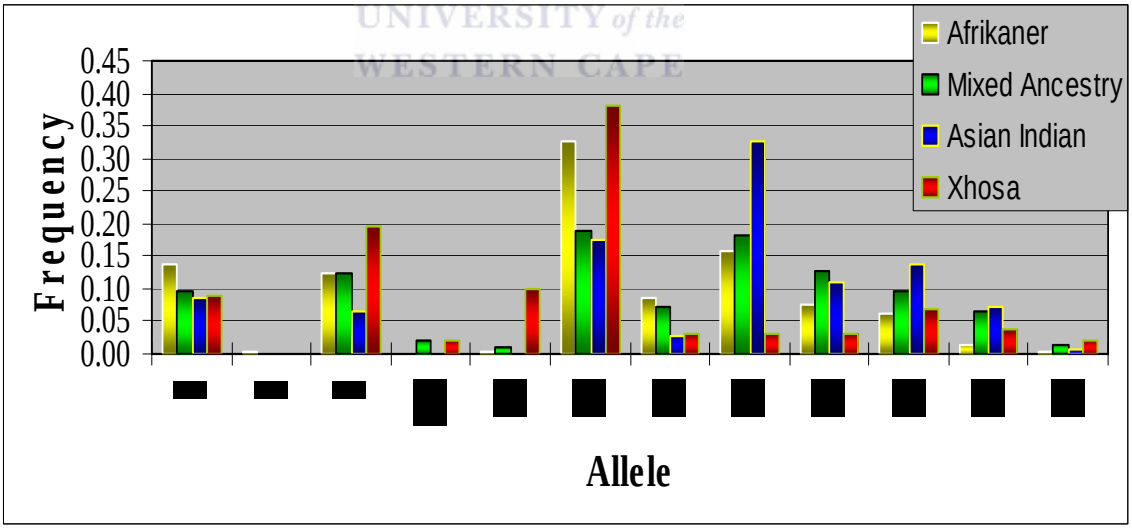


Fig. 3.19: Histogram of allele frequency distribution of the D9S2157 locus within South African populations

3.3 Microsatellite Allele Variability

Genetic variation is necessary to allow organisms to adapt to constantly changing environments. However, it is actually the variation in alleles that is critical. Alleles are different versions of the same gene. They are distinguished from other alleles of the same gene by differences in nucleotide sequences. New alleles appear in a population by random mutations and natural mutations and changes in allele frequencies occur regularly as a result of genetic drift, mutation and selection. Since every population needs variation, the measure of the amount of heterozygosity across all genes can be used as a general indicator of the amount of genetic variability within a population. The number of alleles scored is a valuable indicator of the future usefulness of the marker for forensic use.

3.3.1 Heterozygosity

According to Nei (1975), heterozygosity is a good measure of genetic diversity of polymorphic loci. The higher the heterozygosity values the broader the genetic diversity. The heterozygosity values (H) indicates the proportion of heterozygote individuals in the population. Heterozygosity is calculated by the equation:

$$H = \frac{\text{Number of samples containing heterozygote alleles}}{\text{Total number of samples}}$$

Table 3.9 Summaries the total number of alleles and total observed alleles for each MiniSTR observed across the Afrikaner, Mixed Ancestry, Asian Indian and Xhosa populations. The total number of alleles over all populations ranged from 100 to 204, with the Mixed Ancestry

population contributing the most alleles and the Xhosa population contributing the least. The mean number of alleles per population and the heterozygosities detected are good indicators of the genetic polymorphisms within the populations. Generally, the mean number of alleles is highly dependant on the sample size because of the presence of unique alleles in populations that occur in low frequencies, also because the number of alleles tends to increase with increase in the population size. This explains the low values contributed by the Xhosa population. Allelic richness however, is the number of alleles corrected for different sample sizes when comparing several populations. Allelic richness is therefore not dependant on the sample size and can be used to further investigate the genetic variation of the four populations and the informativeness of the loci. The D4S2364 showed the lowest amount of variation with 3.612 whereas the D9S2157 locus showed the highest variation with 10.307. Among the populations, the Afrikaner showed the least amount of variation and the Mixed Ancestry the highest, with values of 6.686 and 7.476 respectively (Table 3.10).

The overall heterozygosity of the miniSTR loci ranged from 0.436 – 0.807. Maximum heterozygosity was observed for D10S1248 and minimum heterozygosity observed for locus D4S2364 (Table 3.11). A high heterozygosity value (>0.70) generally indicates more allele diversity and therefore less chance of random sample matching. It therefore makes sense that the Mixed Ancestry population showed the most heterozygosity, with an average heterozygosity of 0.742, as this population is thought to be the most diverse.

Table 3.9: Variability parameters pooled for the four populations

Locus Name	Amount of alleles observed for each locus over all populations	Total number of alleles over all populations	
		Minimum Alleles	Maximum Alleles
D12ATA63	10	100	204
D22S1045	9	102	204
D10S1248	8	102	204
D14S1434	10	102	200
D9S2157	12	102	196
D4S2364	4	102	204
D1S1677	9	100	192
D2S441	11	100	196

Table 3.10: Allelic Richness per locus and population

Locus	Afrikaner	Mixed Ancestry	Asian Indian	Xhosa	Overall per locus
D1S1677	6.720	7.537	7.515	7.000	7.107
D2S441	6.876	8.440	7.093	7.000	8.242
D4S2364	3.000	3.490	3.980	3.000	3.612
D9S2157	9.415	10.579	8.525	11.000	10.307
D12ATA63	7.521	8.360	8.359	7.000	8.097
D22S1045	7.018	7.359	7.569	9.000	8.052
D14S1434	6.935	6.654	7.142	5.000	6.679
D10S1248	6.003	7.392	7.273	7.961	7.326
Overall per population	6.686	7.476	7.182	7.120	

Table 3.11: Observed Heterozygosity

Locus Name	Afrikaner	Mixed Ancestry	Asian Indian	Xhosa	Average Heterozygosity per locus
D12ATA63	0.716	0.853	0.711	0.740	0.755
D22S1045	0.747	0.765	0.817	0.765	0.774
D10S1248	0.800	0.814	0.733	0.882	0.807
D14S1434	0.680	0.608	0.731	0.627	0.662
D9S2157	0.786	0.878	0.761	0.745	0.793
D4S2364	0.490	0.520	0.538	0.196	0.436
D1S1677	0.760	0.698	0.711	0.780	0.737
D2S441	0.804	0.796	0.602	0.800	0.751
Average Heterozygosity per population	0.723	0.742	0.701	0.692	

Table 3.12: Number of Alleles Sampled

Locus	Afrikaner	Mixed Ancestry	Asian Indian	Xhosa
D1S1677	7	8	9	7
D2S441	7	9	8	7
D4S2364	3	4	4	3
D9S2157	11	12	9	11
D12ATA63	8	9	9	7
D22S1045	8	8	9	9
D14S1434	8	7	9	5
D10S1248	7	8	8	8

Table 3.13: Total Number of Alleles Observed Across all Populations

Locus Name	Afrikaner	Mixed Ancestry	Asian Indian	Xhosa
D12ATA63	190	204	180	100
D22S1045	198	204	186	102
D10S1248	200	204	180	102
D14S1434	200	194	186	102
D9S2157	196	196	184	102
D4S2364	192	204	182	102
D1S1677	192	192	180	100
D2S441	194	196	176	100

3.3.2 Polymorphic Information Content (PIC)

The Polymorphic Information Content (PIC) of a marker is the probability that a given offspring of a parent carrying a rare allele at a locus will allow deduction of the potential genotype at the locus. PIC is determined by summing the mating frequencies multiplied by the probability that an offspring will be informative (Botstein *et al.*, 1980).

The PIC values for the MiniSTRs are presented in table 3.14. The PIC values ranged from 0.43 to 0.80 in Afrikaner population; in Mixed Ancestry it ranged from 0.50 to 0.86; in Asian Indian from 0.52 to 0.79 and in Xhosa from 0.23 to 0.79. The average PIC for the markers ranged from 0.42 to 0.81 with D4S2364 being the least informative and D9S2157 showing the highest amount of polymorphism. The average PIC for each population across all loci was 0.69, 0.73, 0.69 and 0.68 in the Afrikaner, Mixed Ancestry, Asian Indian and Xhosa populations respectively. Thus, indicating that the Mixed Ancestry population shows the highest genetic polymorphism across all populations. While the Xhosa population shows the lowest.

Table 3.14: PIC values for each marker and population

Locus Name	Afrikaner	Mixed Ancestry	Asian Indian	Xhosa	Average PIC for each loci
D12ATA63	0.78	0.80	0.78	0.71	0.77
D22S1045	0.69	0.77	0.69	0.77	0.73
D10S1248	0.69	0.76	0.73	0.79	0.74
D14S1434	0.63	0.70	0.66	0.65	0.66
D9S2157	0.80	0.86	0.79	0.77	0.81
D4S2364	0.43	0.50	0.52	0.23	0.42
D1S1677	0.73	0.67	0.69	0.71	0.70
D2S441	0.74	0.74	0.68	0.79	0.74
Average PIC for each population	0.69	0.73	0.69	0.68	

3.3.3 Fixation Indexes

F-Statistics were introduced by Wright (1951) to describe the average progress of subpopulations towards fixations of alleles, and hence he called it Fixation Indexes. F-Statistics describe the amount of inbreeding – like effects within subpopulations (F_{ST} or θ), among populations (F_{IS}), and within the entire population (F_{IT}) (Butler, 2000). If $\theta = 1$, it implies that all subpopulations have reached fixation at this locus, possibly for different alleles in different subpopulations. $\theta = 0$ indicates that allele proportions are the same in all subpopulations, and therefore the population is homogenous. Factors such as admixture (e.g. when two previously distinct populations merge) and inbreeding can modify the effects of genetic drift. Population subdivision and inbreeding both increase the probability of shared ancestry of the two alleles at a locus and are one of the main potential causes of deviations from HWE. However, their effect on forensic match probabilities can be accounted for using Fixation Indexes. Deviation from HWE due to assortive mating or selection is expected to be small for forensic loci. It is expected to be limited to only one or two loci of the DNA profile (Hartl and Clark, 1989).

F-Statistics

The F_{IS} values for each population are given in Table 3.15. The mean values of inbreeding coefficient F_{IS} for Afrikaner, Mixed Ancestry, Asian Indian and Xhosa were 0.005, 0.025, 0.044 and 0.033 respectively. The highest within-population fixation index (F_{IS}) was observed for the loci D12ATA63 in Afrikaner, D14S1434 for Mixed Ancestry, D2S441 in Asian Indian and D4S2364 in the Xhosa population. The lowest values were found for loci D10S1248 in Afrikaner and Xhosa, D2S441 in Mixed Ancestry, and D22S1045 in Asian Indian population.

The negative values of F_{IS} for some of the loci, mainly in the Afrikaner and Mixed Ancestry populations, signify that the mates were less related in comparison with in the average population. Comparative to the other three populations, Asian Indian showed more inbreeding coefficient, indicating less gene exchange between it and other population.

The fixation indices (F_{IT}, F_{ST} and F_{IS}) values for each locus are shown in Table 3.16. The mean F_{IT}, F_{ST} and F_{IS} values over all the population across all loci are found to be 0.047, 0.022 and 0.025, respectively. The low values which are very close to zero indicate low level of inbreeding within (F_{IS}) and among (F_{ST}) the populations and also point towards low genetic differentiation between the populations.

Table 3.15: F_{IS} Values (inbreeding coefficient) for 8 MiniSTR loci in four South African populations

Locus	Afrikaner	Mixed Ancestry	Asian Indian	Xhosa
D1S1677	0.004	0.000	0.030	-0.022
D2S441	-0.037	-0.030	0.152	0.026
D4S2364	-0.031	0.074	0.064	0.213
D9S2157	0.043	-0.004	0.073	0.065
D12ATA63	0.118	-0.031	0.117	0.024
D22S1045	-0.019	0.048	-0.116	0.052
D14S1434	0.020	0.189	-0.024	0.112
D10S1248	-0.077	-0.022	0.048	-0.069
All	0.005	0.025	**0.044**	0.033

Table 3.16: F-Statistics analysis for 8 MiniSTR loci across Afrikaner, Mixed Ancestry, Asian Indian and Xhosa populations

Locus	F_{IT}	F_{ST}	F_{IS}
D1S1677	0.017	0.012	0.006
D2S441	0.038	0.015	0.023
D4S2364	0.080	0.029	0.053
D9S2157	0.069	0.030	0.040
D12ATA63	0.091	0.034	0.058
D22S1045	0.013	0.025	-0.012
D14S1434	0.084	0.013	0.072
D10S1248	-0.010	0.017	-0.027
All	0.047	0.022	0.025

3.3.4 Forensic Parameters

Table 3.17 shows the statistical forensic parameters for the non-CODIS loci. These parameters provide an indication of the usefulness of the loci for forensic casework. The definitions are as follows:

Power of Discrimination (PD)

Fisher (1951) first described the Power of Discrimination (PD). PD is equal to 1, minus the square of the genotype frequencies, i.e., $PD = 1 - P_1$

Power of exclusion (PE)

This parameter was first described by Fisher (1951). PE is determined by the formula:

$PE = H^2 (1 - (1 - H) H^2)$ where H is heterozygosity.

Probability of Identity (PI)

The Probability of Identity indicates the likelihood that two individuals selected at random will have an identical genotype at the tested locus. P_I is calculated by adding the squares of the genotype frequencies.

Paternity Index (PI)

The Paternity Index indicates the chance that the genetic alleles obtained by the child supports the assumption that the tested man is the true biological father rather than an untested randomly selected unrelated man.

Table 3.17: Statistical forensic parameters of eight non-CODIS miniSTRs

	D12ATA63				D22S1045			
	Afrikaner	Mixed Ancestry	Asian Indian	Xhosa	Afrikaner	Mixed Ancestry	Asian Indian	Xhosa
Power of Discrimination	0.930	0.936	0.915	0.886	0.874	0.920	0.867	0.912
Power of Exclusion	0.453	0.701	0.446	0.493	0.505	0.535	0.631	0.535
Typical paternity index	1.76	3.40	1.73	1.92	1.98	2.13	2.74	2.13
	D14S1434				**D10S1248**			
	Afrikaner	Mixed Ancestry	Asian Indian	Xhosa	Afrikaner	Mixed Ancestry	Asian Indian	Xhosa
Power of Discrimination	0.831	0.894	0.871	0.860	0.871	0.918	0.900	0.920
Power of Exclusion	0.398	0.301	0.478	0.325	0.599	0.625	0.482	0.760
Typical paternity index	1.56	1.28	1.86	1.34	2.50	2.68	1.88	4.25
	D1S1677				**D2S441**			
	Afrikaner	Mixed Ancestry	Asian Indian	Xhosa	Afrikaner	Mixed Ancestry	Asian Indian	Xhosa
Power of Discrimination	0.905	0.875	0.880	0.891	0.911	0.896	0.882	0.914
Power of Exclusion	0.528	0.425	0.446	0.562	0.607	0.591	0.294	0.599
Typical paternity index	2.09	1.66	1.73	2.27	2.55	2.45	1.26	2.50
	D4S2364				**D9S2157**			
	Afrikaner	Mixed Ancestry	Asian Indian	Xhosa	Afrikaner	Mixed Ancestry	Asian Indian	Xhosa
Power of Discrimination	0.682	0.745	0.760	0.395	0.940	0.955	0.940	0.927
Power of Exclusion	0.179	0.205	0.223	0.029	0.573	0.750	0.529	0.501
Typical paternity index	0.98	1.04	1.08	0.62	2.33	4.08	2.09	1.96

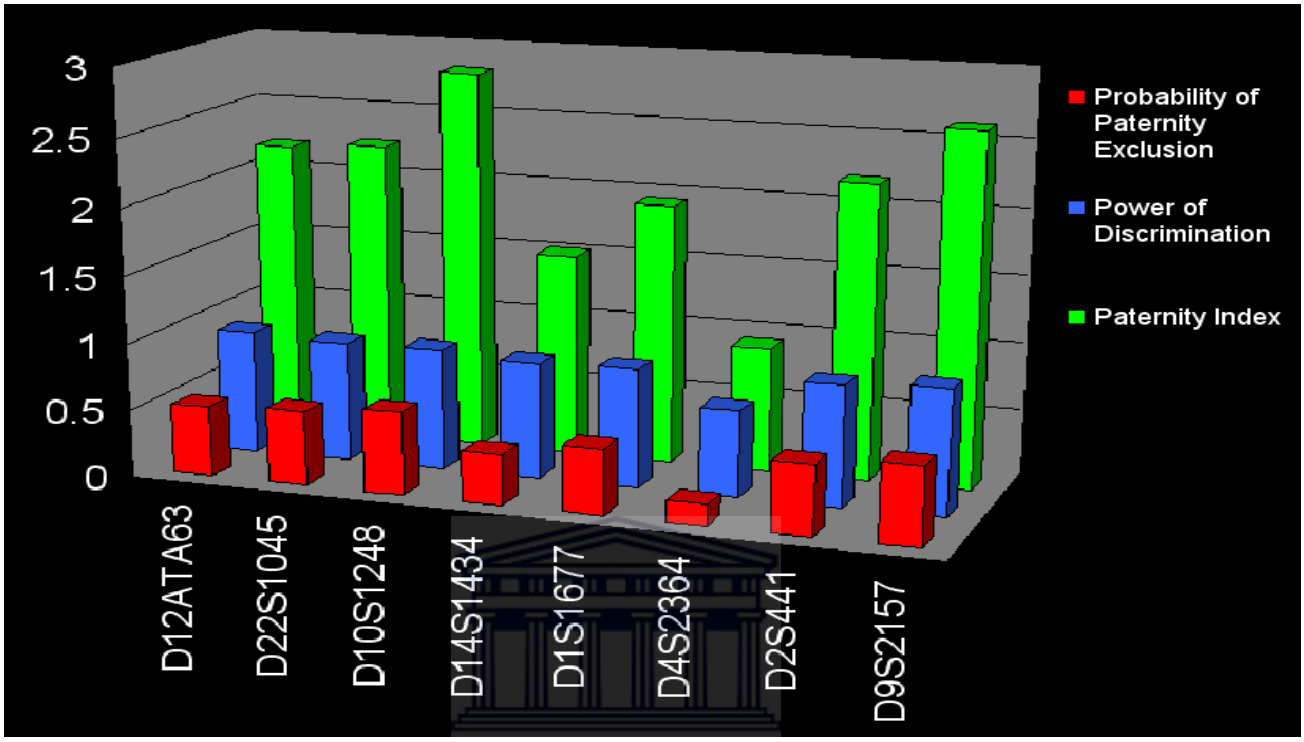


Fig. 3.20: Graphical representation of statistical parameters of the eight non-CODIS miniSTRs. The D4S2364 loci showed the lowest variation across all parameters.

3.3.5 Genetic Distance

The relationship among the populations can be analyzed using the genetic distances. The genetic distance between two samples is described as the proportion of genetic elements (alleles, genes, gametes, genotypes) that the two samples do not share. The eight loci were used for the measurement of genetic distances between the four populations. A good way to measure genetic distances is to estimate the differences in the frequency of different alleles at the loci. Populations, which have common alleles at different frequencies or different alleles all together, are farther apart. The simplest way to estimate a gene order phylogeny is to carry out a pairwise analysis, using a distance matrix. A distance matrix is constructed, where each entry in the matrix is the distance between a pair of gene orders. The matrix is then used as input to a tree-building algorithm. Saitou and Nei (1987) is the best known of these algorithms. The relationships between populations can then be summarized using a phylogenetic tree (Fig. 3.22).

In this study the genetic relationships between the populations was examined. A phylogram following Neighbor Joining (NJ) clustering was generated by means of Nei's (1972) standard genetic distance formula (Fig. 3.21) using Populations V1.2.30 (http://ftp.bioinformatics.org/pub/populatios/).

$$D = \frac{-\ln\left[\sum X_i Y_j\right]}{\sqrt{\sum X_i Y_j}}$$

Fig. 3.21: Genetic distances were estimated using Nei's (1972) standard genetic distance formula where D = Nei's standard genetic distance, X_i and Y_j are frequencies of the i^{th} and j^{th} allele respectively drawn in populations X and Y.

The NJ tree based on genetic distances showed that the four populations were separated into two groups. Afrikaner and Asian Indian were separated into one group, and the Xhosa and Mixed

Ancestry were separated into another. The Xhosa population showed smaller distance than the other populations and the mixed ancestry showed the most. In summary, the eight miniSTR loci clearly segregated in the four populations, by which genetic affiliations can be depicted. The data reported in the present study suggests that the combination of eight miniSTRs provide moderate resolution for the reconstruction of recent human evolutionary history.

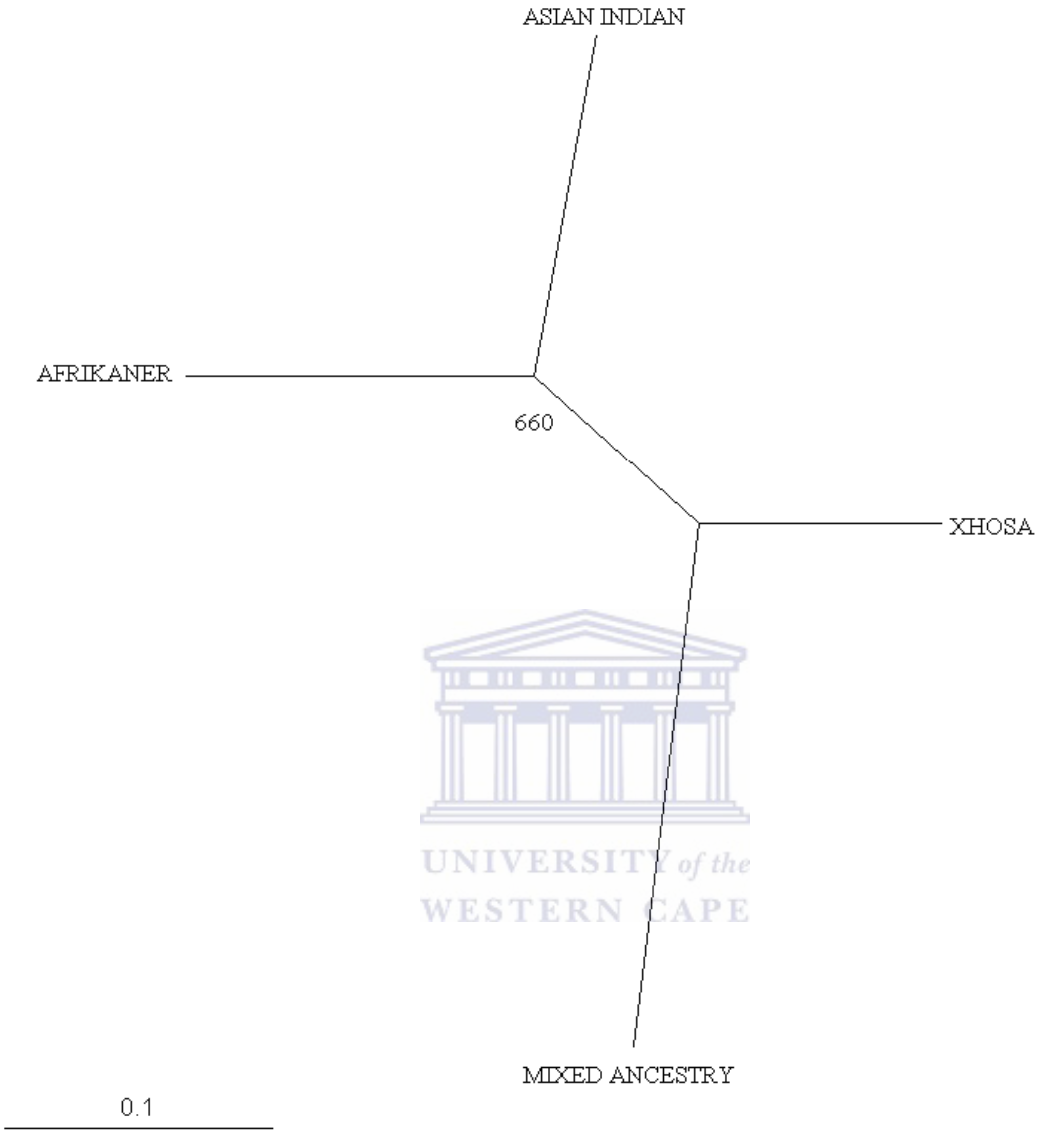


Fig. 3.22: Unrooted phylogenetic tree of the Four South African populations elaborated by the Neighbour-Joining method

CONCLUSION

All of the eight miniSTR loci studied could be amplified successfully in the four populations i.e. Afrikaner, Mixed Ancestry, Asian Indian and Xhosa. The following conclusions can be drawn from the results presented and discussed in the text:

Genetic Diversity in South African populations:

- In Afrikaner, these miniSTRs exhibited 3 to 11 alleles.
- In Mixed Ancestry, these miniSTR loci exhibited 4 to 12 alleles.
- In Asian Indian, these miniSTR loci exhibited 4 to 9 alleles.
- In Xhosa, these miniSTRs exhibited 3 to 11 alleles.
- The D4S2364 loci proved to be the least polymorphic loci across all populations.
- The D9S2157 loci proved the highest variation across all populations.

The population genetic variability parameters were:

- Afrikaner: heterozygosity values of 0.490 to 0.804 and PIC values of 0.63 to 0.80.
- Mixed Ancestry: heterozygosity values of 0.520 to 0.878 and PIC values of 0.50 to 0.86.
- Asian Indian: heterozygosity values of 0.538 to 0.817 and PIC values of 0.52 to 0.79.
- Xhosa: heterozygosity values of 0.196 to 0.882 and PIC values of 0.23 to 0.7.
- The genetic variability parameters across all loci in Afrikaner, Mixed Ancestry, Asian Indian and Xhosa were estimated to be in the range of 3 to 12 alleles, the

total number of alleles over all loci ranged from 100 to 204, the allelic richness ranged from 3.612 to 10.307 and the heterozygosity ranged from 0.4360 to 0.8073.

Inter-Population relationships

- The low mean F_{IS} (-0.027) and F_{IT} (-0.010) and F_{ST} (0.017) values across the populations indicated low level of inbreeding within and among the populations. Comparative to the other three populations, Asian Indian showed higher inbreeding coefficient, indicating less gene exchange between it and other populations.

The Non-CODIS miniSTRs will play a crucial role in future forensic investigation. In addition to the usefulness of these miniSTRs in the analyses of degraded DNA, they will also prove valuable in cases where more genetic information is needed (e.g. paternity cases and mass disaster reconstruction cases. The small sizes of these miniSTRs may have great potential in application to high-speed mass spectrometric approaches. In general, the NC-miniSTRs should offer a new potential tool for recovering useful information from samples that generated negative results or partial profiles with present multiplexes. With the exception of the D4S2364 loci, these microsatellites were highly polymorphic and proved very useful for genetic investigations and assessing admixture in South African populations.

The ability to characterise the genetic structure of populations continues to be central to all aspects of population and quantitative genetics, and measures based on correlations of pairs of alleles continue to be as important now as they were when

discussed by Wright (1951). The need for the characterisation of the genetic structure of populations, especially of humans, has increased with recent large-scale disease association studies. These eight miniSTRs serve as good candidates for such characterisation as well as for forensic casework.

Appendix I: Protocols

DNA Extraction from oral swabs using Chelex Extraction Method

1. Suspend the swab sample in 1ml of deionized water in a 1.5ml microcentrifuge tube.

2. Incubate at room temperature for 30 minutes to rehydrate the sample.

3. Use an autoclaved toothpick to tease the fibers apart on the inside of the tube. Twirl the swab and toothpick for 2 minutes to release the cells from the swab.

4. Remove the substrate and toothpick. Spin in a microcentrifuge for 2 minutes at 10, 000-15,000 x g at room temperature.

5. Without disturbing the pellet, remove and discard all but 25µl of the pellet.

6. Resuspend the pellet in the remaining 25 µl by stirring with an autoclaved pipette tip.

7. Add 5% Chelex to final volume of 200 µl.

8. Add 2 µl of 10mg/ml Proteinase K. Mix gently.

9. Incubate the sample at 56ºC for 15-30 minutes.

10. Vortex the sample at high speed for 5-10 minutes.

11. Spin the sample in a microcentrifuge for 10-20 seconds at 10, 000-15, 000 x g at room temperature.

12. Incubate the sample in a boiling water bath for 8 minutes.

13. Vortex the sample at high speed for 5-10 seconds.

14. Spin the sample in a microcentrifuge for 2-3 minutes at 10, 000-15, 000 x g at room temperature. The sample is now ready for DNA quantitation and the PCR amplification process.

15. Store the remainder of the sample at either 2-6 ºC or -15 ºC to -25 ºC.

16. To reuse, thaw the sample at room temperature and repeat steps 7-8.

DNA Purification using QIAamp® Blood Mini Kit

1. Add an equal volume of Buffer AL to the sample. Mix by pulse-vortexing for 15 s.
e.g., a 400 µl sample will require 400 µl Buffer AL

2. After mixing, briefly centrifuge the 1.5 ml microcentrifuge tube to remove drops from lid

3. Carefully apply the mixture to the QIAamp Spin Column without wetting the rim, close the cap.

4. Centrifuge at 6000 x *g* (8000 rpm) for 1 min. Place the QIAamp Spin Column in a clean 2 ml collection tube (provided), and discard the tube containing the filtrate.

5. Add 500 µl Buffer AW1 (first wash buffer) to the column.

6. Centrifuge at 6000 x *g* (8000 rpm) for 1 min. Place the QIAamp Spin Column in a clean 2 ml collection tube (provided), and discard the collection tube containing the filtrate.

7. Add 500 µl Buffer AW2 (second wash buffer) to the column. Close the cap and centrifuge at full speed (20,000 x *g*; 14,000 rpm) for 3 min.

8. Place the QIAamp Spin Column in a clean 1.5 ml microcentrifuge tube, and discard the collection tube containing the filtrate. Carefully open the QIAamp Spin Column and add 60µl Buffer AE or distilled water.

9. Incubate at room temperature for 5 min, and then centrifuge at 6000 x *g* (8000 rpm) for 1 min.
Samples are now ready for PCR amplification

Protocol for sequencing of homozygotes

Set up sequencing reaction as follow:

ExoSap IT	2ul
PCR Product	5ul

37°C – 15 minutes
80°C – 15 minutes

Set up Sequencing PCR as follow:

ExoSap Template	3.5ul
Ready Mix	4.0ul
H_2O	1.5ul
Primer (5uM)	1.0ul

Cycling done on 24 well block as follow:

96°C – 1 minute
96°C – 10 seconds
50°C – 5 seconds
60°C – 4 minutes
4°C – 60 minutes

Clean-up of sequencing products were done as follow:

[1] 125mM EDTA mixed in equal parts with 3M Sodium Acetate.
[2] 2ul added to each 10ul sequencing reaction.
[3] 25ul of 100% EtOH added to each sequencing reaction.
[4] Incubated at room temp for 15min
[5] Spun at max for 30min (room temp)
[6] Supernatant removed
[7] 70ul 70% EtOH added.
[8] Spun at max for 15 min (room temp)
[9] Supernatant removed and air dried for +-40min in fume hood.
[10] Resuspended in 1.5ul loading mix (5ul formamide to 1ul Blue Dye).
[11] Denatured for 4 min.
[12] Snap cooled on ice.
[13] Load around 0.8 to 1.2ul.
[14] Run on standard 36cm gel at using filter set E at 1200 scans per hour.

Genotyping with the ABI377

Pouring the gel

1. Place a clean set of assembled plates in the gel pouring area. Ensure that the glass plates have been moved as far down the cassette as possible. All the cassette clamps should be in place except the one that holds the beam trap. Ensure that the 48 tooth well forming comb is clean and accessible.

2. Ensure all glassware and syringes have been rinsed with deionised water.

3. Transfer 25ml of gel mix to a clean 50ml beaker. This allows for easy access to the gel mix with a 50ml syringe.

4. Suck up gel mix with the 50ml syringe and filter through a 0.22 micron filter directly into the side-arm flask.

5. Degas gel mix with intermittent gentle agitation. Degas for +/- 5 min or until no bubbles appear after agitation. The pump must be running first. Always ensure that the rubber stopper is in place. Set or break the vacuum only at the rubber hose/side arm junction.

6. Transfer from side arm flask to a clean 50ml beaker.

7. Add 125µl (AMPS) and 17.5µL N, N, N', N' Tetramethyl-EthyleneDiamine (TEMED) to opposite sides of the gel mix in the beaker. You should mix soon after adding the AMPS and TEMED but make sure you don't introduce bubbles into the mix. Once mixed the gel takes a long time to set so there is no rush.

8. Pour gel.

9. Insert comb and clamp.

10. Seal bottom of the plates with cling-wrap.

11. Clamp the beam trap down.

12. Let the gel polymerise for 2hrs.

Running the ABI377

1. Start by turning the ABI377 on and let the instrument go through its start up routine.
2. Check that the laser fan is blowing.

3. Once the ABI377 has gone through its start up routine turn on the PC.

4. Start cleaning the gel so that it can be loaded onto the ABI. Remove the clamps but leave the comb in place. Remove the cling-wrap from the bottom of the gel.

5. Remove the gel from the cassette for cleaning. Pour deionised water over the bottom and top of the gel to remove any residual gel. After the initial rinse remove the comb drawing it backwards in the plane of the gel. Doing this helps draw any polyacrylamide away from the wells. Use paper towel to dab the residual polyacrylamide from the gel.

6. Once the gel has been thoroughly rinsed, dry it with paper towel. Place it back in the cassette.

7. Lift the beam trap and clean the laser scan area.

8. Place the cleaned gel into the instrument. Clip it into place and ensure that it is correctly aligned.

9. Ensure the collection software is open. Open Genescan Run dialogue box. Set the Prerun options tab to PlateCheck X.

10. Click the PlateCheck Run option. With any of the run options if the instrument does not comply simple cancel the run and restart it.

11. Examine the scan and gel windows and cancel the run. If you're happy with the result continue with the rest of the setup. If not, clean the scan area again and repeat the plate check.

12. If the plate check is satisfactory, place the upper buffer chamber onto the instrument. Pour 1X TBE into the chamber and check for leaks.

13. If no leaks occur then add the heat plate to the gel assembly.

14. Pour 1X TBE buffer into the lower buffer chamber.

15. Flush the well area and insert the 48 tooth comb so that approximately 1.5mm of the arches is exposed.

16. Put the lid of the upper chamber on and close the instrument door.

17. Select GS PRF 2400 X from the Prerun options tab. Click on the prerun tab.

18. Open the status window and check that the 35 second initialisation has been initiated. If not cancel the run and click on the prerun tab again.

19. The moment you hear the water pump activate, pause the run, open the instrument door and check that there are no leaks.

20. If there is no leaking simply close the instrument door and allow it to heat up to 50°C.

21. While the instrument is heating up the samples can be prepared for loading.

Preparation of samples for loading on the ABI377

1. Prepare a sample sheet with the names of each sample.

2. Place the amplification-tubes in one continuous row.

3. In front of each amplification-tube place a clean 0.2ml PCR tube (loading tubes). Number the tubes with the relevant well number.

4. Prepare sufficient loading mixture (loading mixture in the following ratio: 5µL Formamide, 1.5µl ROX internal size ladder, 1.5µl Blue dye).

5. Place 1µl loading dye in each loading tube.

6. Transfer 1µl of each amplified sample to the appropriate loading tubes.

7. Place the loading tubes into the GeneAmp 2700 Thermocycler for 5 min to denature. Place odd numbered tubes consecutively into an appropriate number of rows. Then place the even numbered tubes consecutively into an appropriate number of rows. This simplifies the loading process when odd wells are loaded first, run in then even wells are loaded.

8. During this time the paused prerun will brought the gel to 50°C. Cancel the prerun and set up the Run parameters (48 tooth comb, GS RUN 2400 X module, Operator, Filter set, Sample Sheet). Click the run tab. Pause the instrument as soon as you hear the water pump.

9. With the instrument paused open the instrument door and flush the wells with a 5µl syringe. Load 0.8µl of each sample into the relevant well. Load all odd numbered samples first, run the instrument for 2 minutes, pause it and load the even samples. Restart the run. It should take 2hrs 15 minutes for all the relevant bands to be detected.

10. As soon you've seen the 400bp band you can terminate the run and process the results.

Processing the results from the ABI377

1. Open the Gel processor software. Auto track the lanes. Each lane should be visually inspected and any tracking errors corrected.

2. Once tracked the lanes can be extracted.

3. Open Genescan, start a new project and import the relevant run files. Before analyzing each sample install a matrix file and an appropriate size standard. Good initial peak height threshold settings for the colour channels are B: 100, G: 100, Y: 100, R: 50.

4. Click the analyse tab to analyse the samples.

5. Once the initial processing is done check each sample individually to see that all the peaks has been detected. Where necessary raise the peak height cut off threshold values to avoid stutter or lower the threshold to make sure that all allele peaks are detected.

6. The project file can now be saved and is ready to be typed.

Genotyping of Samples

The samples can either be typed manually using a ladder or by using the Genotyper Software as follows:

1. Open Genotyper Macro

2. Click File, Import from Gensecan files. Search for appropriate project file and add it to the Genotyper document and save.

3. Click inside the Dye/Lanes window, click Edit and Select all.

4. Double click on the Label Peak Sizes macro.

5. Select each colour and ensure that the peaks have been properly labeled with sizes as follows: Click on one of the coloured squares next to the Dye/Lanes window and draw a virtual box around the peaks. Press Ctrl R to zoom in on peaks and Ctrl Y to show individual plots. If peaks have been missed or labeled incorrectly, change this by clicking on the appropriate peaks. Press Ctrl H to zoom back out. Do this for all other colours as well and save.

6. Click inside the Dye/Lanes window, click Edit and Select all.

7. Double click on the Label Peaks Categories macro.

8. Select each colour and ensure that the peaks have been properly labeled with the allele numbers as follows: Click on one of the coloured squares next to the Dye/Lanes window and draw a virtual box around the peaks. Press Ctrl R to zoom in on peaks and Ctrl Y to show individual plots and check. Press Ctrl H to zoom back out. Do this for all other colours as well and save.

9. Double click on the Make Table macro and save. All samples are now typed.

Appendix II: Composition of Buffers and Solutions

0.5 M EDTA pH 8-8.5 Commercial Stock

10% (w/v) *N*-lauroyl sarcosine (50ml)

5g *N*-lauroyl sarcosine

Add *N*-lauroyl sarcosine to 50ml DI H_2O.

20 $mgml^{-1}$ Proteinase K (5ml)

100g Proteinase K

Add Proteinase K to 5ml DI H_2O.

5% (w/v) Chelex (50ml)

2.5g Chelex

Add Chelex to 50ml DI H_2O.

10X TBE Buffer (1*l*)

108g Tris Base
EDTA
~55g Boric Acid

Add Tris Base, EDTA and Boric Acid to 800ml DI H_2O and mix to dissolve. Adjust to 1*l* with additional DI H_2O once dissolved.

1X TBE Buffer (1*l*)

100ml 10X TBE Buffer

Add TBE to 900ml DI H_2O.

Polyacrylamide Gel Mix (50ml)

18g Urea
5ml Long-Ranger Gel Solution
5ml 10X TBE

Add required DI H_2O to Urea, Long-Ranger Gel Solution and 10X TBE get 50ml.

REFERENCES

Asamura H, Tsukada K, Ota M, Sakai H, Takayanagi K, Kobayashi K, Saito S, Fukushima H. 2006. **Population study of small-sized short tandem repeats in Japan and its application to analysis of degraded samples.** *International Congress Series*; 1288:610-612

Balding D J. 2005. **Weight-of-evidence for Forensic DNA Profiles. Illustrated Edition.** *John Wiley and Sons*

Bender K, Schneider P M, Rittner C. 2000. **Application of mtDNA sequence analysis in forensic casework for identification of human remains.** *Forensic Science International*; 113:103-107

Bender K, Farfán M J, Schneider P M. 2004. **Preparation of degraded human DNA under controlled conditions.** *Forensic Science International*; 139:135-140

Bennet P. 2000. **Microsatellites Demystified…** *J Clin Pathol: Mol Pathol*; 53:177-183

Botstein D, White RL, Skolnick M, Davis R W. 1980. **Construction of a genetic linkage map in man using restriction fragment length polymorphisms**. Am J Hum Genet. 32(3):314-31.

Budowle B, Baechtel F S. 1990. **Modifications to improve the effectiveness of restriction fragment length polymorphism typing.** *App. Theor Electrophoresis*; 1:181-187

Budowle B, Moretti T M, Niezgoda S J, Brown B L. June 1998. **CODIS and PCR-Based Short Tandem Repeat loci: Law enforcement tools.** *Proceedings of the Second European Symposium on Human Identification.* Innsbruck, Austria, Madison, WI: Promega Corporation, 73-88

Burger J, Hummel S, Herrmann B, Henke W. 1999. **DNA preservation: A microsatellite-DNA study on ancient skeletal remains.** *Electrophoresis*; 20:1722-1728

Butler J M. 2000. **Forensic DNA Typing.** *Academic Press: San Diego, CA.*

Butler J M, Shen Y, McCord B R. 2003. **The development of reduced size STR amplicons as tools for analysis of degraded DNA.** *J Forensic Sci*; 48(5): 1054-1064

Chung D T, Drábek J, Opel K L, Butler J M, McCord B R. 2004. **A study on the effects of degradation and template concentration on the amplification efficiency of the STR Miniplex primer sets.** *J Forensic Sci*: 49(4): 733-740

Chung U, Shin K, Park M, Kim N, Yang W, Cho S, Lee H. 2007. **Population data of nine miniSTR loci in Koreans.** *Forensic Science International.* 168(2); e51-e53

Coble M D, Butler J M. 2005. **Characterisation of new miniSTR loci to aid analysis of degraded DNA.** *J Forensic Sci*; 50(1): 43-51

Collins M J, Nielsen-Marsh C M, Hiller J, Smith I, Roberts J P. 2002. **The survival of organic matter in bone: A review.** *Archaeometry*: 44(3): 383-394

Dean L. 2005. Blood Groups and red cell antigens: The ABO blood group. Chapter 5. NCBI; available online at http://ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=rbcantigen&part=ch05ABO

Dent B B, Forbes S L, Stuart B H. 2004. **Review of human decomposition processes in soil.** *Environmental Geology*; 45: 576-585

Dixon L A, Dobbins A E, Pulker H K, Butler J M, Vallone P M, Coble M D, Parson W, Berger B, Grubwieser P, Mogensen H S, Morling N, Nielsen K, Sanchez J J, Petkovski E, Carracedo A, Sanchez-Diz P, Ramos-Luis E, Briōn M, Irwin J A, Just R S, Loreille O, Parsons T J, Syndercombe-Court D, Schmitter H, Stradmann-Bellinghausen B, Bender K, Gill P. **Analysis of artificially degraded DNA using STRs and SNPs--results of a collaborative European (EDNAP) exercise.** *Forensic Sci Int.* 2006 Dec 1;164(1):33-44. Epub 2005 Dec 15

Dizdaroglu M, Jaruga P, Birincioglu M, Rodriguez H. 2002. **Free-radical induced damage to DNA: Mechanisms and measurement.** *Free Radical Biology & Medicine*; 32(11): 1102-1115

Egyed B, Furedi S, Angyal M, Boutrand L, Vandenberghe A, Woller J, Padar Z. 2000. **Analysis of eight STR loci in two Hungarian populations.** *Forensic Science International*; 113: (1-3): 25-7

Eglinton G, Logan G A, Ambler R P, Boon J J, Perizonius W R K. 1991. **Molecular Preservation and Discussion.** *Phil. Trans. R. Soc. Lond.*; 333: 315-328

Fisher RA. 1951. **Standard calculations for evaluating a blood group system.** *Heredity*; 5:95-102

Friedberg E C, Walker G C, Siede W. 1995. **DNA repair and mutagenesis.** *ASM Press, Washington, DC.*; 698

Forster P, Calì F, Röhl A, Metspalu E, D'Anna R, Mirisola M, De Leo G, Flugy A, Salerno A, Ayala G, Kouvatsi A, Villems R, Romano V. 2002. **Continental and sub-continental distributions of mtDNA control region types.** *Int L Legal Med*; 116: 99-108

Frank R, Köster H. 1979. **DNA chain length markers and the influence of base composition on electrophoretic mobility of oligodeoxyribonucleotides in polyacrylamide gels.** *Nucl Acids Res*; 6: 2069-2087

Gill P. 2001. **Biological Evidence (Including Hair).** *13th Interpol Forensic Science Symposium, Lyon France*; 16-19

Gill P, Fereday L, Morling N, Schneider P M. 2006. **The Evolution of DNA databases – Recommendations for new European STR loci.** *Forensic Science International*; 156: 242-244

Golenberg E M, Bickel A, Weihs P. 1996. **Effect of highly fragmented DNA on PCR.** *Nucleic Acids Research*; 24(24): 5026-033

Graw M, Weisser H-J, Lutz S. 2000. **DNA typing of human remains found in damp environments.** *Forensic Science International*; 113: 91-95

Grubweiser P, Muhlmann R, Berger B, Niederstätter H, Pavlic M, Parson W. 2006. **A new "miniSTR-multiplex" displaying reduced amplicons lengths for analysis of degraded DNA.** *Int J Leg Med*; 120:115-120

Hartl D L, Clark A G. 1997. **Principles of Population Genetics, 3 ed.** *Sinauer Associates, Sunderland, Massachusetts*. 542 pp.

Hellman A, Rohleder U, Schmitter H, Wittig M. 2001. **STR typing of human telogen hairs – a new approach.** *Int J Legal Med*; 114:269-273

Higuchi R, Bowman B, Freiberger M, Ryder O A, Wilson A C. 1984. **DNA sequences from the quagga, an extinct member of the horse family.** *Nature*; 312:282-284

Hill C R, Coble M D, Butler J M. 2006. **Characterization of 26 New miniSTR Loci**. *Poster #44 at Promega Meeting, Nashville, TN*

Copy of poster available:

http://www.cstl.nist.gov/biotech/strbase/pub_pres/Promega2006_Hill.pdf

Hoff-Olsen P, Mevag B, Staalstrom E, Hovde B, Egeland T, Olaisen B. 1999. **Extraction of DNA from decomposed human tissue. An evaluation of five extraction methods for short tandem repeat typing.** *Forensic Science International*; 105:171-183

Hoff-Olsen P, Jacobsen S, Mevåg B, Olaisen B. 2001. **Microsatellite study in human post- mortem tissues**. *Forensic Science International*; 119:273-278

Hofreiter M, Serre D, Poinar H N, Kuch M, Pääbo S. 2001. **Ancient DNA.** *Nature Reviews Genetics*; 2: 353-359

Holland M M, Fisher D L, Mitchell L G, Rodriquez W C, Canik J J, Merril C R, Weedn V W. 1993. **Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War.** *J Forensic Sci*; 38:542-553

Holland M M, Parsons T J. 1999. **Mitochondrial DNA sequence analysis – validation and use for forensic casework.** *Forensic Sci Rev*; 11: 21-50

Höss M, Jaruga P, Zastawny T H, Dizdaroglu M, Pääbo S. 1996. **DNA damage and DNA sequence retrieval from ancient tissues.** *Nucleic Acids Research*; 24(7): 1304-1307

Hummels S, Schultes T, Bramanti B, Herrman B. 1999. **Ancient DNA profiling by megaplex amplification.** *Electrophoresis*; 20: 1717-1721

Jeffreys A J, Wilson V, Thein S L. 1985. **Hypervariable 'minisatellite' regions in human DNA**. *Nature*; 314(6006):67-73

Martìn P, Garcia O, Albarran C, Garcia P, Yurrebasoi I., Alonso A. 2007. **Allele frequencies of six miniSTR loci (D10S1248, D14S1434, D22S1045, D4S2364, D2S441 and D1S1677) in a Spanish population.** *Forensic Science International.* 169(2); 252-254

McCord B. 2005. **The development of Miniplex primer sets for the analysis of degraded DNA.** *U. S. Department of Justice Report: Document no.210037*

Mitchell D, Willerslev E, Hansen A. 2005. **Damage and repair of ancient DNA.** *Mutation Research*; 571:265-276

Nanodop® Technologies. 2004. **NanoDrop® ND-1000 Spectrophotometer User's Manual V3.0.1.** Available online at:

http://www.cvm.tamu.edu/dnacore/Documents/nd-1000-users-manual.pdf

National Institute of Justice (NIJ) Report. 2000. **The future of forensic DNA typing: Predictions of the research and development-working group.** *US Department of Justice*

Nei M. 1972. **Genetic distance between populations.** *American Naturalist.* 106: 283-292

Nei M. 1975. **Molecular genetic population and evolution.** pp. 128-170. Amsterdam, North-Holland Publishing.

Nilsson M. 2007. **Mitochondrial DNA in sensitive forensic analysis (Digital dissertation).** *Uppsala University*

Opel K L, Chung D T, Drábek J, Tatarek N E, Jantz L M, McCord B R. 2006. **The application of Miniplex primer sets in the analysis of degraded DNA from human skeletal remains.** *J Forensic Sci*; 51 (2): 351-356

Pääbo S. **Molecular cloning of ancient Egyptian mummy DNA.** 1985. *Nature*; 314: 644-645

Pääbo S, Poinar H, Serre D, Jaenicke-Desprës V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M. 2004. **Genetic analysis from ancient DNA.** *Annu Rev Genet*; 38: 645-679

QIAGEN Worldwide. 2001. **QIAamp® 96 DNA Blood Kit.** Available online at: http://www.ebiotrade.com/buyf/productsf/qiagen/1019159_HBQA96DNA Blood_1001.pdf

Raymond M., Rousset F. 1995. **Genepop (version 1.2), population genetics software for exact tests and ecumenicism.** *Journal of Heredity 86: 248-249.*

Ricci U, Giovannucci M L, Klintschar M. 1999. **Modified primers for D12S391 and a modified silver staining technique.** *Int J Legal Med*; 112: 342-344

Rogan P K, Salvo J J. 1990. **Study of nucleic acids isolated from ancient remains.** *Yearbook of Physical Anthropology*; 33: 195-214

Rozen S. Skaletsky H J. 1998. **"Primer3: A software component for picking PCR primers." 1996-present.** *Source code available at* https://sourceforge.net/project/showfiles.php?group_id=112461 ; public WWW server at http://fokker.wi.mit.edu/primer3/input.htm

Saitou N, Nei M. 1987. **The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees.** *Mol. Biol. Evol.* 4(4); 406-425

Satoh M, Kuroiwa T. 1991. **Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell.** *Exp Cell Res*; 196(1):137-40.

Schumm J W. 1997. **Why use a size marker and allelic ladders in STR analysis?** *Profiles in DNA, Geneprint™*; 11-13

Takahashi M, Kato Y, Mukoyama H, Kanaya H, Kamiyama S. 1997. **Evaluation of five polymorphic microsatellite markers for typing DNA from decomposed human tissues – Correlation between the size of the alleles and that of the template DNA.** *Forensic Science International*; 90:1-9

Tautz D. 1993. **Notes on the definition and nomenclature of tandemly repetitive DNA sequences.** *EXS*; 67: 21-8

Thomas M, Gilbert P, Hansen A J, Willerslev E, Rudbeck L, Barnes I, Lynnerup N, Cooper A. 2003. **Characterisation of genetic miscoding lesions caused by postmortem damage.** *Am J Hum Genet*; 72:48-61

Utsuno H, Minaguchi K. 2004. **Influence of template DNA degradation on the genotyping of SNPs and STR polymorphisms from forensic materials by PCR.** *Bull Tokyo Dent Coll*; 45(1): 33-46

Vass A. 2001. **Beyond the grave – understanding human decomposition.** *Microbiology Today*; 28: 190-192

Wallin J M, Holt C L, Lazaruk K D, Nguyen T H, Walsh P S. 2002. **Constructing universal multiplex PCR systems for comparative genotyping.** *J Forensic Sci*: 47(1); 52-65

Watson S, Allsop R, Foreman L, Kelsey Z, Gill P. 2001 **Sequenced allelic ladders and population genetics of a new STR multiplex system.** *Forensic Science International;* 115: 207-217

Wiegand P, Kleiber M. 2001. **Less is more – Length reduction of STR amplicons using redesigned primers.** *Int J Legal Med*; 114: 285-287

Wright S. 1951. **Genetical structure of populations.** *Annals of Eugenics;* 15:323 – 354

Yong R, Gan L, Coble M, Yap E. 2006. **Allele frequencies of six miniSTR loci of three ethnic populations in Singapore.** *Forensic Science International*; 166(2): 240-243

UNIVERSITY of the
WESTERN CAPE