Development and Validation of a Non-CODIS miniSTR genotyping system suitable for forensic case work in South Africa

A thesis submitted in partial fulfilment of the requirements for the degree of Magister Scientiae in the Department of Biotechnology, University of the Western Cape

Zainonesa Abrahams

March 2010

Supervisor: Dr. Mongi Benjeddou (PhD)

Keywords

Forensics

Mini short tandem repeats (miniSTRs)

Non-CODIS (NC)

Loci

Polymerase chain reaction (PCR)

DNA sequencing

Degraded DNA

Internal validation

Population Studies

Allele frequencies

Forensic parameters

List of Publication(s)

Allele frequencies of six non-CODIS miniSTR loci (D1S1627, D3S4529, D5S2500, D6S1017, D8S1115 and D9S2157) in three South African populations – Forensic Science International:Genetics (*In press*)

Abstract

Development and Validation of a Non-CODIS miniSTR genotyping system suitable for forensic case work in South Africa

Z. Abrahams

MSc Thesis, Department of Biotechnology, University of the Western Cape

24/02/2010

The objective of this study was to develop and validate a six Non-CODIS miniSTR genotyping system and to determine its suitability for forensic casework in South Africa. In Non-CODIS miniSTR genotyping systems, smaller PCR products are amplified and the primers are positioned as close as possible to the repeat region. For this reason, these systems can be valuable in a variety of scenarios including complex paternity cases, missing persons work, and mass fatality disasters.

After the successful implementation of the genotyping system in the laboratory, allele size range was determined for each of the loci and allelic ladders were constructed. The entire repeat regions of the six loci under investigation were successfully sequenced. Consequently, allele repeat number, structure and observed size were determined for each locus.

An internal validation study of the six Non-CODIS miniSTR genotyping system was conducted following the SWGDAM guidelines. A comprehensive population study, covering five population groups from South Africa was also carried out.

The genotyping system produced consistent, accurate and precise genetic profiles for low concentrations of template DNA. When analyzing mixed DNA samples, successful

differentiation of minor and major DNA components was identifiable. Amplification products were observed in non-human DNA studies but in all instances complete genotype profiles were not obtained.

Allele frequencies and forensic parameters were determined for the system in five South African population groups (i.e. Afrikaner, Asian-Indian, Mixed Ancestry, Xhosa and Cape Muslim). No deviation from Hardy-Weinberg equilibrium was observed in any of the populations. Furthermore, all populations displayed a high power of discrimination and a high power of exclusion.

The six Non-CODIS miniSTR genotyping system has shown a good potential to aid in the analysis of degraded DNA samples. This system can be further improved by including additional loci. Even in its current form, it can certainly provide additional discrimination in complex paternity and/or missing person cases.

Declaration

I declare that 'Development and Validation of a Non-CODIS miniSTR genotyping system suitable for forensic case work in South Africa' 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 have been indicated and acknowledged by complete references.

Full Name: Zainonesa Abrahams

Signature:

Date: 24/02/2010

Acknowledgements

First and foremost I would like to thank my Creator for granting me the opportunity and then helping me fulfill a dream of doing research in DNA forensics. To my parents, I thank you for supporting and allowing me to pursue all my dreams no matter how big or small. Lateefah, Sudique and Silvana, the adults in my life, thank you for your patience and understanding. Nabeelah, Mariam and Shaakir, the bright stars in my dark night sky- thank you for the laughter you bring to my heart but most of all thank you for loving me even when I get it wrong sometimes.

Megan, my thesis buddy, thank you for the friendship, support and motivation which you so willingly gave. The ladies who kept me sane, from lab work to thesis writing and all the way down to heart ache, Thielshaat, Sindisiwe and Emmerentia, thank you for your friendship and guidance without it I would not be me.

My supervisor, Dr. Benjeddou, the girls-Tasneem, Aneesah and Shafieka, and other members of the lab, thank you for always being readily available to help me understand and accept the complexities of being a Masters student. The rest of the technical staff members, thank you for making me laugh and listening to me especially those times when I sounded brainless. To the National Research Foundation, I thank you for funding this research project.

To the rest of my family and friends, thank you for keeping me grounded, motivated and inspired.

Last and most definitely not the least, Shakeel, my soul mate, thank you for everything - your love, your compassion, your friendship. You complete me.

List of Abbreviations

DNA	Deoxyribonucleic acid
RFLP	Restriction fragment length polymorphism
SNP	Single nucleotide polymorphism
STR	Short tandem repeat
mt	Mitochondrial
PCR	Polymerase chain reaction
BSA	Bovine serum albumin
RFU	Relative fluorescence units
CODIS	Combined DNA Index System
NC	Non-CODIS
EDNAP	European DNA profiling group
ENFSI	European network of Forensic Science Institute
SWGDAM	Scientific working group on DNA analysis methods
ISFG	International Society for Forensic Genetics
HWE	Hardy-Weinberg equilibrium
Ho	Observed heterozygosity
He	Expected heterozygosity
MP	Match probability
PD	Power of discrimination
PE	Power of exclusion

CPD	Combined power of discrimination
CPE	Combined power of exclusion
PIC	Polymorphism information content
TPI	Typical paternity index
CTPI	Combined typical index

List of Tables

List of Figures

Figure 5.7

Tri-allelic pattern for the locus D8S1115

Figure 5.8

Tri-allelic pattern for the locus D9S2157

Table of Contents

Chapter 1 – Literature Review

Appendix I

Chapter 1: Literature Review

1.1 Overview of genetic identity testing

Genetic identity testing involves identifying patterns of genetic material that are unique to almost every individual (Saad, 2005). It takes advantage of the 0.3% DNA sequence variations in the human genome to distinguish between individuals (Butler, 2005). Different techniques are used to test for these variations. These techniques are mostly known as DNA fingerprinting, DNA profiling, and DNA typing. Although there are some technical differences between these tests, the terms have been used interchangeably (Saad, 2005). These include, among others, restriction fragment length polymorphisms (RFLP) analysis, single nucleotide polymorphisms (SNPs), short tandem repeat sequences (STRs), and mitochondrial (mt) DNA sequencing (Rudin & Inman, 2002; Butler, 2005). These techniques are indeed valuable in a number of situations including the determination of perpetrators of violent crime such as murder and rape, resolving un-established paternity, identifying remains of missing persons or victims of mass disasters (Butler, 2005).

Most human identity testing is performed using DNA markers on the autosomal chromosomes, and gender determination is done with markers on the sex chromosomes (Butler, 2005). The chromosomal position or location of the gene or a DNA marker is commonly referred to as a locus (Butler, 2005). A genetic marker can therefore be described as a variation which may arise due to mutation or alteration in the genomic loci that can be observed (Vicente & Fulton, 2004). A genetic marker may be a short DNA sequence, such as a sequence surrounding a single base-pair change, e.g. SNP, or a long

one like minisatellites (Vicente & Fulton, 2004). These polymorphic (variable) markers that differ among individuals can be found throughout the human genome (Butler, 2005).

1.2 Major forms of DNA polymorphism

1.2.1 Forms of DNA polymorphism

DNA polymorphism was defined by Strachan and Read (2004) as: "the existence of two or more variants (alleles, phenotypes, sequence variants, chromosomal structure variants) at significant frequency in a particular population". It can exist in two forms; either a sequence polymorphism or a length polymorphism. Sequence polymorphisms are indicated by a change in one or more bases in the DNA sequence at a particular locus. A length polymorphism occurs where there is a variation in the length of a repetitive DNA sequence (Rudin & Inman, 2002; Butler, 2005). Different techniques are used to test for DNA polymorphisms: i) restriction fragment length polymorphisms (RFLP) analysis, ii) single nucleotide polymorphisms (SNPs), iii) short tandem repeat sequences (STRs) and iv) mitochondrial (mt) DNA sequencing.

1.2.2 RFLP analysis

RFLP analysis was one of the first techniques adapted for forensic DNA analysis (Rudin & Inman, 2002; Budowle *et al*, 2000). RFLPs are genetic markers wherein a fragment of DNA sequence containing a variable restriction site is amplified (Strachan & Read, 2004). Therefore, in the amplified product (i.e. the amplicon), the restriction site for the restriction enzyme will either be present or absent (Strachan & Read, 2004). RFLP analysis entails the following: i) extraction of DNA, ii) digestion of the DNA into

fragments with restriction enzymes and iii) electrophotoretic separation of fragments based on size (Budowle *et al*, 2000). The restriction enzyme will cut the DNA surrounding the repeat polymorphism (Butler, 2005). Each individual will show different distances for the sites of cleavage for the restriction enzyme (Butler, 2005). So doing, it will produce DNA fragments of variable lengths when comparing from one person to another (Rudin & Inman, 2002; Budowle *et al*, 2000).

RFLP loci used for forensic purposes were selected on the basis that they indicated hundreds of variations at each locus (Rudin & Inman, 2002). Furthermore, different restriction enzymes are available to digest restriction sites. Because of this, RFLP is normally the first technique of choice to be used in DNA analysis since it generates the highest degree of discrimination per locus (Rudin & Inman, 2002). Therefore, not only does RFLP analysis have a high power of discrimination, but it also enables the usage of minimal loci to discriminate between individuals (Rudin & Inman, 2002).

RFLP is not the best technique to use for forensic purposes since it requires large amounts of high quality DNA and the technique is fairly time consuming and labour intensive. Forensic evidence is often old or degraded or of limited quantity. Furthermore this technique is difficult to automate and its usage in the building and participation in large DNA databases is also very limited (Rudin & Inman, 2002).

1.2.3 SNP analysis

Single nucleotide polymorphisms *(*SNPs*)* are the most abundant form of genetic variation in the human genome (Sobrino *et al*, 2005). Generally, SNPs are used as markers to identify genes which may underlie complex diseases (this is also known as disease gene-

mapping) (Shastry, 2007; Caulfield, 2001). SNPs are suitable for forensic purposes since they display slow mutation rates and this makes them useful for parentage testing (Sobrino *et al*, 2005; Sanchez *et al*, 2003). SNP markers also have the advantage that only small DNA fragments are needed for typing (Sanchez *et al*, 2003; Dixon *et al*, 2005). Generally forensic casework samples may have undergone some form of degradation and thus using SNP markers can be advantageous. Furthermore they are suitable for high throughput technologies and this is vital for the creation of automated systems and the building of databases (Sobrino *et al*, 2005; Thacker *et al*, 2004).

SNPs, however, have some limitations. They possess a rather limited polymorphic content and because of this four times more SNPs are required in comparison to the number of STRs to generate conclusive results (Sobrino *et al*, 2005; Dixon *et al*, 2006). Even though the cost of running high throughput SNP systems may work out to be more cost effective to its STR counterparts, the amount of knowledge gained over the past few years with regards to STR systems far surpasses SNPs and thus STR usage might seem to be more feasible (Sobrino *et al*, 2005; Dixon *et al*, 2006). SNPs therefore can be used in combination with STR systems to assess difficult forensic casework samples (Jobling, 2001; Dixon *et al*, 2006). SNPs are also useful in reconstructing Y-chromosome and mtDNA haplogroups for various genetic ancestry studies (Sobrino *et al*, 2005; Lessig *et al*, 2004; Jobling 2001).

1.2.4 STR analysis

The eukaryotic genome is filled with repeated DNA sequences which occur in different sizes (Butler, 2005). The repeated DNA sequences are specified by the length of the core

repeating unit and the number of continual repeat units or by the overall length of the repeating unit (Butler, 2005). Some of these repeats may contain hundreds to thousands of base pairs (Butler, 2005). The repeating DNA sequence is normally ~2-6bp in length and are normally repeated ~5-20 times (Chung *et al*, 2004; Rudin & Inman, 2002; Schoske *et al*, 2003) and are called microsatellites also known as short tandem repeats (STRs). Dinucleotide repeats (e.g. AC-AC-AC) have two bases in the repeat unit, trinucleotide repeats (e.g. ACT-ACT-ACT) have three bases in the repeat unit, tetranucleotide units have four bases in the repeat unit (e.g. ACTG-ACTG), etc.

STRs do not only show differences in the length and the number of repeating units, they also differ in their repeat patterns (Butler, 2005). Some STR markers contain simple repeats wherein the repeat units have identical length and sequence, whilst others contain compound repeats where the repeating unit contains two or more adjacent simple repeats. Other STR markers can contain complex repeats where the repeat unit can be made up of a number of repeat lengths and may contain variable intervening sequences (Butler, 2005). Furthermore, other alleles may contain incomplete repeat units and are known as microvariants (Butler, 2005). An example of a microvariant is the allele 9.3 at the TH01 locus. It contains nine complete AATG repeat units and one incomplete repeat wherein an adenine is missing (Puers *et al*, 1993).

For STR analysis, primers are designed from sequences known to flank the STR locus. This enables the amplification of alleles whose sizes will vary by the number of repeating units they contain (Strachan & Read, 2004). STR markers are not as polymorphic as RFLP loci but thousands of STR loci have been identified in human DNA (Butler, 2005;

Rudin & Inman, 2002). STR loci can be amplified simultaneously (i.e. multiplex PCR), thereby saving: time, sample, reagents and other consumables (Rudin & Inman, 2002).

1.2.5 mtDNA sequencing

In some instances, such as DNA samples which have undergone extensive degradation and also in instances where tiny amounts of DNA are available for amplification, STR loci amplification can be challenging. In these instances, mtDNA sequencing is most likely to produce a useful genetic profile.

In any given cell there are hundreds of copies of mtDNA present in comparison to nuclear DNA. It has a small genome (~16.5kb) which is maternally inherited and generally does not undergo recombination (Forster *et al*, 2002; Carracedo *et al*, 2000).

These characteristics of mtDNA make it an attractive alternative for typing of degraded DNA and it is often used to type: dead cells in hair shafts (Takayanagi *et al*, 2003), bones and teeth (Irwin *et al*, 2007). mtDNA sequencing is more flexible than STR markers in the sense that more distant maternal relatives can be used as genetic references (Budowle *et al*, 2005).

Since mtDNA is only maternally inherited it would not enable discrimination between maternally related individuals (Coble *et al*, 2004). Furthermore mtDNA can display heteroplasmy, i.e. mtDNA extracted from different tissue types within an individual can show variation when comparing one tissue type to another (Carracedo *et al*, 2000). It is also a time consuming technique to employ. In essence mtDNA sequencing is very useful in the construction of haplogroups and analyzing population history, but cannot be

used alone for forensic purposes (Sobrino *et al*, 2005; Irwin *et al*, 2007; Budowle *et al*, 2005).

1.3 STR markers and their usage in forensic DNA typing

Forensic samples can be very challenging to work with. Sexual assault cases will most likely contain a mixture of DNA from both the victim and the suspect. Remains may have been exposed to a number of different factors such as: fire (e.g. in April 1993 where 80 people were killed by a fire which struck the Branch Davidian religious sect in Texas), explosions and collapse of buildings (e.g. the terrorist attack on the World Trade Centre in 2001), water (e.g. the tsunami which struck South East Asia in December of 2004) (Clayton *et al*, 1995; Budimlija *et al*, 2003; Lessig *et al*, 2006). Furthermore, remains are also exposed to environmental factors such as bacterial and fungal contamination, insects and growth of microorganisms (Bender, 2004). All of these play a contributory role in making it difficult in correctly identifying victims.

In most of the instances mentioned above, DNA profiling is most likely to be the method where accurate identification can be obtained. STR loci are the most commonly used loci for human identification (Chung *et al*, 2004; Butler, 2005; Butler *et al*, 2003). This is because these loci are: highly polymorphic, require minimal template DNA to enable genotyping, possess a narrow size range thereby permitting multiplexing and the results generated using STR loci are obtained fairly easily (Hummel *et al*, 1999; Butler, 2005; Chung *et al*, 2004; Opel *et al*, 2006). Dinucleotide, trinucleotide and tetranucleotide STR loci have been used as markers in linkage studies and with the building of the human genetic map (Fornage *et al*, 1992; Edwards *et al*, 1991). Tetranucleotide loci have

however been favoured for usage as forensic markers since they are highly polymorphic and display heterozygosity values greater than 0.90 (Walsh *et al*, 1996).

STR typing has shown to be useful when typing DNA extracted from skeletal remains and compromised samples (Alonso *et al*, 2001; Budimlija *et al*, 2003; Holland *et al*, 2003; Miloš *et al*, 2007). Hereby STR markers displayed their ability to enable the identification of human remains since they could be used in instances where DNA degradation was observed (Hummel *et al*, 1999).

1.4 Multiplex amplification of STR markers

Multiplex PCR allows for the amplification of two or more DNA fragments simultaneously. This technique facilitates the co-amplification of a range of STR loci whilst at the same time reducing the amount of reagents and time needed to amplify uniplex PCRs. This is particularly important in forensic cases where the amounts of sample(s) are generally limited.

Multiplex reactions require that primer pairs which will be run together be compatible since they will be run under the same PCR conditions. Therefore melting temperatures need to be fairly similar, they should not exhibit significant interactions with one another, themselves or with other regions of the template (Schoske, 2003). Primer dimmer formation must be prevented by avoiding complementarities between primers (Schoske *et al*, 2003). Success of multiplex PCR assays is dependant on: relative concentrations of the primers, concentration of PCR buffer, cycling parameters and a balance between deoxynucleotide and magnesium chloride ($MgCl_2$) concentrations (Henegariu, 1997). When many loci are being amplified simultaneously the PCR has a limited supply of

enzyme and nucleotides since all products will be competing for the same supplies (Schoske *et al*, 2003; Henegariu, 1997).

Amplification of STR alleles can produce artifacts known as stutter products (i.e. observation of additional peaks) (Walsh *et al*, 1996). Other complications with STR allele amplification include: non-template addition, microvariance, mutations and polymorphisms (Butler, 2005). All of these complications can make it difficult to generate accurate DNA profiles.

1.4.1 Stutter products

Stutter product formation is observed when a PCR product, one repeat unit smaller than the main product peak is detected (Walsh *et al*, 1996; Gill *et al*, 1997). These minor product peaks are also referred to as "shadow bands" or stutter peaks (Walsh *et al*, 1996; Butler, 2005). Stutter peaks occur as a result of slippage of the *Taq* polymerase during DNA replication (Walsh *et al*, 1996; Gill *et al*, 1997).

During DNA synthesis, the primer and template become separated and upon re-annealing to each other slippage of either the primer or the template has occurred, thereby causing one repeat to form a non-base paired loop (Walsh *et al*, 1996; Butler, 2005). So doing, a PCR product one repeat shorter than the main allele is produced.

Thus, in dinucleotide repeat loci, a stutter band will be 2bp smaller than the main allele and it will be 3bp smaller than the main allele in a trinucleotide repeat and so on (Walsh *et al*, 1996; Mulero *et al*, 2006). Stutter peaks are usually 10% of the main peak (Walsh *et al*, 1996). Longer alleles show a higher degree of stutter than shorter alleles for the same locus (i.e. a 6 repeat allele will show a lower degree of stutter in comparison to a 9

repeat allele of the same locus) (Walsh *et al*, 1996). Furthermore, the greater the repeat unit, the lower the degree of stutter (i.e. pentanucleotide repeat < tetra < tri <di) (Butler, 2005).

Stutter products are especially problematic in mixed DNA samples, where more than one contributor is possible. Because stutter bands have the same size as the main allele, it can be difficult to determine whether the faint band in a mixed sample is an allele from a minor contributor or simply a stutter product from the adjacent allele (Walsh *et al*, 1996; Gill *et al*, 1997; Clayton *et al*, 1998).

1.4.2 Non-template addition

The Taq polymerase used in the PCR reaction mixture can sometimes cause the addition of a non-template adenine residue to the 3' end of the PCR product (Butler, 2005). This process is also referred to as adenylation. This non-template addition of an adenine residue therefore causes the PCR product to be one base pair longer than the actual target DNA sequence (Butler, 2005). It is best to have uniform PCR products instead of a mixture of partially adenylated and non-adenylated and adenylated product, since each would produce products of varying sizes (Butler, 2005). 3' adenine nucleotide addition will be favoured by adding in an additional incubation step of 60°C/72°C post the final cycling steps (Butler, 2005).

1.4.3 Microvariance, mutations and polymorphisms

Sequence variation can be exhibited in the human population and they generally occur as a result of: insertions, deletions or nucleotide changes (Butler, 2005; Huel *et al*, 2007).

These alleles therefore show a slight variation in sequence when compared to the rest of the population and are thus referred to as microvariants. Microvariant alleles are therefore slightly different in comparison to the full allele repeat (Butler, 2005). They are also referred to as "off-ladder" alleles since they will not size correctly with the alleles present in the allelic ladder (Butler, 2005). Microvariant alleles can easily be detected in heterozygous samples because one of the alleles will line up with a fragment in the allelic ladder whilst the other remaining allele will not (Butler, 2005).

Sometimes alleles will show the same length but a sequence variation will exist. In these instances the micorvariant allele can only be detected by sequence analysis (Butler, 2005). Sequence variation can also cause amplification failure and result in the production of null alleles or lower the peak area of the affected alleles (Huel *et al*, 2007). This normally occurs when the sequence variation occurs in the primer binding sites (Huel *et al*, 2007).

1.5 Autosomal STR loci

1.5.1 CODIS STR Loci

Since 1997, STR loci systems have been used by the FBI for the identification of humans (Butler, 2006). The FBI developed a database of STR markers known as the **CO**mbined **D**NA **I**ndex **S**ystem (CODIS) which was comprised of 13 core STR loci. These 13 CODIS loci are: CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S359, D18S51 and D21S11 (Butler, 2005). When the 13 CODIS loci are randomly tested, they display a match probability of one in a trillion in unrelated individuals (Butler, 2005). FGA, D18S851 and D21S11 are the most

polymorphic markers amongst the 13 CODIS loci, whilst TPOX is the least polymorphic (Butler, 2005).

These 13 loci can be amplified simultaneously and were thus developed into commercial kits, i.e. Powerplex® 16 kit from Promega Corporation and the AmpF∫STR® Identifiler™ from Applied Biosystems. These commercial kits generate amplicons in the range of 100-450bp (Grubweiser *et al*, 2006; Butler *et al*, 2003). Furthermore, each kit is provided with an allelic ladder which enables accurate genotyping.

The advantage of generating kits like these is that they allow for database construction which can be used and shared world wide. Since the invent of the CODIS loci, the United Kingdom and United States alone have been able to obtain more than 5 million DNA profiles in their criminal justice DNA database (Butler, 2006). In addition to this another 1 million DNA profiles were obtained using these loci in paternity cases (Butler, 2005). The 13 CODIS loci have therefore become the common currency for exchange of data with regards to human identity (Butler, 2006).

1.5.2 Performance of the CODIS STR Loci

The commercial kits have proved to be very successful in certain forensic casework samples, however problems have been reported when these kits were used to analyze degraded DNA samples or DNA samples which had PCR inhibitors present. The commercial kits were able to generate complete DNA profiles in instances where high quality DNA was obtained (Butler, 2006; McCord, 2005; Chung *et al*, 2004, Grubweiser *et al*, 2006). In contrast, when degraded DNA or small amounts of DNA were amplified, these kits only produced partial profiles where the larger molecular weight markers failed

to draw conclusive results (Opel, *et al,* 2006; McCord, 2005; Grubweiser *et al,* 2006; Chung *et al,* 2004; Butler, 2006; Romano *et al,* 2006).

In instances where STR loci are unable to generate accurate DNA profiles, mtDNA sequencing is normally employed. However, because of the disadvantages associated with mtDNA sequencing (i.e. maternal inheritance and heteroplasmy) it is not always feasible to use as a method for discriminating between individuals especially maternally related individuals (Holland & Parsons, 1999; Pakendorf & Stoneking, 2005).

1.5.3 Improvement of performance of genotyping systems on degraded DNA

A few hours after death DNA degradation will begin (Butler, 2005). During DNA degradation the template DNA is broken and so doing becomes fragmented. The rate of DNA degradation can vary depending on light, humidity and temperature (Bender *et al,* 2004, Burger *et al,* 1999). Damp environments in particular cause the most damage and show an increased rate of degradation (Graw *et al,* 2000, Burger *et al,* 1999). When amplifying degraded DNA, successful amplification is not always obtained, especially in the higher molecular weight ranges since the loss of template DNA or a break in template DNA or the presence of PCR inhibitors will not enable annealing of primers thereby resulting in amplification failure (Takahashi *et al,* 1997; Bender, 2004; Chung, *et al,* 2004; Grubweiser *et al,* 2006). Therefore, allelic dropout, negative amplification and allelic imbalance can be observed (Walsh, 1992; Butler, 2005).

Allelic imbalance, i.e. the larger allele is fainter than the small allele, is generally the first indication of DNA degradation (Utsumo & Minaguchi, 2004**).** This therefore implies that smaller alleles are more effectively amplified in comparison to their larger counterparts.

Takahashi *et al* (1997) determined that with an increase in size of loci alleles, an increased amount of template DNA was required to enable detectable amplification. Bender *et al* (2004) made use of the AmpF∫STR Profiler^{TM} kit and applied it to the amplification of artificially degraded DNA. Their results indicated that STR loci >250bp either produced incomplete or no results. In addition to this, strong peak imbalance and even allelic dropout (i.e. the complete loss of one allele in a heterozygous sample) were observed. Chung *et al* (2004) observed a loss of peak intensity with an increase in amplicon size with the PowerPlex® 16 system.

These kits, therefore produced a decay curve where peak height was inversely proportional to the length of the amplicon (Figure 1.1) (Chung *et al*, 2004; Opel *et al*, 2006). The larger amplicons show a lowered sensitivity and can in some instances fall below the detection threshold thereby generating partial genetic profiles (Chung *et al*, 2004; Opel *et al*, 2006).

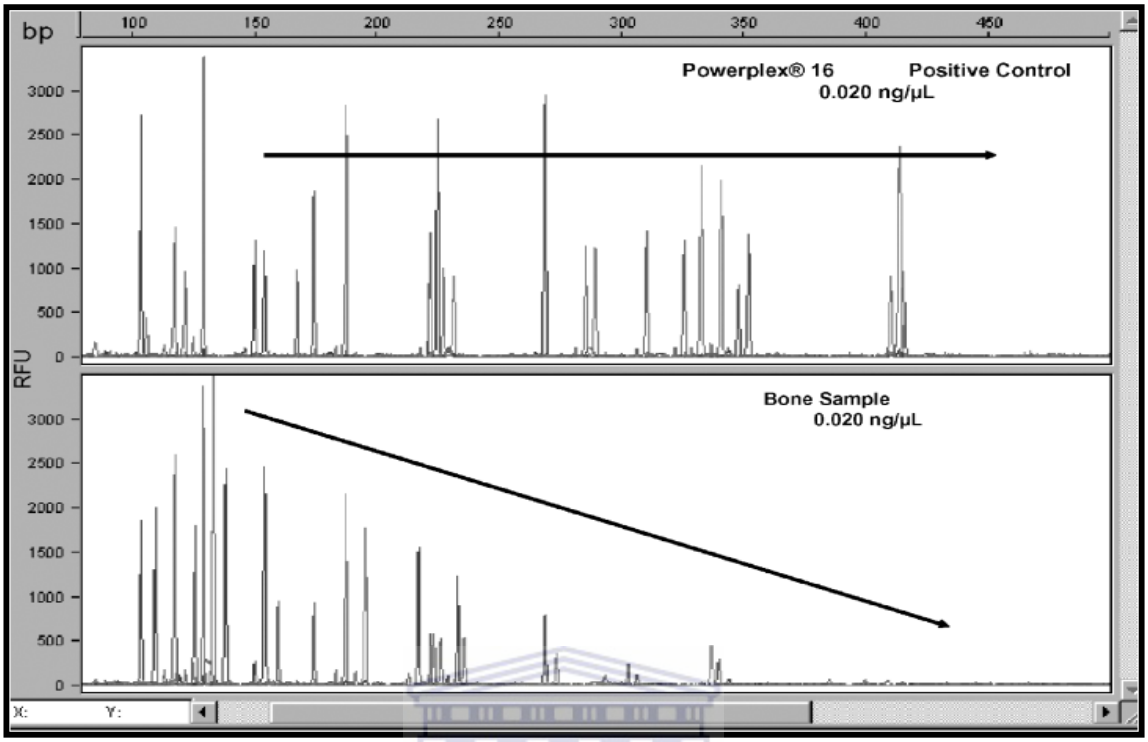


Figure 1.1: Decay curve of degraded DNA. Top panel: positive control DNA sample 9947A amplified with the Powerplex® 16 system. Bottom panel: bone sample that has been exposed to the environment for 3 years (therefore containing degraded DNA) amplified with Powerplex® 16 system. In the degraded bone sample amplicon size shows an inverse relationship in comparison to relative peak fluorescent intensity when amplified with the Powerplex® 16 system. Taken from Opel *et al* (2006).

The development of STR primers which would be able to successfully generate genetic profiles from degraded DNA therefore became a priority. Ricci *et al.* (1999) and Grubweiser *et al* (2003) demonstrated that by redesigning STR loci primers to produce shorter amplicons, they were able to successfully type samples which previously failed to produce product(s) with the normal full length primers. Smaller amplicons were achieved simply by moving the primers closer to the STR repeat region thereby generating shorter amplicon fragments.

Grubweiser *et al.* (2006) designed a STR multiplex system using the loci TH01, D2S1338, D18S51, D16S539 and FGA wherein amplicon sizes were reduced. Fragment length reduction ranged between 109 to 162bp for the selected STR loci. Thirty five casework samples which showed loss of loci when analysed with the SGMplus kit, showed successful amplification with the reduced STR loci. As expected, the SGMplus kit showed loss of signal when amplifying compromised DNA. Not only did the reduced STR loci cause an increase in discrimination power of the STR loci, but they also demonstrated reproducible results with 100pg of genomic DNA. Hellmann *et al.* (2001) redesigned the primers of the STR loci: FES, TH01 and TPOX so that amplicons produced sizes less than 110bp. These newly designed primers were tested on telogen hairs and an increase from 21.7% to 65.3% of successfully typed single telogen hairs were observed. Wiegand and Kleiber (2001) reduced the amplicon size of the loci: TH01, D10S2325, DYS19 and DYS391. With these primers, they were able to successfully type DNA form formalin-fixed tissue and 6 year old mummified tissue. Turrina *et al* (2007) made use of re-designed STR primers: CSF1P0, D8S1179 and D13S317 to obtain genetic profiles from tissue which were stored in Bouin's fluid for 10 years. Successful typing of stain samples was observed by Wiegand *et al* (2006) when making use of a short amplicon STR multiplex.

This therefore proved that by moving the primers as close as possible to the STR repeat region, typing efficiency was increased in degraded DNA samples, where the normal STR alleles would result in allelic drop out, allelic imbalance or complete loss of signal. When designing these reduced STR primers care should be taken so that null allele production is avoided (Grubweiser *et al*, 2006).

Using this strategy of reducing the amplicon size, Butler *et al*. (2003) developed new primers dubbed "miniSTR" based on the 13 CODIS STR loci. The 13 CODIS loci were selected for their study since these loci were the most widely used loci and were therefore considered to be the perfect candidates for the study. By redesigning the primer binding sites of the CODIS loci, each STR marker was made as small as possible (Figure 1.2). The newly designed primers were combined into multiplexes and referred to as miniplexes. Each locus was allocated a single dye colour. Thus one drawback of these miniplex systems is that unlike the normal STR commercial kits where 13 loci can be simultaneously amplified, now four to five amplifications are required to enable typing of all the core loci (Butler *et al*, 2003) and in some instances this proved to be counter productive as was experienced by Parsons *et al*. (2007).

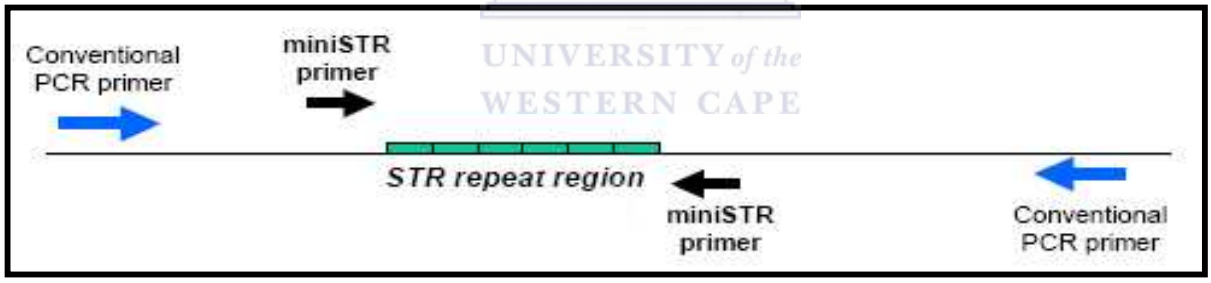


Figure 1.2: Diagrammatic representation of miniSTR primers. MiniSTR is a reduced sized amplicon of the conventional PCR primers. MiniSTR enables high recovery of information from degraded DNA samples. Taken from Coble & Butler (2005).

1.5.4 The CODIS miniSTR system and beyond

Using their newly designed miniSTRs Butler *et al* (2003) was able to obtain improved genetic profiles from partially degraded DNA when amplified with the miniSTR and compared with commercial STR kits. They were able to obtain fully concordant results

between the miniSTRs and the commercial kits. Opel *et al* (2006) conducted a study on naturally degraded DNA extracted from human skeletal remains using the miniSTR primers designed by Butler *et al* (2003). Sixty four percent of the samples tested produced complete genetic profiles using the miniplex sets whilst only 16% of samples generated full profiles with the Powerplex® 16 kit. McCord (2005) conducted a validation study using the miniplex primer sets. In this study degraded DNA was applied to the miniplex primer sets as well as the normal CODIS multiplex primers. The results generated showed that the miniplex sets, which resulted in smaller amplicon products, had a higher degree of success with generating genetic profiles when using degraded DNA. Chung *et al* (2004) used the miniplex primer sets on both enzymatically degraded DNA and naturally degraded DNA. In both types of degraded DNA the miniplex primer sets were capable of producing more complete genetic profiles in comparison to its commercial counterpart.

Currently, commercial kits have become available that contains the miniSTR loci as designed by Butler *et al* (2003). The AmpFʃSTR® Minifiler™ kit contains the reduced STR loci of the Identifiler® kit. Hill *et al* (2007) used AmpFʃSTR® Minifiler™ kit and the Identifiler® kit in a concordance study. They obtained concordant results but also obtained null alleles in some instances where a heterozygous sample would show allelic dropout and therefore appear to be a homozygote. The Mini-SGM multiplex kit, (which contains the miniSTR loci: TH01, FGA, D18S51, D16S539 and D2S1338), was used by Camacho *et al* (2008) to identify degraded samples. Their results indicated that complete profiles could not be obtained in all the samples.

Therefore, the miniSTR primer sets were able to produce full genetic profiles in the majority of samples containing degraded DNA, however there were still some instances where only partial profiles could be obtained as previously mentioned (Hummel *et al*, 1999; Butler *et al*, 2003; Kleyn, 2009; Romano *et al*, 2006). The underlying reason for this is that not all the CODIS loci could be successfully reduced to enable sufficient reduction in product size that would enable it to successfully amplify degraded DNA. For example: the locus FGA contains large allele ranges, D7S820 and D18S51 contains flanking regions which are not amenable to redesigned primers (Coble & Butler, 2005; Butler *et al*, 2003). To overcome these limitations, a comprehensive project was undertaken by Hill and co-workers (2008) to search for new markers that produce shorter amplicons. Primers were moved as close as possible and often directly next to the STR repeat region.

1.6 Overview of Non-CODIS genotyping systems

Coble and Butler (2005) characterized six novel miniSTR loci to complement the 13 CODIS loci especially in instances where degraded DNA samples were being investigated. The STR loci selected were: D1S1677, D2S441, D4S2364, D10S1248, D14S1434 and D22S1045. Selected miniSTR loci were arranged into two miniplexes each containing three loci. Uncompromised DNA samples produced reliable results with DNA concentrations as low as 100pg. DNA extracted from degraded bone samples showed a higher success rate of amplification using the new miniSTR loci in comparison to the Powerplex 16 commercial kit which showed partial profiles for the majority of samples (Coble and Butler, 2005). Furthermore an increase amount of genetic

information was found by Eisenberg *et al* (2006) when analyzing skeletal remains. A significant increase in the power of discrimination and the power of exclusion as well as a reduction in random match probabilities when combining results from the CODIS loci with these six minSTR loci were observed (Eisenberg *et al*, 2006).

The success of these studies thus showed that investigating additional miniSTR loci would prove beneficial to the forensic community. Hill *et al* (2006a, 2008) characterized an additional 20 miniSTR loci. These newly developed miniSTR loci were referred to as the Non-CODIS loci. The development of these 26 Non-CODIS loci therefore now enabled the use of additional chromosomal locations which could be vital in the analysis of degraded DNA samples, as well as missing persons investigations and parentage testing especially in instances where a limited amount of reference samples were available (Hill *et al*, 2006a; 2008; Goodwin *et al*, 2004; Ricci *et al*, 2007). Non-CODIS loci were preferentially designed on chromosomes that differed from the 13 CODIS loci. Instances where Non-CODIS loci and CODIS loci occur on the same chromosome, they were placed approximately 50Mb apart thereby eliminating linkage (Hill *et al*, 2008). Figure 1.3 illustrates chromosomal position of Non-CODIS loci in relation to CODIS loci. All candidate loci primers were designed as close as possible to the repeat regions, in some instances, right next to the repeat, to allow for the smallest possible amplicons. Amplicon sizes ranged between 50bp-150bp (Hill *et al*, 2008).

All 26 loci investigated were comparable with the 13 CODIS loci with regards to population statistics and heterozygosity values (Hill *et al*, 2008). Asamura *et al* (2007) developed two miniSTR systems which contained two of the Non-CODIS loci. The newly developed miniSTR systems combined were able to generate power of

discrimination values of 0.999 and power of exclusion values of 0.998. Asamura *et al* (2006 & 2007) determined that combining results generated from commercial kits with results generated from miniSTR systems was a good combination to get effective results from degraded DNA samples.

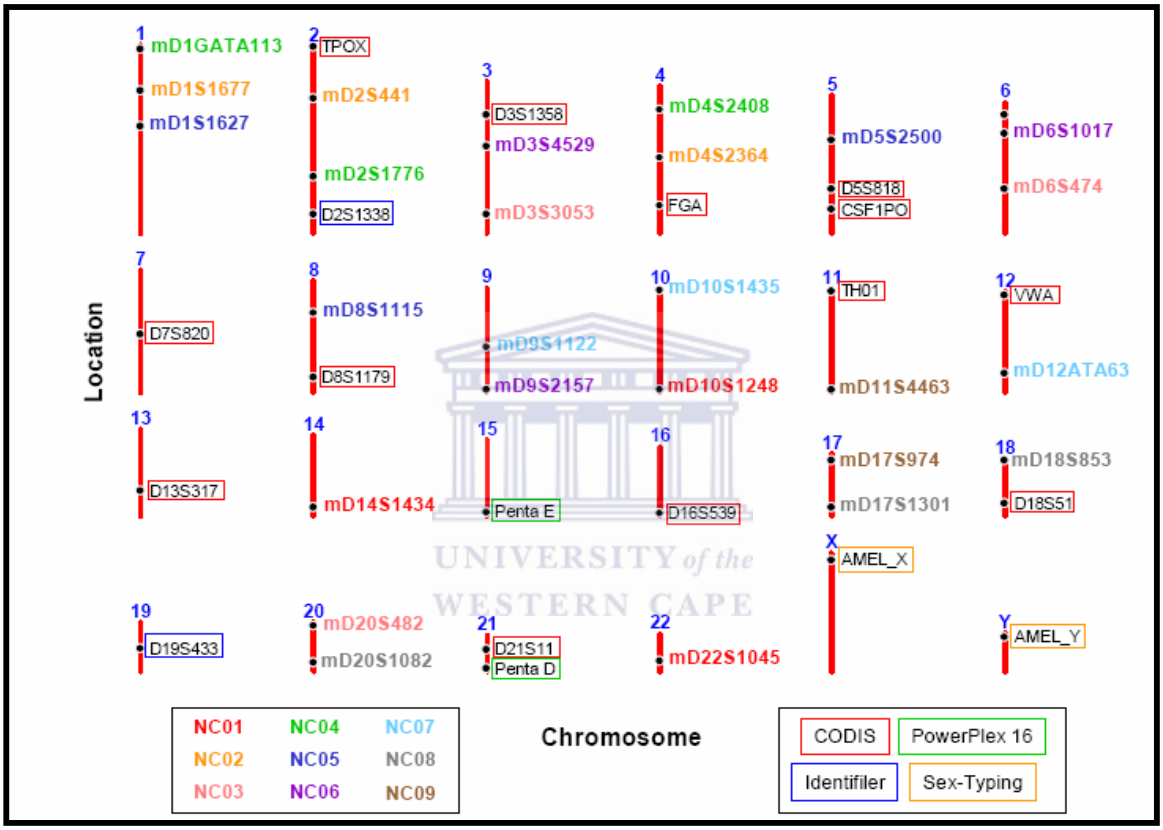


Figure 1.3: **Diagrammatic representation the Non-CODIS loci**. Chromosomal position of Non-CODIS loci in relation to the 13 CODIS loci. Taken from Hill *et al* (2006a).

The European DNA Profiling Group (EDNAP) and the European Network of Forensic Science Institute (ENFSI) has included the use of the Non-CODIS loci: D10S1248, D2S441 and D22S1045 into their core loci database (Gill *et al.*, 2006a and 2006b). Essentially Non-CODIS loci were not designed to replace the 13 CODIS loci, but instead

to complement/extend them especially in instances where the 13 CODIS loci are unable to produce conclusive results (Reid *et al*, 2008).

1.7 Development, internal validation and population studies

1.7.1 Non-CODIS miniSTR loci primer design

As previously stated, the Non-CODIS loci primers were designed as close as possible to the STR repeat region thereby enabling the production of amplicons ranging in sizes of 50bp-150bp. It is thus envisioned that this reduction in size of amplicons will be able to facilitate the successful analysis of degraded or otherwise compromised DNA samples. Furthermore these primers were arranged into triplexes allowing for concurrent amplifications. Triplexes were arranged in such a manner that additional loci could be added to each triplex (Hill *et al*, 2008) thereby increasing the number of loci analyzed per assay allowing for an increased power of discrimination.

1.7.2 Allelic ladders

An allelic ladder functions as the reference or standard used for each STR locus. It contains all the alleles which are present in the human population for that particular STR locus (Bar *et al*, 1997; Butler, 2005). They are amplified using the same primers as the test samples (Butler, 2005). Allelic ladders are thus produced by combining PCR products from different samples, where the sum of all samples will be representative of all the alleles present in the population for that particular STR locus (Figure 1.4) (Butler, 2005; Sgueglia *et al*, 2003; Bar *et al*, 1997). To enable equal representation of each allele in the allelic ladder, the amount of PCR product or template DNA used representing each allele can be varied (Butler, 2005).

Allelic ladders are used so as to enable the sizing of unknown fragments. ABI Prism DNA Genotyper Analysis Software (version 3.7) allows for ±0.5bp variation between the unknown fragment and the allelic ladder fragment (Sgueglia *et al*, 2003). Outliers or variant alleles may occur outside of this specified range (Sgueglia *et al*, 2003). Since the allelic ladder and the sample fragments have the same length and sequence, they will migrate the same distance during the electrophoresis process regardless of any environmental changes (Schumm, 1997). Thus making use of an allelic ladder will enable the identification of off-ladder alleles which generally arises because of unexpected population genetic variation (Huel *et. al*, 2007) an enable accurate genotyping of known alleles.

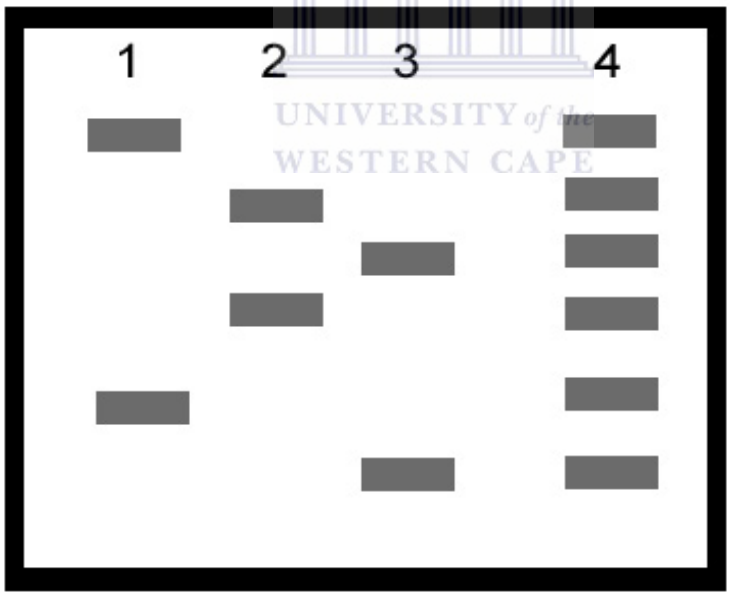


Figure 1.4: Principle of an allelic ladder. Lanes 1-3: STR alleles from individual DNA samples (representative of the common alleles for the locus) separated on a polyacrylamide gel. Lane 4: Combining of individual samples for the generation of an allelic ladder. Adapted from Butler (2005).

1.7.3 Validation studies

Multiplex typing systems must be optimized to the point where they meet certain performance standards. There are several governing bodies that ensure that high typing and analysis standards are maintained. Among these is the International Society for Forensic Genetics (ISFG), the Scientific Working Group on DNA Analysis Methods (SWGDAM), and the European DNA Profiling Group (EDNAP). These organizations have proposed guidelines for the use and validation of multiplex PCR typing systems. Some common validation exercises include: (1) establishing that the typing system is sensitive and performs consistently using freshly prepared and stored DNA, (2) that identical results are obtained irrespective of the type of tissue from which DNA was extracted, (3) that the systems yield consistent results in several laboratories, and (4) that the system performs well when used to analyze samples similar to those encountered in forensic casework (Leat *et al*, 2004a) Validation plays a vital role in the forensic community since there is a constant influx of new DNA technology being developed (Butler *et al*, 2004). There are two types of validation processes; i.e. a developmental validation and an internal validation.

1.7.3.1 Developmental validation

Developmental validation is the more thorough of the two methods and it is used to determine the reliability and limitations of a novel method (Daniels *et al*, 2004). This entails the determination of: accuracy, precision, reproducibility, species specificity, sensitivity, stability, PCR parameters and simulated casework studies (Daniels *et al*, 2004). SWGDAM has recommended a set of developmental validation guidelines which are available from www.cstl.nist.gov/strbase/validation/SWGDAM. Developmental

validations are generally performed by: manufacturers, academic institutions, technical organizations and government laboratories (Butler, 2005).

1.7.3.2 Internal validation

The purpose of an internal validation is solely to demonstrate the ability of a laboratory to perform a previously developmentally validated procedure (Daniels *et al*, 2004). Both developmental and internal validations determine whether the analytical procedure will be adequate for its intended use (Butler *et al*, 2004).

The SWGDAM guidelines to be followed during the implementation of an internal validation are presented in Appendix I. Validation studies should be complemented with local population background data especially in instances where DNA evidence is presented in court.

1.7.4 Population studies

1.7.4.1 Ethnic diversity in South Africa

When forensic data is presented in court, especially genetic information, it needs to be substantiated with background information on local population structures. Globally, South Africa is referred to as the 'Rainbow Nation' because of the mixture of nationalities that reside within this country. In 2007 the South African population was estimated to be ~47.9 million people. Majority of the South African inhabitants are referred to as 'Black Africans' and they are more accurately defined by their cultures and languages (Figure 1.5). The remaining 20.4% of the total population can be subdivided into: 9.1% white/Caucasian, 8.9% coloured/mixed ancestry and 2.5% Indian/Asian

(www.southafrica.info/about/people/population) (Figure 1.5). The ethnic diversity within South Africa is demonstrated by the eleven official languages of the country of which nine are African (Figure 1.6 shows the break down of South African populations based on linguistic preference and cultures). The objective of population studies is to determine the likelihood that a set number of markers will randomly or non-randomly identify individuals via various statistical measurements (Taroni *et al*, 2007).

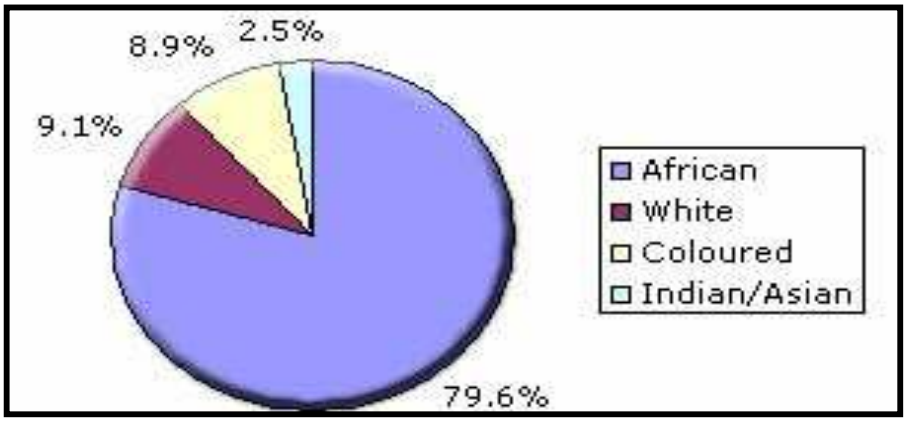


Figure 1.5: Graphical illustration of racial distribution within South Africa. Taken from: www.southafrica.info/about/people/population.

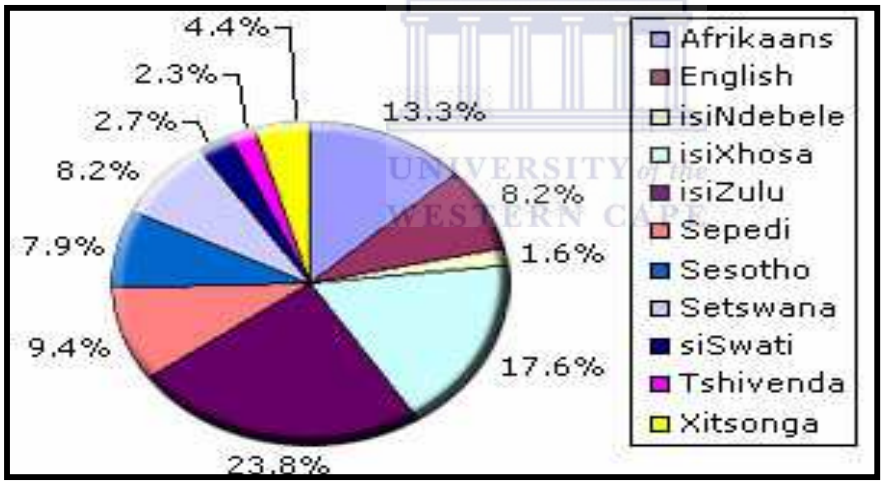


Figure 1.6: **Linguistic preference and culture of South African population groups**. Black South Africans include the: Zulu, Xhosa, Ndebele, Swazi, Sotho, Tswana, Tsonga and Venda people. Caucasian, coloured and Asian Indian people generally speak a combination of English and Afrikaans. Taken from: www.southafrica.info/about/people/population.

1.7.4.2 Statistical analysis

In addition to the number of people typed, population data sets should include the main statistical parameters such as *p*-value(s) for Hardy-Weinberg equilibrium testing, heterozygosity value(s), discrimination power and chance of exclusion. Additional quality criteria such as the use of sequenced allelic ladders could also be presented.

1.7.4.2.1 The Hardy-Weinberg equilibrium

When observed and expected genotype and allelic frequencies are similar, the population is said to be in Hardy-Weinberg Equilibrium (HWE). The Hardy-Weinberg principle assumes that allele and genotype frequencies will remain constant in a population from one generation to the next (Butler, 2005; Nei & Kumar, 2000). In order for any population to be within HWE the following criteria needs to be met: random mating, an infinitely large population, no mutations, and no migrations into or out of the population and no selection for any genotypes (Butler, 2005; Nei & Kumar, 2000). In reality no such population exists. However, the HWE is still used to determine genotype frequencies (Table 1.1).

1.7.4.2.2 Observed heterozygosity

The number of homozygotes plus the numbers of heterozygotes equals 100% of the samples tested in a population. Heterozygosity (Table 1.1) displays the proportion of heterozygous individuals in the population. The higher the heterozygosity value the more allele diversity exists for that particular locus (Butler, 2005). This therefore means that there is less of a chance of a random sample matching.

1.7.4.2.3 Match probability, Power of discrimination and Power of exclusion

Match probability (MP) is defined as the number of individuals which needs to be surveyed before obtaining the observed DNA pattern in a randomly selected individual (Huston, 1998). MP is used to calculate the power of discrimination of a locus (Halder & Akhteruzzaman, 2006). The power of discrimination (PD) and the power of exclusion (PE) are used to exclude falsely accused individuals (Table 1.1). The greater the PD of a locus, the greater the likelihood that it will successfully discriminate between individuals of a population (Halder & Akhteruzzaman, 2006). The PE represents the percentage of individuals within a population who would not share the same genotype as that of the father in a paternity case (Halder & Akhteruzzaman, 2006). Thereby inferring that the higher the PE value of the locus the more non-fathers are excluded. Multiple loci can be used to increase the strength of PD and PE values. These are referred to as the Combined PD (CPD) and the Combined PE (CPE).

1.7.4.2.4 Polymorphism information content and Typical paternity index

The polymorphism information content (PIC) is a measure of polymorphism for a locus used in linkage analysis. It is the probability whereby a rare allele at a locus shown in a child can be traced back to the parent carrying the rare allele (Table 1.1) (Butler, 2005). PIC values can range from 0–1. A PIC value of zero means that the locus only has one allele; whilst a PIC value of one means that the locus has an infinite number of alleles (Hildebrand *et al*, 1994). PIC values ≥0.7 are considered to be highly informative (Hildebrand *et al*, 1994).

The typical paternity index (TPI) displays the likelihood that the alleles shown by the child were obtained from the true biological father instead of a random unrelated male in the same population (Table 1.1) (Butler, 2005). TPI is a ratio of two numbers where the numerator is the probability that the sperm that fertilized the mother's ovum came from the biological father whilst the denominator is the probability that the sperm that fertilized the mother's ovum came from a random male (Butler, 2005). The greater the TPI value of a locus, the better the locus ability to identify the biological father (http://www.dnacenter.com/science-technology/paternity). To improve the odds, a combined TPI (CTPI) is used in paternity cases. CTPI is the product of several loci's TPI values.

1.7.4.2.5 Population differentiation

Population differentiation is determined from allelic frequencies (Raymond & Rousset, 1995b). As its title suggests it is used to determine whether genetic differentiation exists between populations or not. It varies between 0–1 where zero is indicative of no differentiation existing between populations, whilst one indicates complete differentiation (Raymond & Rousset, 1995b). When the null hypothesis, i.e. no differentiation amongst populations, demonstrates a *p*-value <0.05 it is said to show significant differentiation (Raymond & Rousset, 1995b).

Table 1.1: Summary of genetic formulae

HWE formula	$\mathbf{p^2 + 2pq + q^2 = 1,}$ where: p = frequency of the dominant allele and, q = frequency of the recessive allele
Ho	$\mathbf{1 - \sum_{i}^{n} P_i^2,}$ where P_i = frequency of i^{th} allele in a population of n samples
MP	$\sum_{i=1}^{n} \mathbf{x_i^2,}$ where n = number of samples and x_i = frequency of i^{th} genotype
PD	**1-MP**
PE	$\mathbf{H^2 (1 - 2 \times H \times h)^2}$, where H = heterozygosity and h = homozygosity
CPD	$\mathbf{1 - (MP_{locus\ 1} \times MP_{locus\ 2} \times ……..)}$
CPE	$\mathbf{1 - (PE_{locus\ 1} \times PE_{locus\ 2} \times ……..)}$
PIC	$\sum_{1}^{n} \mathbf{P_i^2 - (\sum_{1}^{n} P_i^2)^2 + \sum_{1}^{n} P_i^4,}$ where P_i is frequency of i^{th} allele in a population of n samples
TPI	$\frac{\mathbf{H + h}}{\mathbf{2h}}$ where H = heterozygosity and h = homozygosity
CTPI	$\mathbf{1 - (TPI_{locus\ 1} \times TPI_{locus\ 2} \times ……..)}$

Genetic formulae taken from Butler, 2005; Huston, 1998.

1.8 Objectives of the study

This study is part of a comprehensive project aiming at developing and validating Non-CODIS miniSTR genotyping systems suitable for South Africa. The development of such genotyping systems will help in forensic casework involving DNA samples and in particular those who are highly degraded. The Non-CODIS miniSTRs can be valuable in a variety of scenarios including complex paternity cases, missing persons work, and mass fatality DNA identification testing. In these Non-CODIS miniSTR systems, smaller sized PCR products are amplified, and PCR primers are moved as close as possible to the STR repeat regions. This has led to a decrease in the incidence of allele or/and locus-dropout that may occur in degraded DNA samples.

It is believed that Non-CODIS miniSTRs can in the future provide a potential alternative to mitochondrial DNA sequencing in the forensic analysis of degraded DNA. Indeed, mtDNA testing is a laborious and cost-prohibitive procedure for most forensic laboratories to utilize. Additionally, given the haploid and maternal-only transmission of mtDNA, the power of discrimination is not as powerful for identification as a full multi-locus STR match. The Non-CODIS miniSTR assays should offer a new potential tool for recovering useful information from samples that generated partial profiles with present STR multiplexes.

The present study covers six Non-CODIS miniSTR loci amplified in two miniplex sets. This includes D1S1627, D5S2500 & D8S1115 amplified in miniplex 1 and D3S4529, D6S1017 & D9S2157 amplified in miniplex 2.

In the initial part of the study the six Non-CODIS miniSTR genotyping system described above was evaluated on recent DNA samples (Chapter 2). The amplification conditions

were fine tuned to obtain the optimal results. Allelic ladders necessary for the system were also developed.

For proper allelic designation, two homozygous alleles per locus were fully sequenced (Chapter 3). The sequence and the repeat structure of each of the loci were also confirmed.

The exercise of internal validation of the Non-CODIS genotyping system was carried out (Chapter 4).

A comprehensive population study, covering five population groups from South Africa, was conducted (Chapter 5).

Chapter 2: Laboratory development and implementation of a six Non-CODIS miniSTR genotyping system

2.1 Introduction

The six Non-CODIS miniSTR genotyping system investigated here includes six loci amplified in two miniplex sets. Miniplex 1 includes D1S1627, D5S2500 & D8S1115 and miniplex 2 includes D3S4529, D6S1017 & D9S2157. Primer sequence and amplification conditions were previously described by Hill and co-workers (2008). The system was evaluated on recent DNA samples in this part of the study. The amplification conditions were fine tuned to obtain the optimal results. Allelic ladders necessary for the system were also developed.

2.2 Materials and methods

2.2.1 PCR amplification

Primers were synthesized by Applied Biosystems and all forward primers were labelled with either 6FAM™, VIC™ and NED™ dyes which enabled colour separation on an ABI 377 genetic analyzer. All reverse primers were unlabelled and had an additional 5' guanine base added to produce fully adenyalted PCR products (Hill *et al*, 2008). The chromosomal position, location, repeat type, repeat motif, allele range and the size range for each locus is presented in Table 2.1. Complete primer sequences and final concentrations of each of the loci used in this study are presented in Table 2.2.

Amplification of samples were performed in reaction volumes of 10µl using a master mix containing: 1X Supertherm PCR buffer (containing 1.5mM $MgCl_2$); 250µM of each

deoxynucleotide triphosphate (Roche, dNTP's: dATP, dCTP, dGTP, dTTP); 200pg of template DNA, 160µg/ml BSA and 1U Supertherm Gold Taq DNA polymerase. Primers were added to produce final concentrations as indicated in Table 2.2. Amplification reactions were performed using a 96-well Gene-Amp® PCR system 9700 thermal cycler (Applied Biosystems). Thermal cycling conditions for miniplex 1 were: 95°C for 10 minutes; 35 cycles of: 94°C for 1 minute, 59°C for 1 minute and 72°C for 1 minute followed by a final incubation at 60°C for 45 minutes. Thermal cycling conditions for miniplex 2 were: 95°C for 10 minutes; 35 cycles of: 94°C for 1 minute, 55°C for 1 minute and 72°C for 1 minute followed by a final incubation at 60°C for 45 minutes.

Table 2.1: Information on six Non-CODIS miniSTR loci examined in this study. Adapted from Hill *et al* (2008).

STR locus	Chromosomal position	Chromosomal location	Repeat type	Repeat motif	Allele range	Size range (bp)*
D1S1627	Chr 1 106.676 Mb	1p21.1	Tri	ATT	10-17	84-105
D3S4529	Chr 3 85.935 Mb	3p12.1	Tetra	ATYT	11-19	113-145
D5S2500	Chr 5 58.735 Mb	5q11.2	Complex tetra	GRYW	13-24	83-127
D6S1017	Chr 6 41.785 Mb	6p21.1	Tetra	ATCC	7-18	84-128
D8S1115	Chr 8 42.656Mb	8p11.21	Tri	ATT	7-21	60-102
D9S2157	Chr 9 133.065Mb	9q34.2	Tri	ATA	7-17	74-104

*Amplicon size obtained in this study according to GS500 LIZ size standard

R=G/A
Y=T/C
W=A/T

Table 2.2: Dye labeled primer sequences used in miniplex 1 and miniplex 2 reactions

STR locus	Miniplex set	Primer sequence (5'-3')	Primer concentration (μM)
D1S1627	1	F-[VIC]-CATGAGGTTTGCAAATACTATCTTAAC R-GTTTTAATTTTCTCCAAATCTCCA	0.3
D3S4529	2	F-[VIC]-CCCAAAATTACTTGAGCCAAT R-GAGACAAAATGAAGAAACAGACAG	0.1
D5S2500	1	F-[NED]-CTGTTGGTACATAATAGGTAGGTAGGT R-GTCGTGGGCCCCATAAAATC	0.1
D6S1017	2	F-[VIC]-CCACCCGTCCATTTAGGC R-GTGAAAAAGTAGATATAATGGTTGGTG	0.1
D8S1115	1	F-[6FAM]-TCCACATCCTCACCAACAC R-GCCTAGGAAGGCTACTGTCAA	0.1
D9S2157	2	F-[6FAM]-CAAAGCGAGACTCTGTCTCAA R-GAAAATGCTATCCTCTTTGGTATAAAT	0.5

2.2.2 Allelic ladder design

To facilitate consistent genotyping from one gel run to another, allelic ladders were designed for each locus. The allelic ladders were designed using a combination of individual DNA samples which were representative of the most common alleles present in the populations (Coble and Butler, 2005). Amplification of allelic ladders was performed using the same set of primers. From this a single amplification contained a range of reference amplicons which could be used for comparison when typing unknown samples. Two alleles for each ladder were selected for sequencing to verify accurate allele designation, repeat structure and sequence. DNA sequencing is discussed in detail in Chapter 3.

DNA samples used for ladder construction were diluted to concentrations of 1ng. 1µl of each necessary DNA sample for the ladder was combined in a 1.5ml eppendorf tube and made up to a final volume of 10µl thereby producing DNA mixtures at a final concentration of 100pg. The PCR master mix contained the same concentrations of reagents as previously described for miniplex 1 & 2 amplfication (see 2.2.1) and the DNA mix was added to produce a final concentration of 200pg in a 10µl reaction. Ladder samples for miniplex 1 used the following cycling conditions: 95°C for 10 minutes; 35 cycles of: 94°C for 1 minute, 59°C for 1 minute and 72°C for 2 minutes followed by a final incubation at 60°C for 60 minutes and 4°C forever. Ladder samples for miniplex 2 used the following cycling conditions: 95°C for 10 minutes; 35 cycles of: 94°C for 1 minute, 55°C for 1 minute and 72°C for 2 minutes followed by a final incubation at 60°C for 60 minutes and 4°C forever. The 72˚C extension time and the final extension time was increased since more PCR products had to be created (Coble and Butler, 2005).

2.2.3 Fragment analysis on ABI 377 Genetic Analyzer

Amplified samples were electrophoretically separated using an ABI Prism® 377 Genetic Analyzer (Applied Biosystems) and GeneScan™ 500 LIZ™ size standard (Applied Biosystems, Warrington, UK). Samples were prepared for analysis by mixing 1µl of dye mix (i.e. 5µl de-ionized formamide, 1µl dextran blue and 1µl of GS500 LIZ size standard [Applied Biosystems]) with 1µl of PCR product. Mixtures were heat denatured on a thermocycler at 95°C for 5 minutes and then immediately snap cooled on ice. 1µl of

sample was loaded on a 36cm gel, (for gel preparation refer to Appendix I), and filter set G at 2400 scans per hour.

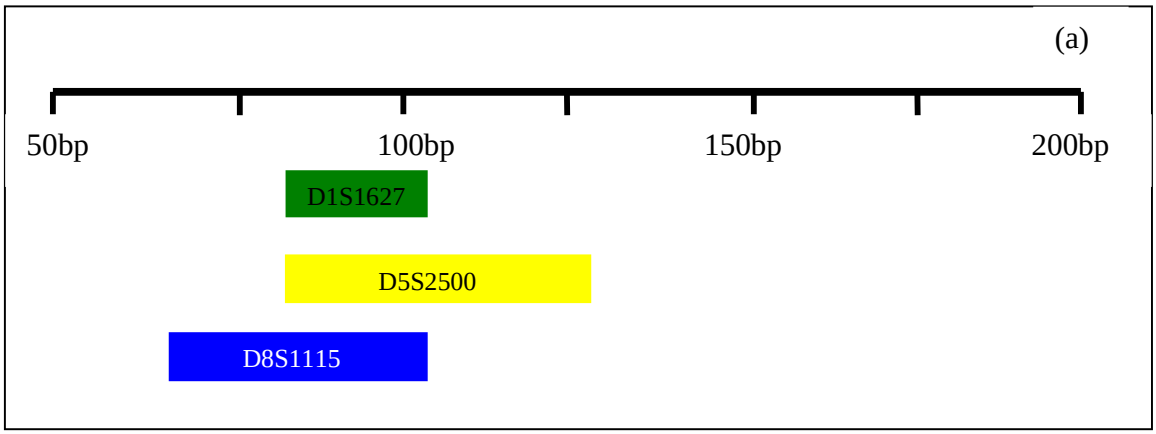
All data was collected with the ABI 377 collection software (Applied Biosystems) and analyzed using the GeneScan 3.1 (Applied Biosystems) software. Size fragments were converted to allele numbers by making use of the Genotyper 2.5 (Applied Biosystems) software. Allele numbers were assigned to samples according to the known allele numbers in the allelic ladders.

2.3 Results and Discussion

The performance of six Non-CODIS miniSTR loci were evaluated in this study. Observed allele size ranges for each of the loci was determined (Figure 2.1 (a) and Figure 2.2 (a)). In miniplex 1, locus D1S1627 displayed amplification products in the size range of 84-105bp, D5S2500 showed a size range of 83-127bp and D8S1115 showed a size range of 60-102bp. In miniplex 2, locus D3S4529 displayed amplification products in the size range of 113-145bp, D6S1017 showed a size range of 84-128bp and D9S2157 showed a size range of 74-104bp.

All six loci were consistently amplified in the test samples. Loci generally showed balanced profiles (Figure 2.1 (b) and Figure 2.2 (b)), but in some instances D9S2157 did not amplify as strongly as the other two loci in miniplex 2. In most cases the RFU value of D9S2157 did not fall below the threshold limit 150RFU. The presence of PCR artefacts was generally detected in D9S2157 (indicated by arrow in Figure 2.2 (b)). PCR artefacts are normally caused by interactions between primers in the multiplex reaction.

These artefacts however fell outside of the size range of the alleles of interest and were thus disregarded during analysis.



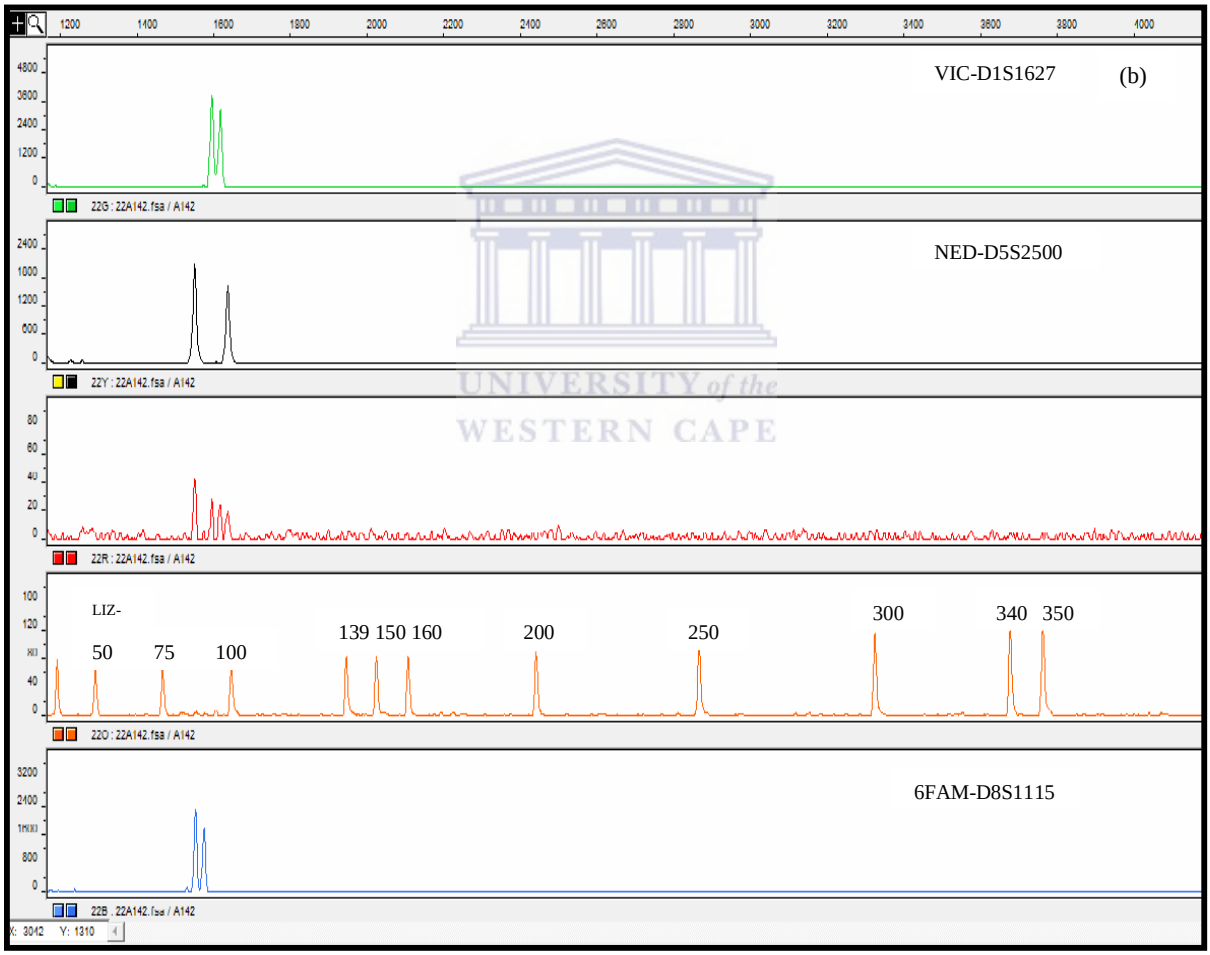


Figure 2.1: DNA fragment analysis for miniplex 1. (a) Observed allele size range of the three loci amplified in miniplex 1. (b) An example of an electropherogram for miniplex 1 illustrating colour separated allele peaks.

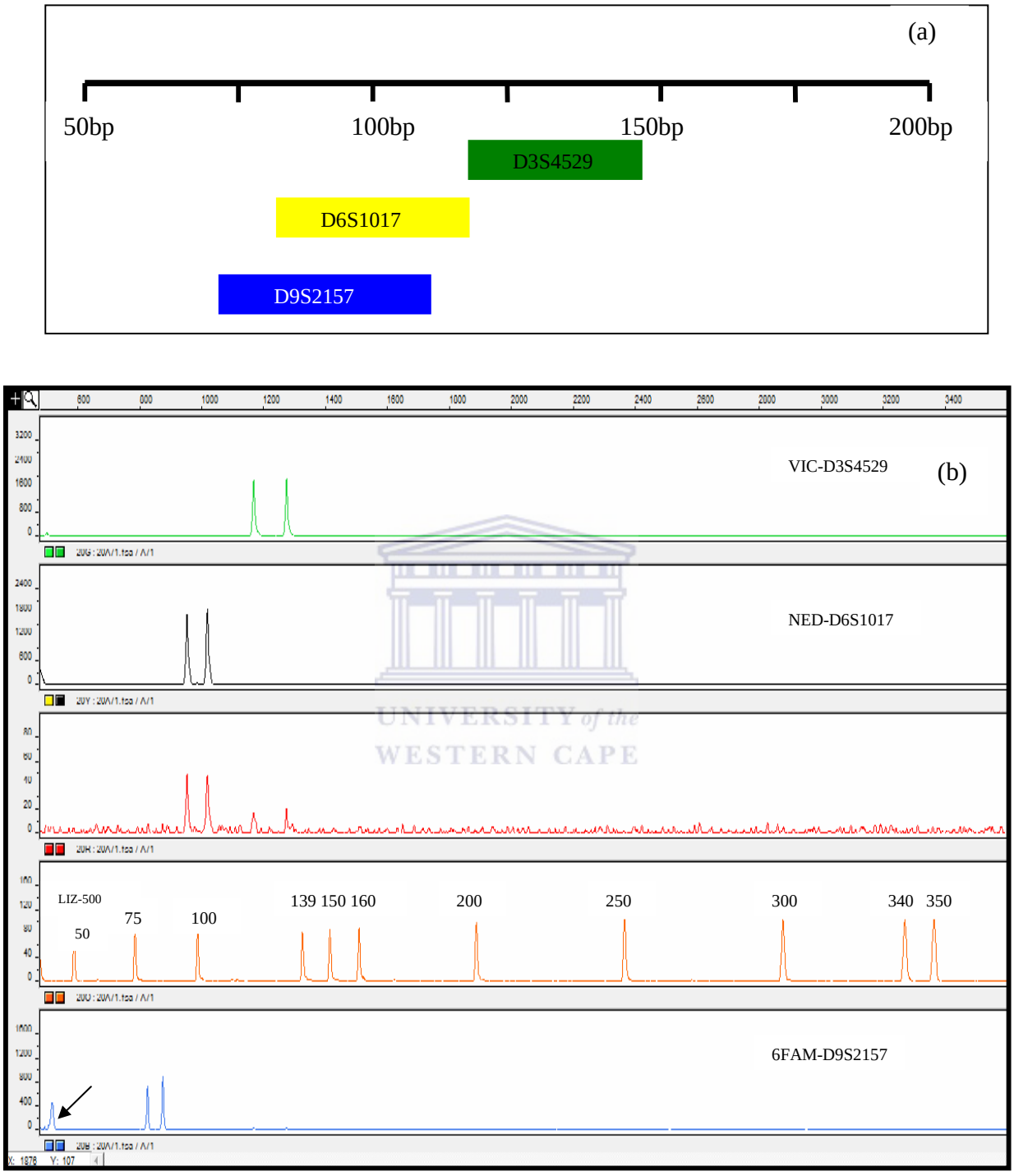


Figure 2.2: DNA fragment analysis for miniplex 2. (a) Observed allele size range of the three loci amplified in miniplex 2. (b) An example of an electropherogram for miniplex 2 illustrating colour separated allele peaks.

Prior to the design of allelic ladders, sequencing analysis of individual samples were conducted to ensure that correct allelic designations were made in accordance with the recommendations of the DNA Commission of the ISFG (Bar *et al*, 1997). The DNA sequencing procedure is explained in detail in Chapter 3. Allelic ladders were designed for each locus to enable accurate genotyping of unknown samples from one gel run to another. Figure 2.3 (a)-(c) and Figure 2.4 (a)-(c) depicts allelic ladder electropherograms for each of the loci.

After the completion of the population study additional rare alleles not present in the current allelic ladders were detected. In miniplex 1 these additional rare alleles were, repeats 17 for locus D1S1627, repeats 13, 16 and 22 for locus D5S2500 and repeats 12 and 20 for the locus D8S1115. In miniplex 2 these additional rare alleles were repeats 19, 14 & 18, and 9.1 for the loci D3S4529, D6S1017 and D9S2157 respectively. The rare allele of locus D9S2157 is known as a microvariant allele. This microvariant allele displayed by locus D9S2157, has also been reported in USA populations and in Singaporean populations (Hill *et al*, 2006b; Yong *et al*, 2009). The microvariant allele was confirmed by sequencing with flanking primer pairs that included the miniSTR amplicon. The repeating structure contained eight complete repeating units of ATA followed by one repeat of ATTA. Thus an additional 'T' base is present in the final repeating unit (Figure 2.5 displays the sequencing of a homozygous 9.1 allele).

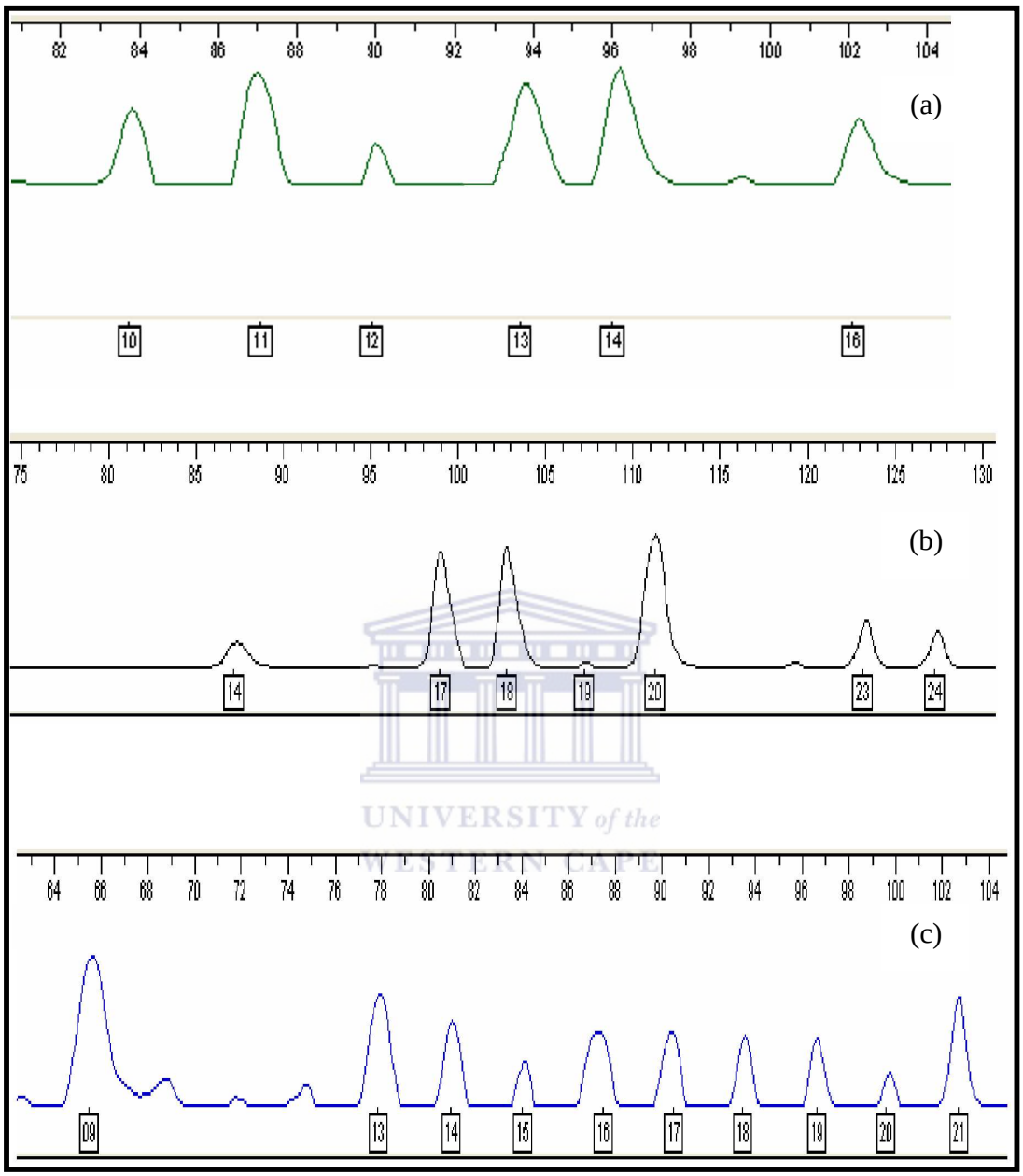


Figure 2.3: Miniplex 1 allelic ladders. Alleles from individual DNA samples representative of the common alleles for each locus present in miniplex 1. (a) Genotyped allelic ladder profile for locus D1S1627. (b) Genotyped allelic ladder profile for locus D5S2500. (c) Genotyped allelic ladder profile for locus D8S1115.

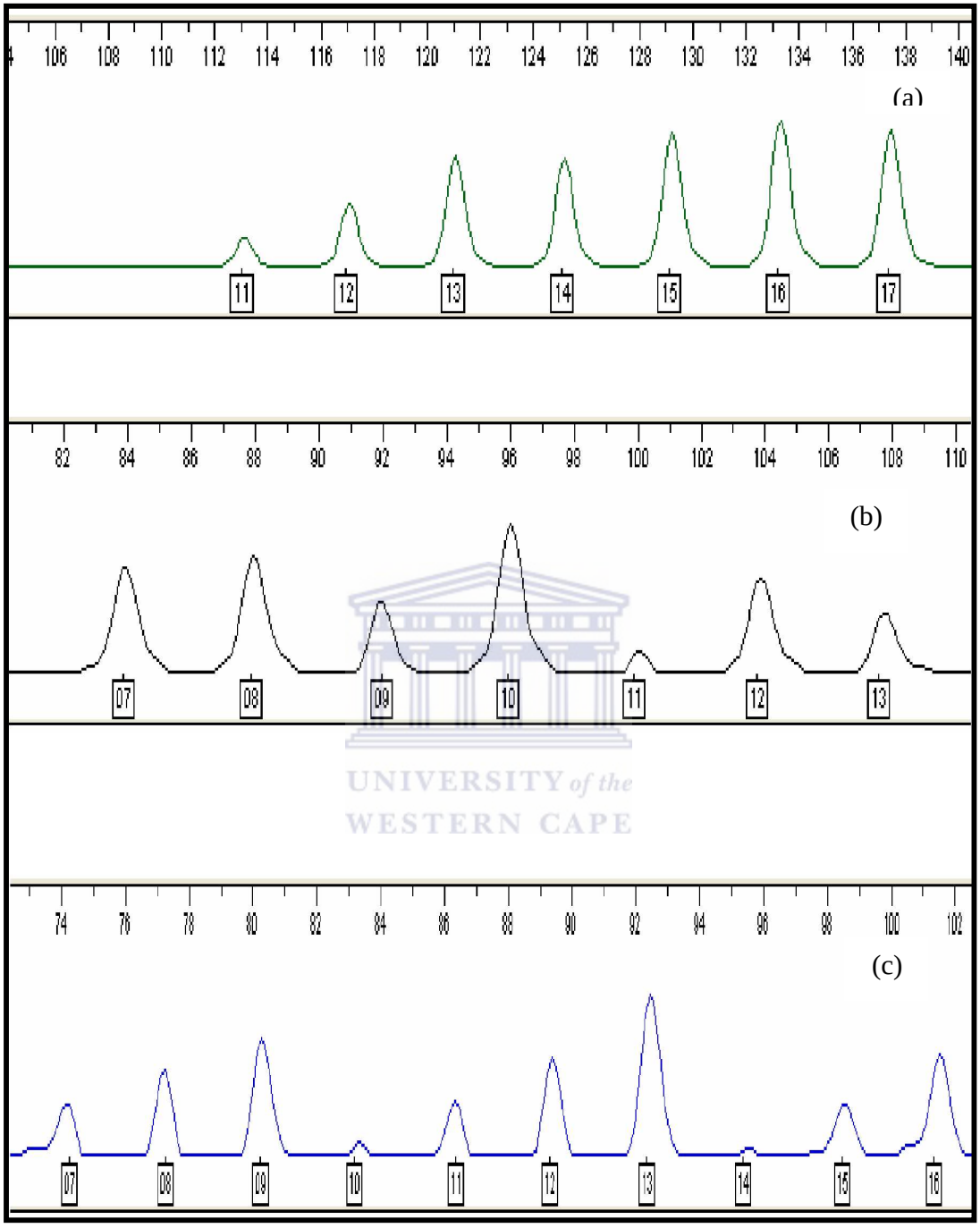


Figure 2.4: Miniplex 2 allelic ladders. Alleles from individual DNA samples representative of the common alleles for each locus present in miniplex 2. (a) Genotyped allelic ladder profile for locus D3S4529. (b) Genotyped allelic ladder profile for locus D6S1017. (c) Genotyped allelic ladder profile for locus D9D2157.

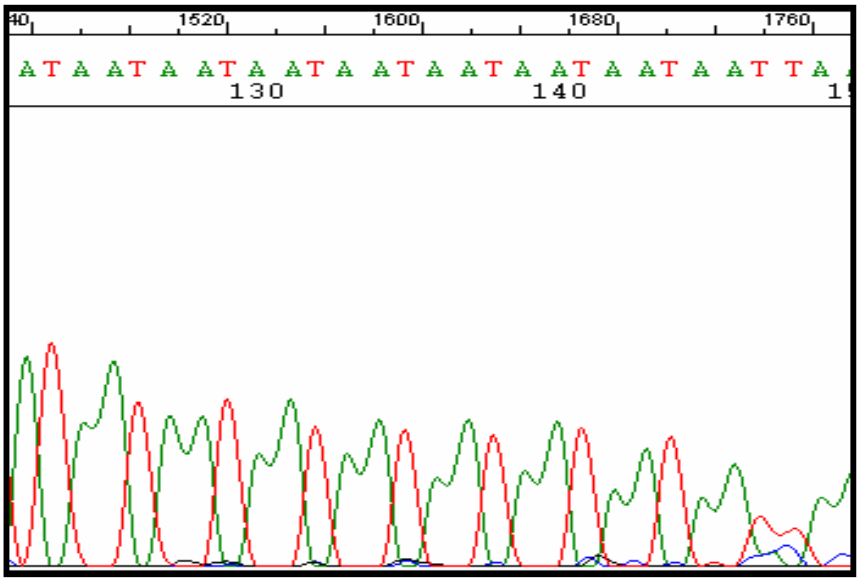


Figure 2.5: **Sequencing of 9.1 microvariant allele for the locus D9S2157.** The sequencing product of a homozygous sample identified for the marker, D9S2157 with 9.1 repeating units of the trinucleotide ATA. Eight complete repeating units were counted and one repeat unit containing and additional 'T' nucleotide was observed.

2.4 Summary

The six Non-CODIS miniSTR genotyping system was successfully implemented in the laboratory following minor optimizations of the loci amplification conditions. Allele size range was determined for each of the loci and allelic ladders were constructed. In both miniplex sets balanced profiles were obtained for all loci except for the locus D9S2157 in miniplex 2. To overcome its amplification weakness the concentration of D9S2157 in the primer mix was increased to aid its successful amplification.

Following the promising preliminary results, the system was further investigated to assess its suitability for forensic casework in South Africa. This included internal validation and population studies.

Chapter 3: DNA sequencing of six Non-CODIS miniSTR loci

3.1 Introduction

Variations within STR loci are displayed by the number of repeating units and the repeating structure. Thus, when using these loci in human identification tests the sequence and structure of the repeating unit must be determined. Repeat structure and numbers per locus are normally verified by means of DNA sequencing (de Kock, 2006). A minimum of two homozygous samples for each miniSTR locus was sequenced to verify repeat structure, repeat number and observed size. To ensure that the entire repeat region was being sequenced, sequencing primers were designed outside of the target miniSTR amplicon regions. This enabled the sequencing of the full repeat region.

3.2 Materials and methods

3.2.1 Primer design

Unlabelled sequencing primers were designed by extending the sequences on both ends of the miniSTR primers. To determine the additional bases that were needed to be added, miniSTR primer sequences were copied into BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat) and an extended DNA sequence was obtained. Primers were then designed against these reference sequences using a web-based primer design tool, Primer3. BLAST and an electronic PCR programme (http://lpgws.nci.nih.gov/perl/epcr) were used to verify chromosomal location. Refer to Table 3.1 for: sequencing target region for each locus, primer sequences and product size. All sequencing primers were synthesized by Whitehead Scientific (Pty) Ltd.

Table 3.1: Unlabeled primers used for sequencing purposes of the six Non-CODIS miniSTR loci investigated in this study. These primers bind outside of the miniSTR primers and have larger amplicon sizes.

STR locus	Target Region (5'-3')	Product size (bp)
D1S1627	***CTGTAACTGTGGAGTATAAAC***TATTCTTAAGTAGAAAGACTAAAAGATGAACCAAACAAAAGTTTTACTATTTA TGACTTTTTTTGTGTATTCATTGTATATTTTTGGATTCGAGGTTACCATGAGGTTTGCAAATACTATCTTAACA CCCATTATTATTATTATTATTATTATTATTATTATTATTATTGGAGATTTGGAGAAAATTAAAATTAATGTGTG TGCGTGTGTGTGTGTTTATATTCCAGGGTATTGTGCACGTTAACTATTTAACCTATTCAAGACATAGACAGT ACAGTAAGGTATGAATAGA**AATGACCAAAAGTTAAACAGTGGC**	339
D3S4529	***GCCTTATGAGAAAGTCTGAAGTCC***TCTAAGTACCCTAAGGAAGAAGAAATTCTCCCTGACACTG[CCTCTGGAC ATGAGCTGCAACAGAATCTCTTTGGGTCTCCAGCCTACCAGCCTGCCCTGCATTTTTCAAACTTGCCAGACCCC AAAATTACTTGAGCCAATTTATTAAAATAAATATCGATCAATTGATCTATCTATCTATCTATCTATCTATCTAT CTATTTATCTATCTATCTATCTGTCTGTTTCTTCATTTTGTCTCCCTGGAGGCCCCTAACAGTTCCTTTATCCC TGGTTCATTCTTCTGCATAGCACTTGTCCCCTTTTAATGTATGATATAATGTATTTCATTTGCATGTTTCTTTA T**TGTCTGTCTCCTCCACTACACTGT**	395
D5S2500	***TGTTTTATTTATGGAACAACTTTTTG***TTTTTTCTGGAGTTATATATTACCTTCTTTATTTGATTATGTGACATT ATCACCAATTTTTCTAGACGTCTCCAAAACATAATTTTTTAAATTTTAAATTTCTGTTGGTACATAATAGGTAG GTAGGTAGACAGACAGACAGACAGACAGACAGACAGACAGACAGATAGATAGATAGATTGATTGATTTATGGGGCCCA CGAGATATTTTCATACAGGCAAGCAATGCATAATAATATCAGGGTAAACAGGGAATCTATTACTGCAAACATTT ACCCTTTAAGTTACAAACAATTCAATTATATTCTCTTAGTTATTTT**GAGATGTATGATAAACTACTGTTGAC**	368
D6S1017	***ACAACTAGTGCCTTTTTCATCTCC***TTTTAAACATCCATCCATCTATCCATGCATACATCTATCCATTATATCCA CCTTTTCATCCATTATATCCATCCATCCATCCATCCACCCACCATTATATCCACCTTTTCATCTGTCCACCCGT CCATTTAGGCATGCATCCATTATATCCATCCATCCATCCATCCATCCATCCATCCATCCATCCACCAACCATTA TATCTACTTTTTCATGCATCCACCCATCCATTTATGCATGCATCCATTATGTCCATCCATCCATTATATACATG CATCCATCTAATATTTATTGAGCACCTATCCTGTGCCAGGTACTGCATTAGATACTTTATGTTT**AATCCTTACA GAAATTCTCCAGGG**	384
D8S1115	***TTTTCTATATTCAGAAGGGGAACG***GCTGGACCATATCCCATAGTAATTCTATTCTTAATGTTTTCAGGAACCAC CATAATGTTTTCCACAGCGGCCGCACCATTTCACATATCCACCAGAGGCCCACAAGAGTTCCAGTTTCTCCACA TCCTCACCAACACATTATTATTATTATTATTATTATTATTTTGACAGTAGCCTTCCTAGGCCATGGGGCTTTTA AGCTGTATGATTAAGGAAATTTGTGCAAATGCAGTTTTTAAAACTCTCTTGGCCAGGTGCGGTGGCTCATGCCT GTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCAGATCACCTGAGGTCGAGTTCCAGACCAGCCTGACCAACA**T GGTGAAACCCTGTCTCTACTAAA**	393
D9S2157	***AAATACAAAAATTAGCGTGTGTGC***CTGTAAT[CCCAGCTATTCAGGTGGCTGAGGCATGACAATTGCTTGAACC CGGGAGGTGGAGGTTGCAATGAGCTGAGATCACGCCACGGTACTCCAGCCTGGGTGGCAAAGCGAGACTCTGTC TCAATATAATAATAATAATAATAATAATAATAATAATAAATAAAATTTATACCAAAGAGGATAGCATTTTCCCTCTC CCAATCAGCAAGGTAGGTTCATTTTTGTTTTCATAACCTCTGGTATTCCCAGCAGTAGCCCAGGTGTACTACAC GGTGGCTTTGCTGGGTACT]GCTTGGGCAGGGCACATGAATGAGTGACTGATGTGCTGTGGTC**CTTCTGGAAAT AACCCAATGAGTT**	383

Forward sequencing primers
Reverse sequencing primers
Forward miniSTR primers
Reverse miniSTR primers

3.2.2. PCR amplification

Sequencing amplification reactions were performed in reaction volumes of 30 µl using a master mix containing: 1X Supertherm PCR buffer (containing 1.5mM $MgCl_2$); 250µM

of each deoxynucleotide triphosphate (Roche, dNTP's: dATP, dCTP, dGTP, dTTP); 1µl of template DNA with a concentration ranging between 1ng-10ng; 160µg/ml BSA and 1U Supertherm Gold Taq DNA polymerase. Primer concentrations were adjusted to produce visible amplicons. Amplification reactions were performed using a 96-well Gene-Amp® PCR system 9700 thermal cycler (Applied Biosystems). Thermal cycling conditions were: 95°C for 10 minutes; 35 cycles of: 94°C for 1 minute, 57°C/59°C for 30 seconds and 72°C for 30 seconds followed by a final incubation at 60°C for 45 minutes and 4°C forever. Annealing temperatures; final primer concentrations and DNA concentrations used for each primer pair is presented in Table 3.2.

Table 3.2: Amplification conditions for DNA sequencing of the six Non-CODIS miniSTR loci

Sequencing primer pair	Annealing temperature (˚C)	Primer concentration (µM)	DNA concentration (ng)
D1S1627	57	0.3	1.0
D3S4529	57	0.1	1.0
D5S2500	57	0.1	10.0
D6S1017	57	0.1	1.0
D8S1115	59	0.1	10.0
D9S2157	57	0.1	1.0

Post amplification, sequencing products were run on a 2% agarose gel and stained with ethidium bromide to enable visualization under UV light. Agarose gels were run for 30 minutes to 1 hour at a constant current of 80 ampere in 1X TBE buffer.

3.2.3 Sequencing procedure

For the sequencing reaction 1µl of ExoSap IT and 2.5µl of PCR product (see section 3.2.2) were mixed and incubated at 37^{o}C for 15 minutes followed by a further 15 minute incubation at 80^{o}C in a 96-well Gene-Amp® PCR system 9700 thermal cycler (Applied Biosystems). Sequencing PCR was then performed using ABI Prism® BigDye® Terminator (version 3.1) cycle sequencing kit (Applied Biosystems). The sequencing mixture contained: 3.5µl of ExoSap IT template, 4µl BigDye Terminator kit mix (Applied Biosystems), 1.5µl distilled water and 1µl of forward or reverse primer (5µM each) to produce a final volume of 10µl. Thermal cycling was conducted in a 96-well Gene-Amp® PCR system 9700 thermal cycler (Applied Biosystems) with the following parameters: 96°C for 1 minute followed by 25 cycles of: 96°C for 10 minutes; 50°C for 5 minutes and 60°C for 4 minutes followed by 4°C for 60 minutes.

The DNA sequencing product was cleaned up as follows: 10µl of sequencing product was mixed with 2µl of 125mM EDTA:3M Sodium acetate which was previously prepared at a ratio of 1:1. 25µl of 100% ethanol was then added to each sequencing reaction and incubated at room temperature for 15 minutes. Post the 15 minute incubation; samples were spun at maximum speed for 30 minutes at room temperature using a bench top centrifuge. Supernatant was removed and the pellet was washed with 70µl ice cold 70% ethanol. This was followed by a 15 minute spin (at maximum speed) at room temperature. Supernatant was removed and the pellet was allowed to air dry in a fume hood. Once completely dry the pellet was resuspended in 1.5µl of loading mix (which contained: 5µl formamide and 1µl of dextran blue dye [Applied Biosystems]).

Prior to loading, samples were heat denatured for 4 minutes on a thermocycler set at 95˚C and then immediately snap cooled on ice until samples were loaded.

0.8-1.2µl of sample was loaded on a gel (for gel preparations refer to Appendix I). Electrophoresis of sequencing products were performed on the ABI Prism® 377 Genetic Analyzer (Applied Biosystems) using a 36cm gel and filter set E at 1200 scans per hour. Sequencing gels were analyzed using Sequencing Analysis version 3.7 (Applied Biosystems). Forward and reverse sequences were aligned and checked using an online sequence alignment tool, i.e. BioEdit Sequence Alignment Editor which is available from: http://www.mbio.ncsu.edu/BioEdit/page2.html. Allele repeats for each sample was determined and repeat numbers in relation to size were assigned accordingly.

3.3 Results and Discussion

The aim of DNA sequencing was to verify that correct allele designations were being assigned to each locus. Hence, size in bp could accurately be assigned to the number of allele repeats. For example, Figure 3.1 (a) has a size of 84bp which displays 10 repeating units of ATT for the locus D1S1627. Therefore all amplicons with a size of 84bp for the marker D1S1627 will have 10 repeating units attributed to it. Thus, from this data one could then predict the range of allele repeats for the marker using the amplicon length in bp. An increase of 3bp for the marker D1S1627 will result in an extra repeat, thus 11 repeats will be assigned the size of 87bp and a decrease of 3bp for the same marker will result in 9 repeats being assigned to an amplicon length of 81bp.

Applying this correlation to all the markers used in this study, a trinucleotide allele will increase or decrease by 3bp for each repeat and a tetranucleotide allele will increase or

decrease by 4bp for each repeat. In this study the loci: D1S1627, D8S1115 and D9S2157 contain trinucleotide repeats (Figure 3.1 (a), (e) and (f) respectively) and the loci D3S4529, D5S2500 and D6S1017 contain tetranucleotide repeats (Figure 3.1 (b), (c) and (d) respectively).

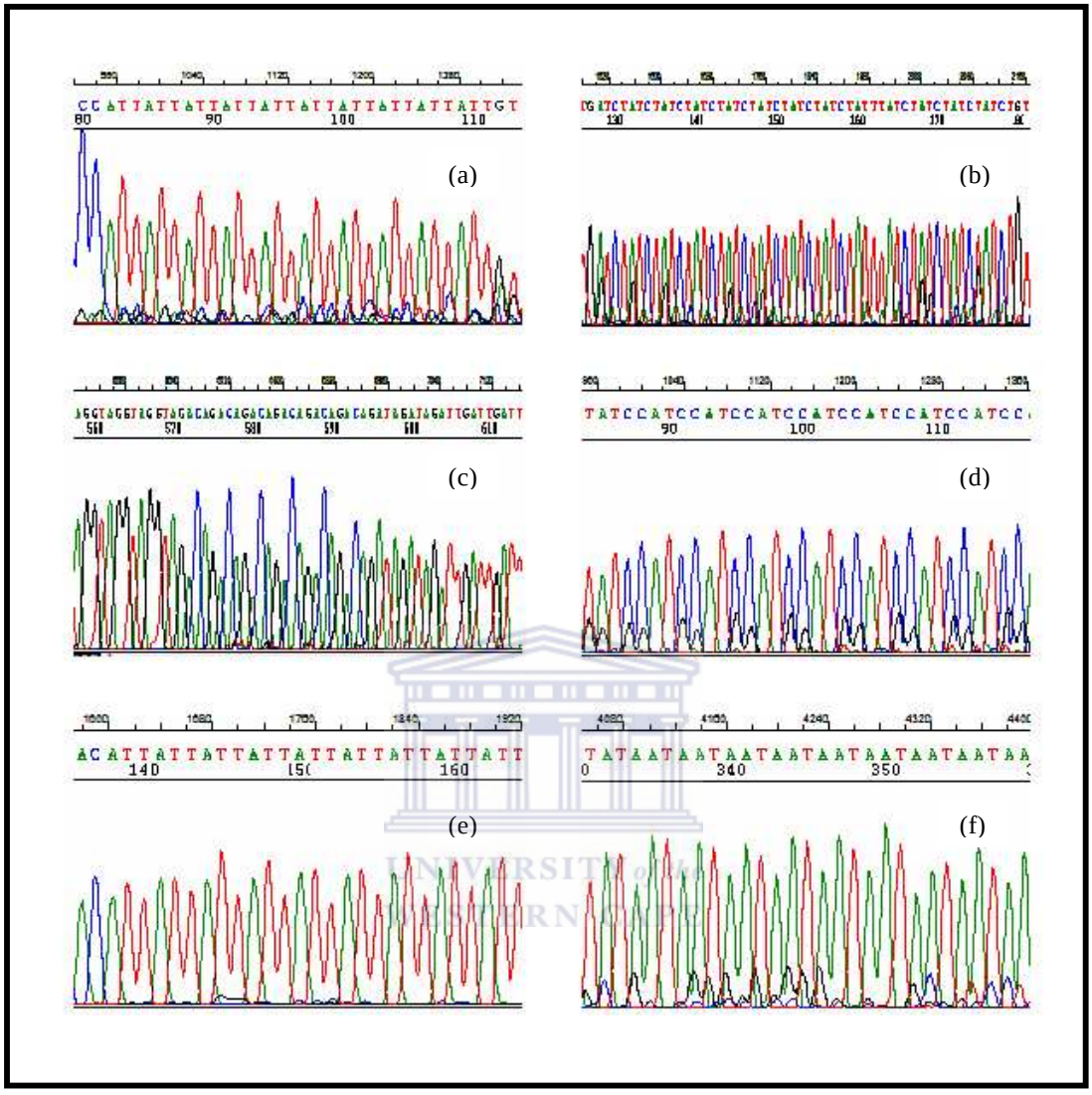


Figure 3.1: DNA sequencing profiles of homozygous samples tested for each locus. (a) Sequencing profile for locus D1S1627 illustrating 10 repeating units of ATT (84bp). (b) Sequencing profile for locus D3S4529 illustrating 13 repeating units of ATCT/ATTT (121bp). (c) Sequencing profile for locus D5S2500 illustrating 14 repeating units of GGTA/GACA/GATA/GATT (87bp). (d) Sequencing profile for locus D6S1017 illustrating 8 repeating units of ATCC (88bp). (e) Sequencing profile for locus D8S1115 illustrating 9 repeating units of ATT (66bp). (f) Sequencing profile for locus D9S2157 illustrating 9 repeating units of ATA (80bp).

Confirmation of accurate allele repeat number designation was tested by comparing the allele repeat numbers (genotype profile) obtained for the standard DNA sample 9947A found in this study to those generated by Hill *et al* (2008). Concordance was obtained for all loci in both miniplex sets (Table 3.3).

Table 3.3: Concordance study of standard reference sample 9947A

		Genotype profile obtained for the sample 9947A in this study	Genotype profile obtained for the sample 9947A by Hill *et al* (2008)
Miniplex 1	D1S1627	13:14	13:14
	D5S2500	14:23	14:23
	D8S1115	9:18	9:18
Miniplex 2	D3S4529	13:13	13:13
	D6S1017	9:10	9:10
	D9S2157	7:13	7:13

3.4 Summary

The entire repeat regions of six Non-CODIS miniSTR loci were successfully sequenced. Allele repeat number, structure and observed size were determined for each locus. Subsequently, allele repeat numbers could accurately be assigned to all samples. The sequencing procedure was also used to evaluate flanking regions of each locus. The sequencing results generated from this study showed consistency in this regard for each of the investigated loci.

Chapter 4: Internal validation study of six Non-CODIS miniSTR loci

4.1 Introduction

Currently forensic analysis is being more frequently used in court cases than in the past (Butler, 2005). In most instances the performance of the genotyping system used comes under scrutiny instead of the science behind the selected methods (Butler, 2005). Thus the validity of any new techniques has to be carefully documented. Since forensic samples can be obtained under a number of varying conditions, it was also necessary to determine the limitations of the method as well as to determine the reliability of the method in producing accurate results. To achieve this, an internal validation study was conducted on the six Non-CODIS miniSTR system in accordance to the SWGDAM guidelines (www.cstl.nist.gov/strbase/validation/SWGDAM).

4.2 Materials and method

4.2.1 DNA extraction

Samples were collected as whole blood or buccal swab samples. DNA was extracted as previously described by Lahiri and Nurnberger (1991) (refer to Appendix I).

4.2.2 DNA quantification

DNA samples were quantified using a Nanodrop ND 1000 UV-Vis spectrophotometer.

4.2.3 PCR amplification

PCR amplification reactions were performed as previously described (Chapter 2; Section 2.2.1)

4.2.4 Sample genotyping

Amplified DNA fragments were analyzed on an ABI 377 Genetic Analyzer (Applied Biosystems). Allele designations were assigned according to recommendations of the DNA Commission of the ISFG (Bar et al., 1997) with the aid of allelic ladders. Allelic ladders were designed by combining products of varying length representative of the common alleles present.

4.3 Results and Discussion

4.3.1 Reproducibility

To test genotyping consistency random human DNA samples were selected for each miniplex set and typed on three separate occasions with a minimum of one week lapse between each experiment. Consistent genotype profiles were obtained for all samples tested within each miniplex set (Table 4.1 and 4.2).

Table 4.1: Genotype profiles obtained for human DNA samples analysed with miniplex 1 loci

Sample ID	Miniplex 1 loci								
	D1S1627			D5S2500			D8S1115		
	Run 1	Run 2	Run 3	Run 1	Run 2	Run 3	Run 1	Run 2	Run 3
9947A	13:14	13:14	13:14	14:23	14:23	14:23	9:18	9:18	9:18
C106	11:14	11:14	11:14	17:18	17:18	17:18	9:14	9:14	9:14
C26	12:14	12:14	12:14	17:18	17:18	17:18	14:16	14:16	14:16
As17	13:14	13:14	13:14	14:18	14:18	14:18	9:16	9:16	9:16
A146	10:13	10:13	10:13	18:23	18:23	18:23	15:16	15:16	15:16
As12	10:13	10:13	10:13	14:18	14:18	14:18	9:16	9:16	9:16

Table 4.2: Genotype profiles obtained for human DNA samples analysed with miniplex 2 loci

Sample ID	Miniplex 2 loci								
	D3S4529			D6S1017			D9S2157		
	Run 1	Run 2	Run 3	Run 1	Run 2	Run 3	Run 1	Run 2	Run 3
9947A	15:15	15:15	15:15	9:10	9:10	9:10	7:13	7:13	7:13
A130	15:17	15:17	15:17	10:13	10:13	10:13	11:12	11:12	11:12
A149	13:15	13:15	13:15	8:10	8:10	8:10	12:15	12:15	12:15
As2	14:16	14:16	14:16	8:10	8:10	8:10	7:13	7:13	7:13
As3	13:14	13:14	13:14	10:12	10:12	10:12	13:14	13:14	13:14
C9	13:14	13:14	13:14	8:9	8:9	8:9	7:10	7:10	7:10

4.3.2 Precision and accuracy

Central to STR genotyping is the ability of the method to measure the size of the amplified product (Mulero *et al*, 2006; Krenke *et al*, 2005). "Sizing precision allows for determining accurate and reliable genotypes" (Mulero *et al*, 2006). In this study precision refers to the reproducibility of the size measured for the amplified product and accuracy refers to the ability of the method to obtain a correct size and genotype (Daniels *et al*, 2004; Krenke *et al*, 2005). The standard DNA sample 9947A was selected for this study.

DNA sample was amplified and typed on three different occasions with a minimum of one week lapse between each typing session. Precision was calculated for each locus and expressed as a standard deviation. Accuracy was expressed as the variation in size (bp) between the allele(s) of the sample compared to those same allele(s) within the allelic ladder of the same gel run. Allele sizes were determined using an internal lane standard, GeneScan LIZ 500.

The precision study indicated that across all loci the standard deviation did not exceed 0.141 bases (Table 4.3). In this study the highest standard deviation of 0.141 bases was shown by the locus D8S1115 (Table 4.3). This is well below the accepted norm of 0.2 bases standard deviation (www.cstl.nist.gov/biotech/strbase/training.htm). Capillary based genetic analysers and DNA sequencers generally report standard deviations below 0.15 bases (www.cstl.nist.gov/biotech/strbase/training.htm). In this study the gel based ABI PRISM 377 DNA sequencer (Applied Biosystems) was used. This method is deemed to be less accurate than the capillary based systems (www.cstl.nist.gov/biotech/strbase/training.htm). Thus this study proved that the miniplex systems investigated could generate reproducible results with high precision since standard deviation across all markers fell well below the 0.2 base limit.

In the accuracy test alleles across all loci fell within ±0.5bp of the same allele in the allelic ladder (Table 4.4). ABI Prism DNA Genotyper Analysis Software (version 3.7) allows for ±0.5bp variation between the unknown fragment and the fragment in the allelic ladder. In each set of analysis conducted, the standard sample, 9947A, fell well within the ±0.5bp variation allowed. Furthermore, the standard sample, 9947A, did not show the presence of any variant alleles.

Table 4.3: Precision study showing average allele size and standard deviation of the control sample 9947A for six Non-CODIS miniSTR loci

Locus	Mean size allele 1 (bp)	Standard deviation allele 1 (bp)	Mean size allele 2 (bp)	Standard deviation allele 2 (bp)
D1S1627	93.027	0.086	96.113	0.045
D3S4529	120.957	0.098	120.957	0.098
D5S2500	86.687	0.133	123.083	0.062
D6S1017	91.857	0.082	95.733	0.058
D8S1115	65.560	0.141	93.310	0.037
D9S2157	74.120	0.100	92.270	0.135

Table 4.4: Accuracy study presenting the variation between allele sizes of the control sample 9947A in comparison to the same allele sizes in the allelic ladders for each of the Non-CODIS miniSTR loci

Locus	Run number	Allele size(s) (bp) of sample 9947A	Allele size(s) (bp) within the allelic ladder	Variation of size (bp)*
D1S1627	1	93.13; 96.17	93.21; 96.27	0.08; 0.1
	2	93.03; 96.06	93.08; 96.23	0.05; 0.17
	3	92.92; 96.11	92.65; 96.06	-0.27; -0.05
D3S4529	1	120.83	120.91	0.08
	2	120.97	120.94	-0.03
	3	121.07	120.87	-0.2
D5S2500	1	86.76; 123.03	86.81; 123.14	0.05; 0.11
	2	86.80; 123.17	86.96; 123.02	0.16; -0.15
	3	86.50; 123.05	86.50; 122.83	0.0; -0.22
D6S1017	1	91.75; 95.72	91.87; 95.78	0.12; 0.06
	2	91.95; 95.67	91.98; 95.95	0.03; 0.28
	3	91.87; 95.81	91.87; 95.95	0.0; 0.14
D8S1115	1	65.66; 93.36	65.74; 93.37	0.08; 0.01
	2	65.66; 93.30	65.30; 93.35	-0.36; 0.05
	3	65.36; 93.27	65.35; 93.37	-0.01; 0.1
D9S2157	1	74.07; 92.31	74.17; 92.37	0.1; 0.06
	2	74.26; 92.43	74.14; 92.63	-0.12; 0.2
	3	74.03; 92.28	74.02; 92.29	-0.01; 0.01

* Variation of size (bp) = Allelic ladder size – Allele size of sample 9947A

4.3.3 Sensitivity and stochastic studies

DNA obtained from crime scenes are generally compromised in one form or another. In some instances the DNA recovered can be both degraded and the amount of template can also be low. It is envisioned that by shortening the size of the amplicon, amplification efficiency of degraded and low copy DNA would be improved. The Non-CODIS miniplex system investigated in this study was designed so as to aid in the amplification of degraded and low concentrations of DNA.

The sensitivity of miniplex 1 (n=9) and miniplex 2 (n=8) was investigated. Decreasing amounts of DNA (500pg, 200pg, 100pg, 50pg, 25pg, and 12.5pg) was added to a standard 10µl PCR reaction. GeneScan analysis threshold was set at a minimum of 150RFU.

The lower limit of template DNA necessary for a complete profile was 25pg and 50pg in miniplex 1 and miniplex 2 respectively (Figure 4.1 and 4.2). At lower concentrations miniplex sets displayed preferential amplification of the smaller allele and/or allele drop out. At concentrations of 500pg allele drop in and pull up was observed in some samples. Good signal intensities with average RFU values ranging between 800-1800 for miniplex 1 loci at a template concentration of 25pg was observed (Figure 4.3). Miniplex 2 loci displayed average RFU values ranging between 800-2500 at a template concentration of 50pg (Figure 4.4).

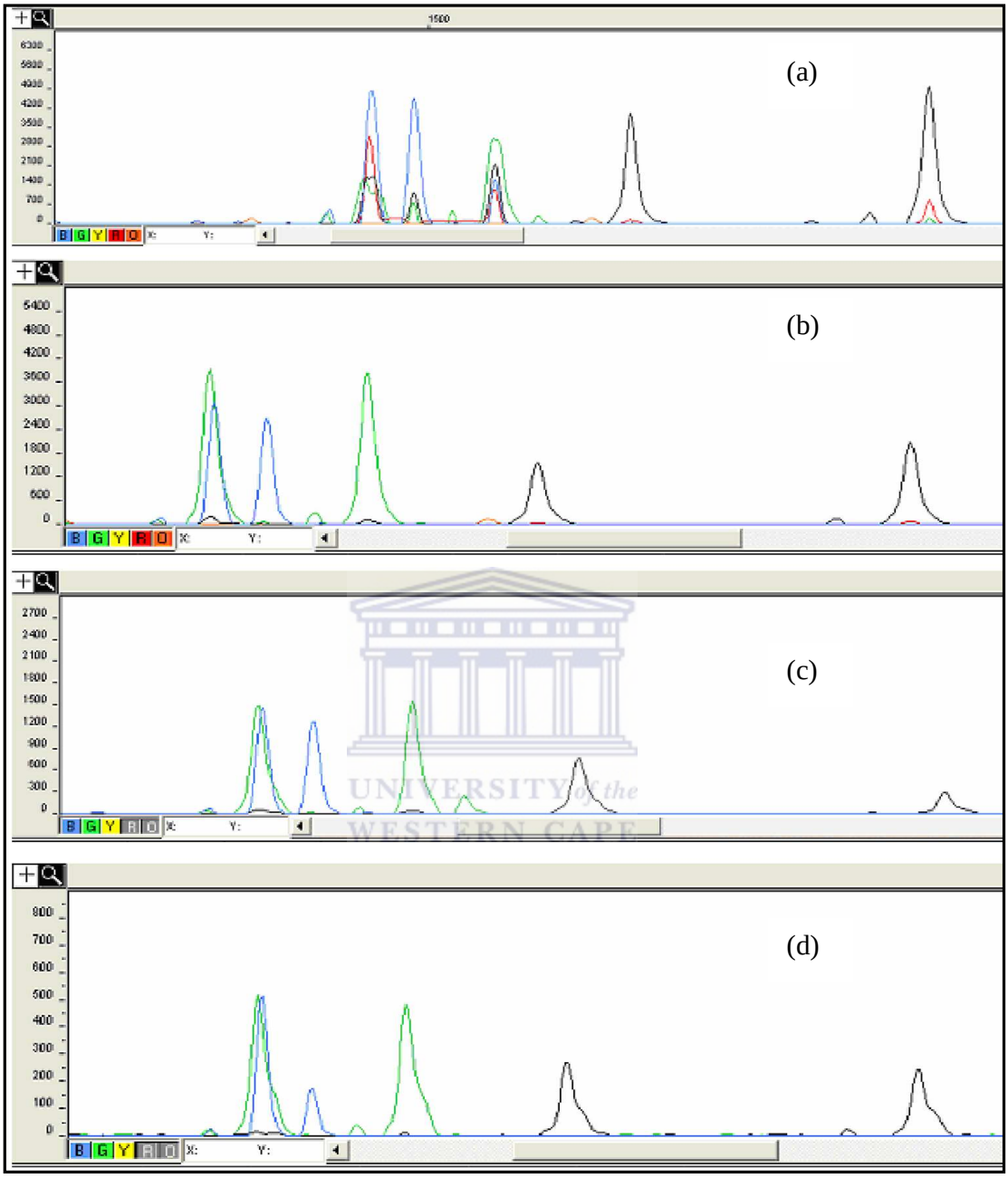


Figure 4.1: Electropherograms illustrating DNA template concentration titration for miniplex 1. (a) Amplification of 500pg of DNA, allele drop in and pull up can be observed. Loci are performing optimally when amplified with DNA concentrations of 200pg and 50pg (panels b and c respectively). (d) Locus D5S2500 (black) is not performing optimally when amplified with 12.5pg of DNA. Sensitivity limit for miniplex 1 was set at 25pg.

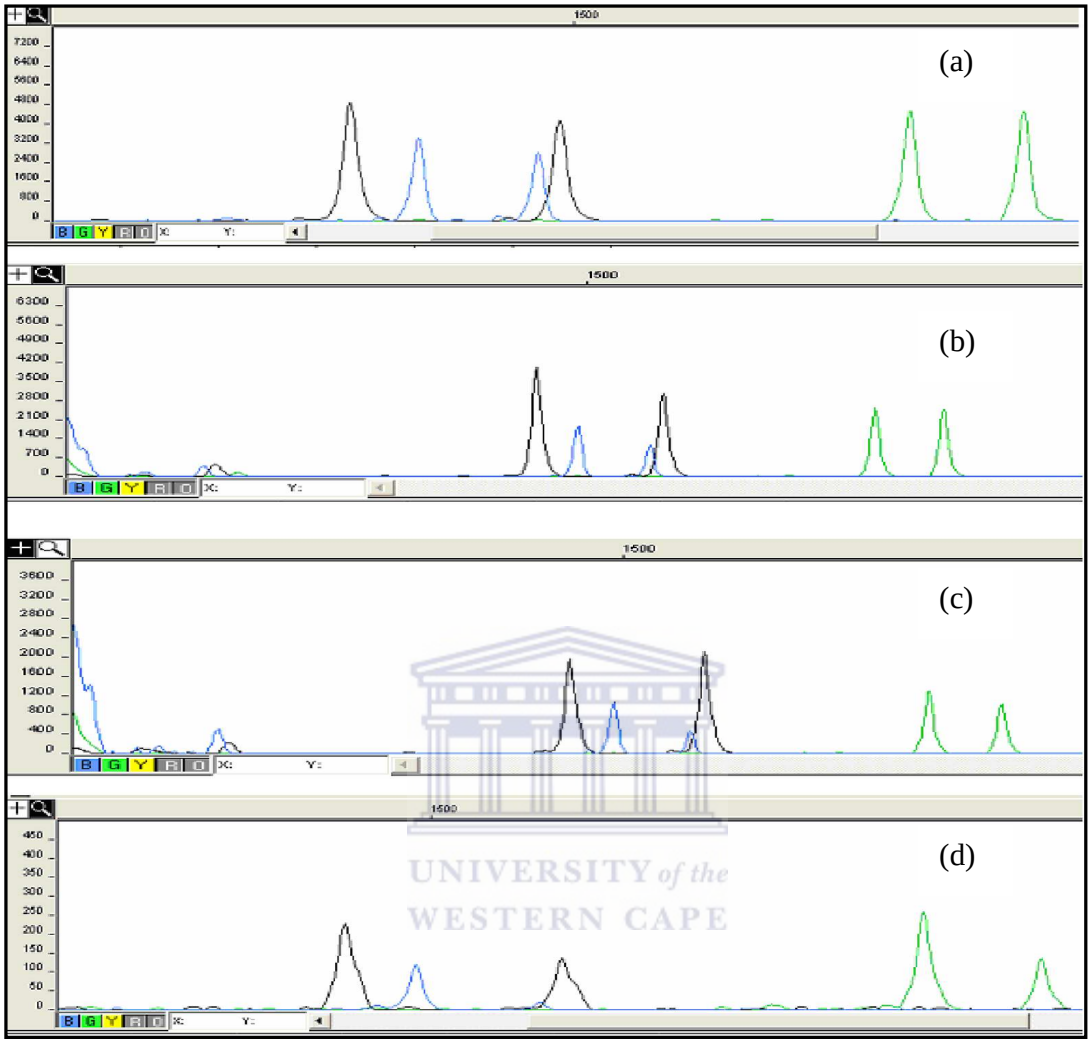


Figure 4.2: Electropherograms illustrating DNA template concentration titration for miniplex 2. Loci are performing optimally when amplified with DNA concentrations of 500pg, 200pg and 50pg (panels a, b and c respectively). (d) At a concentration of 12.5pg, allele drop out was observed for the larger allele of the locus D9S2157 (blue). Sensitivity limit for miniplex 2 was set at 50pg.

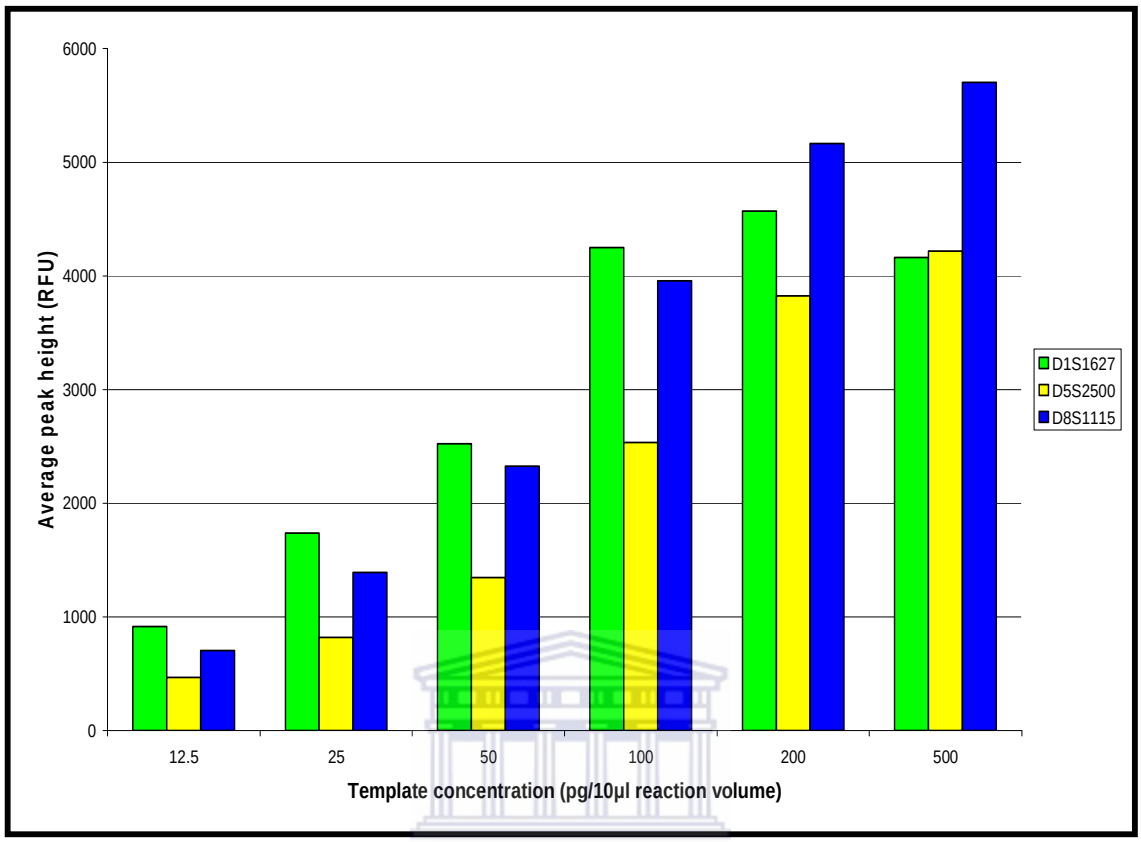


Figure 4.3: Sensitivity and stochastic study for miniplex 1. The average fluorescence signal intensity for D1S1627, D5S2500 and D8S1115 was plotted as a function of template concentration. Good signal intensities ranging between 800-1800RFU were obtained at the minimal template concentration of 25pg. (n=9)

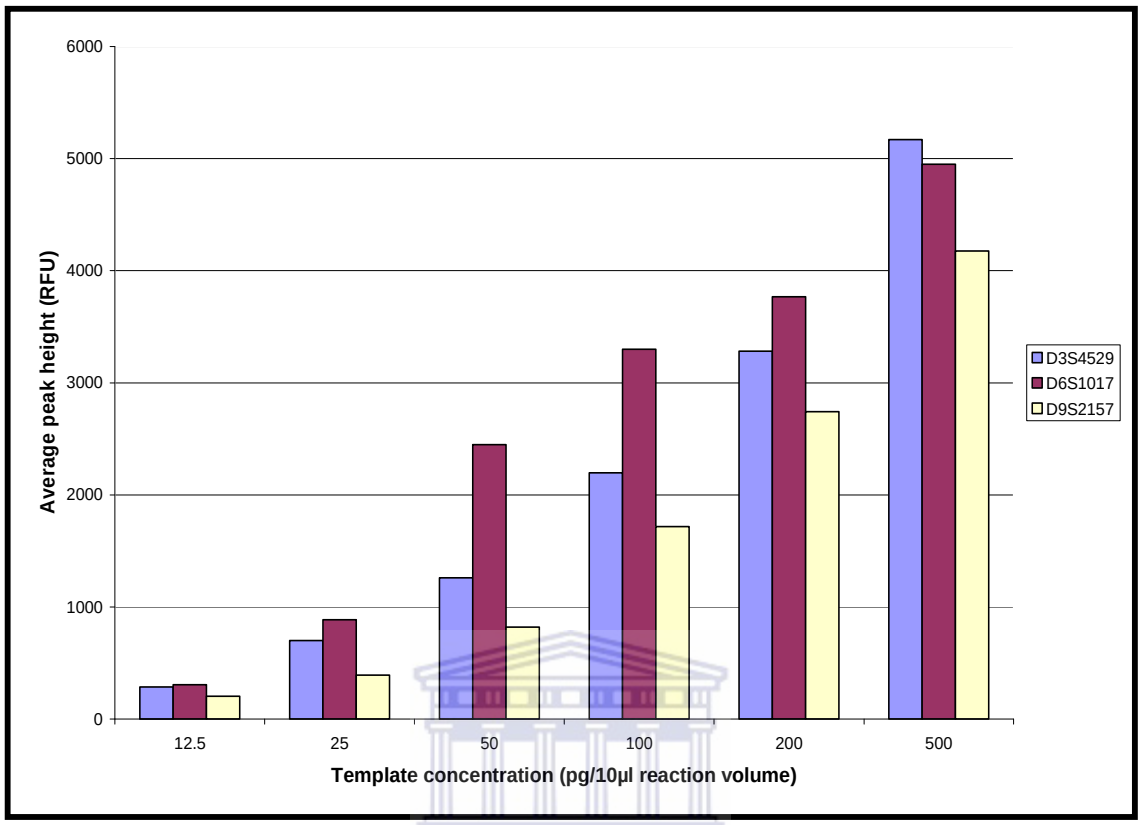


Figure 4.4: Sensitivity and stochastic study for miniplex 2. The average fluorescence signal intensity for D3S4529, D6S1017 and D9S2157 was plotted as a function of template concentration. Good signal intensities ranging between 800-2500RFU were obtained at the minimal template concentration of 50pg. (n=8)

4.3.4 Peak balance studies

When the amount of template DNA added to a PCR reaction is low in concentration, the amplification of heterozygous alleles can be imbalanced due to stochastic effects. Preferential amplification may be observed due to the unequal sampling of the heterozygous sample during PCR (Chung, 2004). During genotyping, balanced allele peaks is of vital importance, therefore the peak balance ratio of heterozygous samples were determined at six DNA concentrations (500pg, 200pg, 100pg, 50pg, 25pg, and 12.5pg) for both miniplex sets in a final reaction volume of 10µl. The peak balance ratio was calculated by dividing the peak height of the smaller peak by the peak height of the larger peak. Samples which displayed allele dropout (i.e. displayed a RFU <150) of one allele was assigned a zero peak balance ratio.

Miniplex 1 displayed good intraloci and interloci balance (ratio ≥0.6) across all concentration ranges, except for locus D5S2500 (Figure 4.5). Locus D5S2500 displayed interloci and intraloci balance of <0.6 at a concentration of 12.5pg (Figure 4.5). Miniplex 2 displayed good intraloci and interloci balance (ratio ≥0.6) at concentrations ≥50pg (Figure 4.6). At 25pg interloci and intraloci balance of <0.6 was obtained for the locus D3S4529. D6S1017 and D9S2157 showed interloci and intraloci balance of <0.6 at a concentration of 12.5pg (Figure 4.6). However, at the determined levels of sensitivity, i.e. 25pg for miniplex 1 and 50pg for miniplex 2, all loci displayed a peak balance ratio >0.6.

The peak balance ratio between loci is also affected by the concentration of primers since each primer will have a varied primer binding efficiency depending on the DNA sequence of each locus (Chung, 2004). Optimal signal intensity and balanced peak

heights could be obtained for each miniplex by making further adjustments to individual primer concentrations.

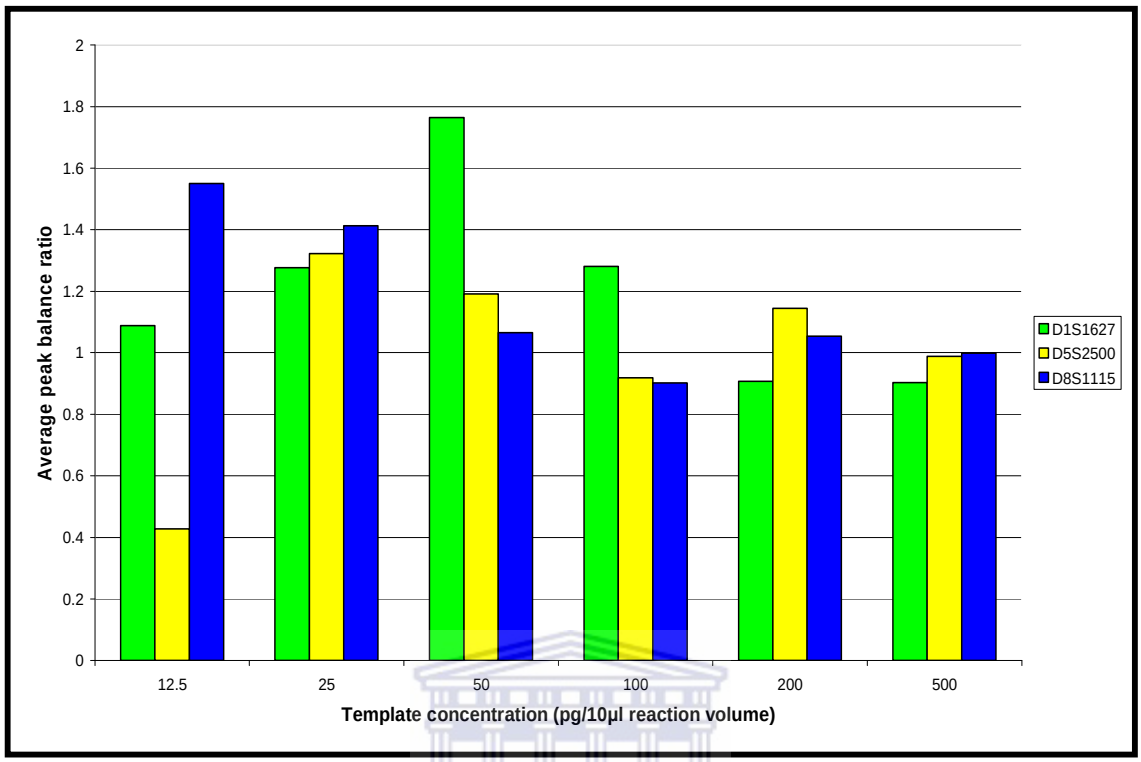


Figure 4.5: Peak balance ratio for miniplex 1. The average peak balance ratio for D1S1627, D5S2500 and D8S1115 was plotted as a function of template concentration. The peak balance ratio remained above the 0.6 criterion for all loci at the sensitivity limit of 25pg. (n=9)

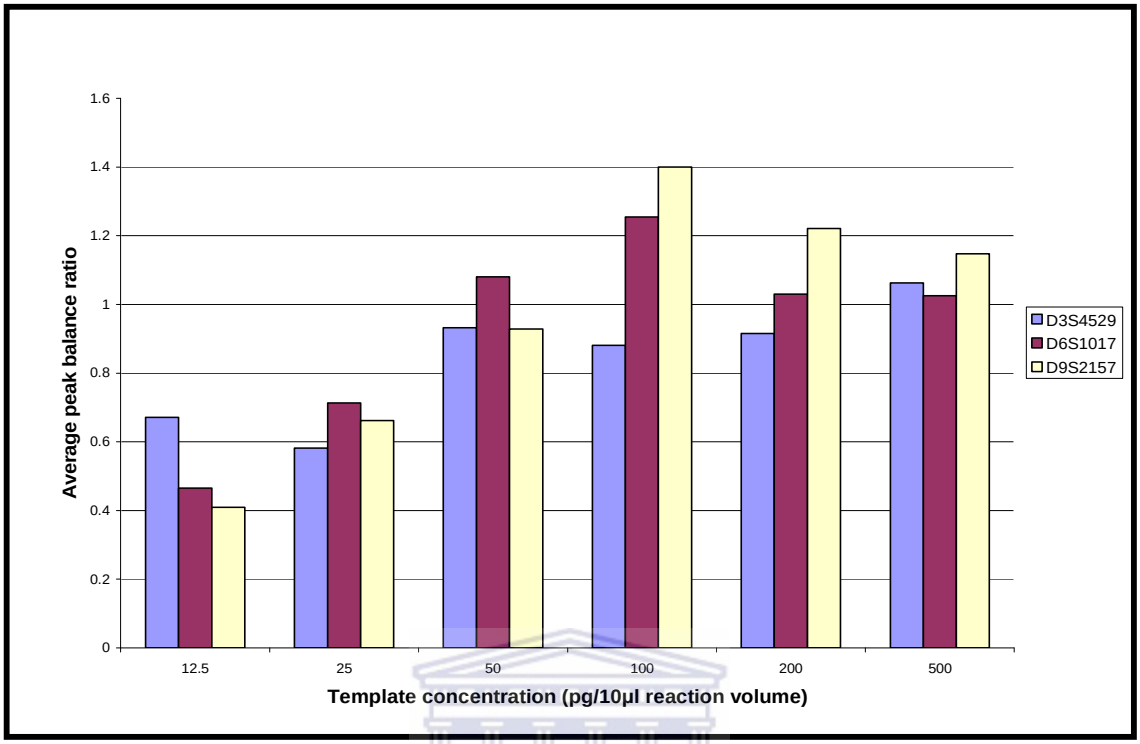


Figure 4.6: Peak balance ratio for miniplex 2. The average peak balance ratio for D3S4529, D6S1017 and D9S2157 was plotted as a function of template concentration. The peak balance ratio remained above the 0.6 criterion for all loci at the sensitivity limit of 50pg. (n=8)

4.3.5 Mixture studies

Casework samples may in some instances contain DNA from more than one contributor. Thus it is important to test the ability of a miniplex system to distinguish between the major and minor contributors. Deciphering the genotypes of individuals in mixed samples can be very difficult especially in instances where the contributors share common alleles. In cases like these, analysis of peak ratios becomes an important tool in differentiating between the DNA originating from the minor and major contributors (Chung, 2004).

For this mixture study, two DNA samples were selected for each miniplex. The samples selected shared no common alleles for the loci tested. DNA samples were prepared at a concentration of 100pg (for miniplex 1) and 200pg (for miniplex 2) and then mixed at ratios of: 1:19, 1:9, 1:4 and 1:1. 1µl of mixed DNA, (for each of the above mentioned ratios), was cycled under the standard PCR conditions in a final reaction volume of 10µl.

The results from this study showed that across all loci (i.e. D1S1627, D5S2500 and D8S1115) and all dilutions (i.e. 1:19, 1:9, 1:4 and 1:1) in miniplex 1 both the minor and major components were detectable within the mixed samples (Table 4. 5).

At a ratio of 1:19 a complete profile was only observed for the major component in miniplex 2 (Table 4.6). The minor component displayed allele drop out for all loci (i.e. D3S4529, D6S1017 and D9S2157) at this ratio (Table 4.6). In miniplex 2 the minor component was detectable for the all loci from a ratio of 1:4 (Table 4.6).

Across all loci in both miniplex sets, the ratio of the minor component alleles to the major component alleles quantitatively reflected the amount of DNA input. The peak height ratio of the minor component increases as the contribution of minor component increases

in the mixed DNA sample (Figure 4.7 and 4.8). These mixture studies displayed that it is possible to discriminate between individuals within a mixed sample for miniplex 1 and miniplex 2. In miniplex 2 this is however dependant on the ratio of genomic DNA within the mix since two complete profiles could only be detected with certainty for all loci from a ratio of 1:4 and less.

Table 4.5: Genotype profiles obtained from the amplification of mixed DNA samples with miniplex 1 loci

Ratio of minor to major DNA component	Genotype profiles obtained for minor DNA component			Genotype profiles obtained for major DNA component		
	D1S1627	D5S2500	D8S1115	D1S1627	D5S2500	D8S1115
Neat	10:12	17:17	9:9	13:13	14:14	16:16
1:1	10:12	17:17	9:9	13:13	14:14	16:16
1:4	10:12	17:17	9:9	13:13	14:14	16:16
1:9	10:12	17:17	9:9	13:13	14:14	16:16
1:19	10:12	17:17	9:9	13:13	14:14	16:16

0 indicative of an allele displaying an RFU value less than the threshold limit of 150.

Table 4.6: Genotype profiles obtained from the amplification of mixed DNA samples with miniplex 2 loci

Ratio of minor to major DNA component	Genotype profiles obtained for minor DNA component			Genotype profiles obtained for major DNA component		
	D3S4529	D6S1017	D9S2157	D3S4529	D6S1017	D9S2157
Neat	15:17	10:13	11:12	13:14	8:9	7:10
1:1	15:17	10:13	11:12	13:14	8:9	7:10
1:4	15:17	10:13	11:12	13:14	8:9	7:10
1:9	0:17	0:0	0:0	13:14	8:9	7:10
1:19	15:0	0:0	0:0	13:14	8:9	7:10

0 indicative of an allele displaying an RFU value less than the threshold limit of 150.

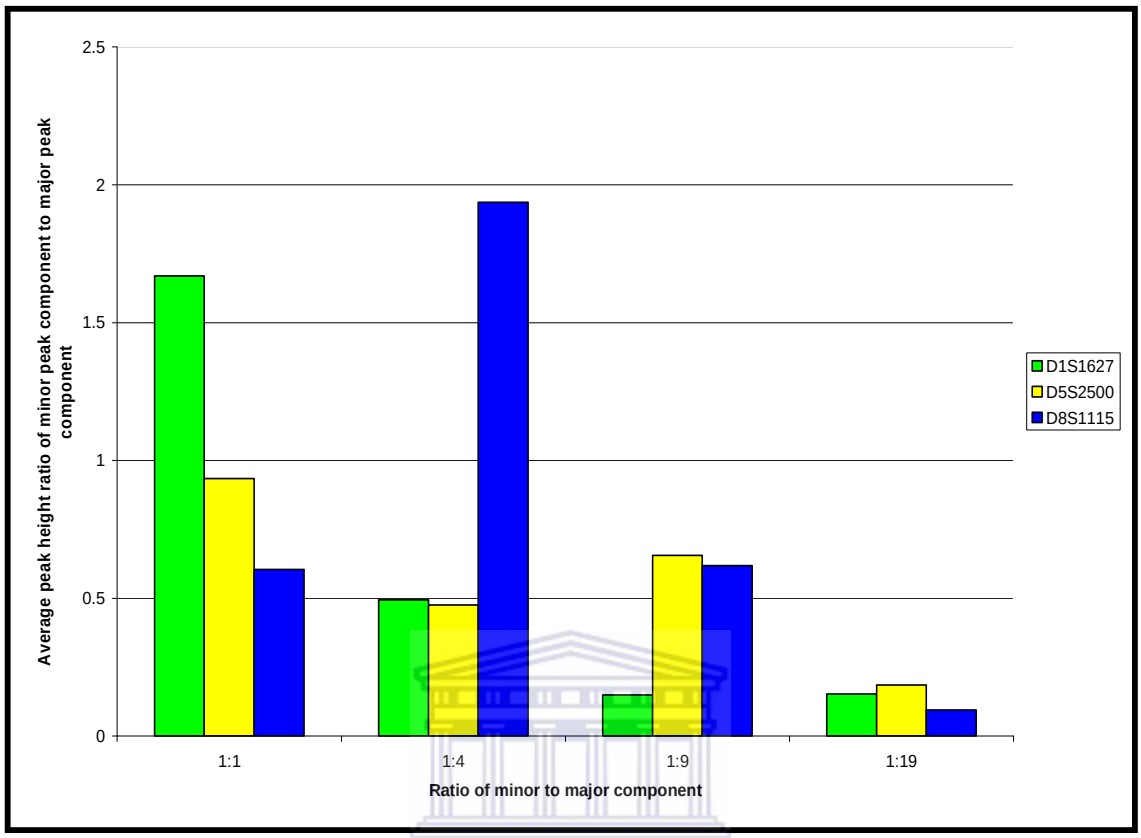


Figure 4.7: Mixture studies conducted for miniplex 1. Two DNA samples were mixed at different ratios whilst keeping the starting concentration of each component at 100pg prior to making DNA mixes. All miniplex loci were detectable for both minor and major components from a ratio of 1:19.

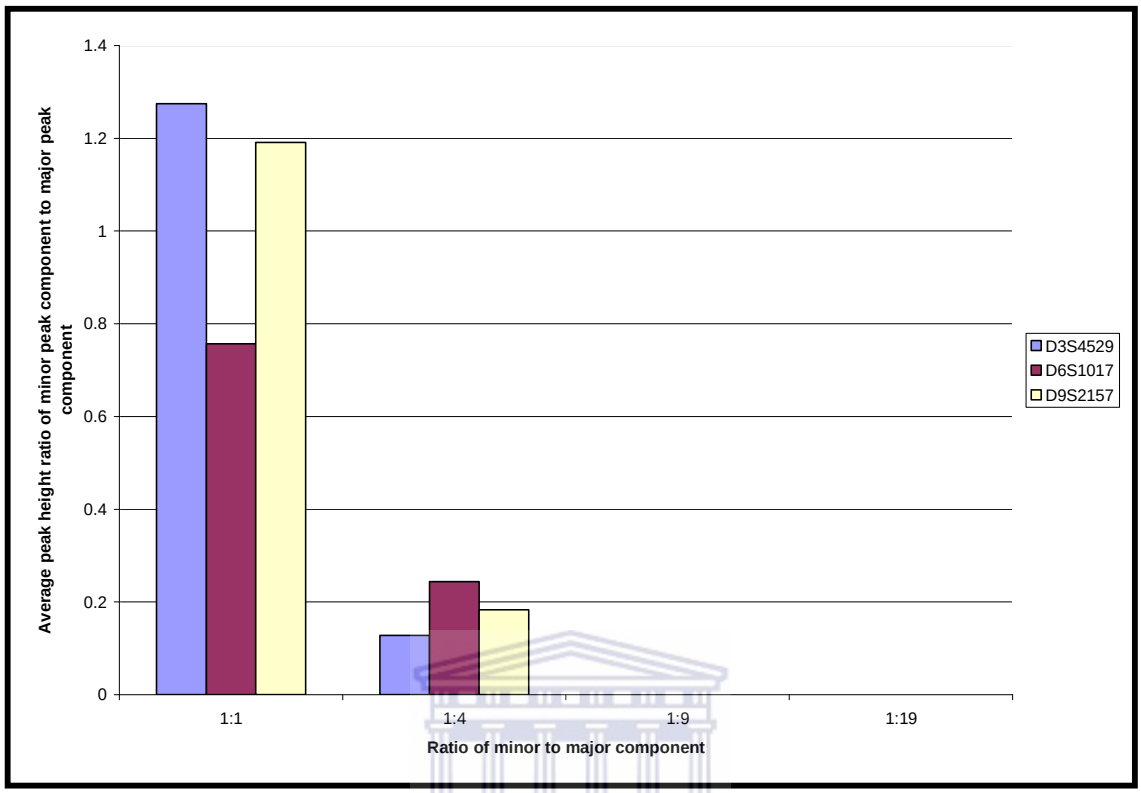


Figure 4.8: Mixture studies conducted for miniplex 2. Two DNA samples were mixed at different ratios whilst keeping the starting concentration of each component at 200pg prior to making DNA mixes. The minor component displayed allele drop out for all loci at ratios of 1:9 and 1:19. Complete profiles were observed for both minor and major components from a ratio of 1:4 and less.

4.3.6 Species specificity

The aim of the study was to determine the ability of the system to amplify non-human DNA in forensic samples. DNA from: dog, cat, baboon, chimpanzee, vervet monkey and gorilla were amplified with the two miniplex sets. Amplifications were performed using 200pg of non-human DNA cycled under the standard PCR conditions in a final reaction volume of 10µl.

Dog, chimpanzee and gorilla DNA displayed amplification products for the locus D1S1627 (Figure 4.9). Cat and gorilla DNA displayed amplification products for the locus D8S1115 where the gorilla DNA showed a tri-allelic pattern (Figure 4.9). Chimpanzee, baboon and vervet monkey DNA displayed amplification products for the locus D9S2157 (Figure 4.10). Baboon DNA also displayed amplification products for the locus D6S1017 (Figure 4.10). However, none of the non-human DNA samples tested produced a complete profile using the six Non-CODIS miniSTR genotyping system. Furthermore, baboon and vervet monkey samples did not display any amplification products for any of the loci within miniplex 1 (Figure 4.9). Dog, cat and gorilla samples did not display any amplification products for any of the loci within miniplex 2 (Figure 4.10).

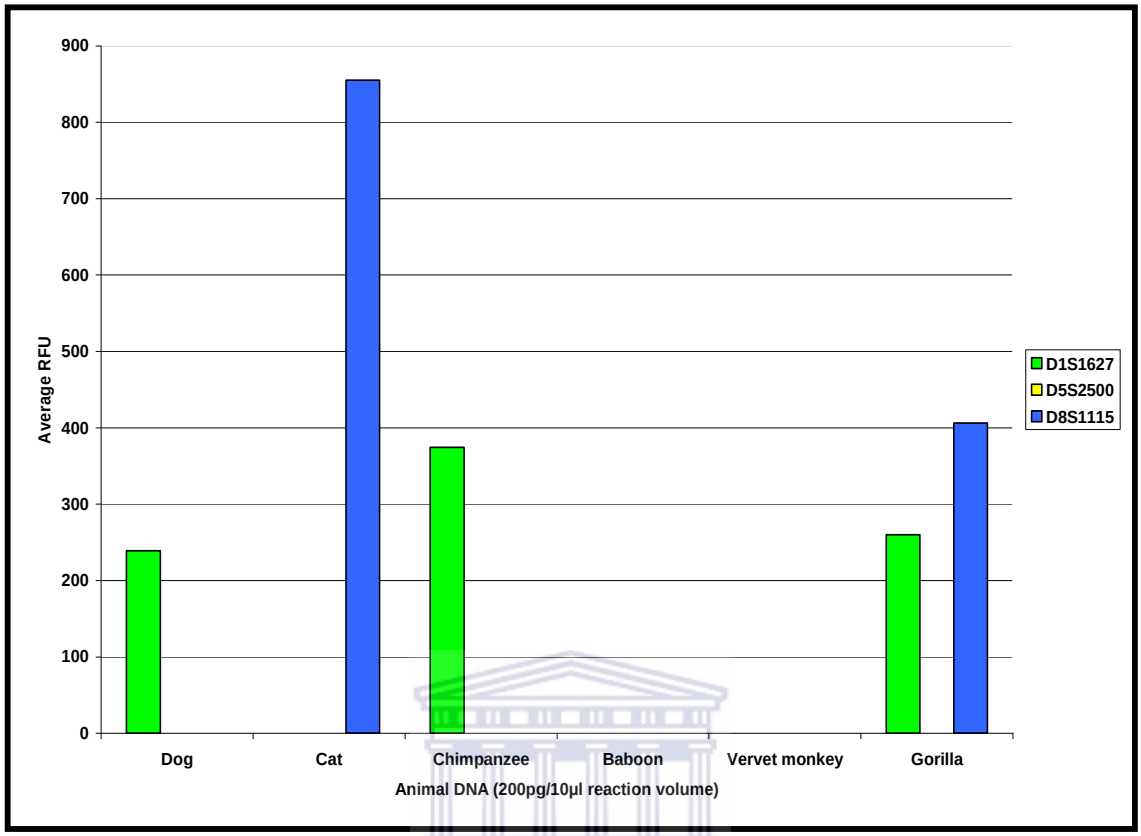


Figure 4.9: Species specificity for miniplex 1: The average RFU was plotted as a function of template concentration. At 200pg of template DNA, amplification products were displayed by dog, chimpanzee and gorilla samples for locus D1S1627. Locus D8S1115 displayed amplification products for cat and gorilla DNA samples.

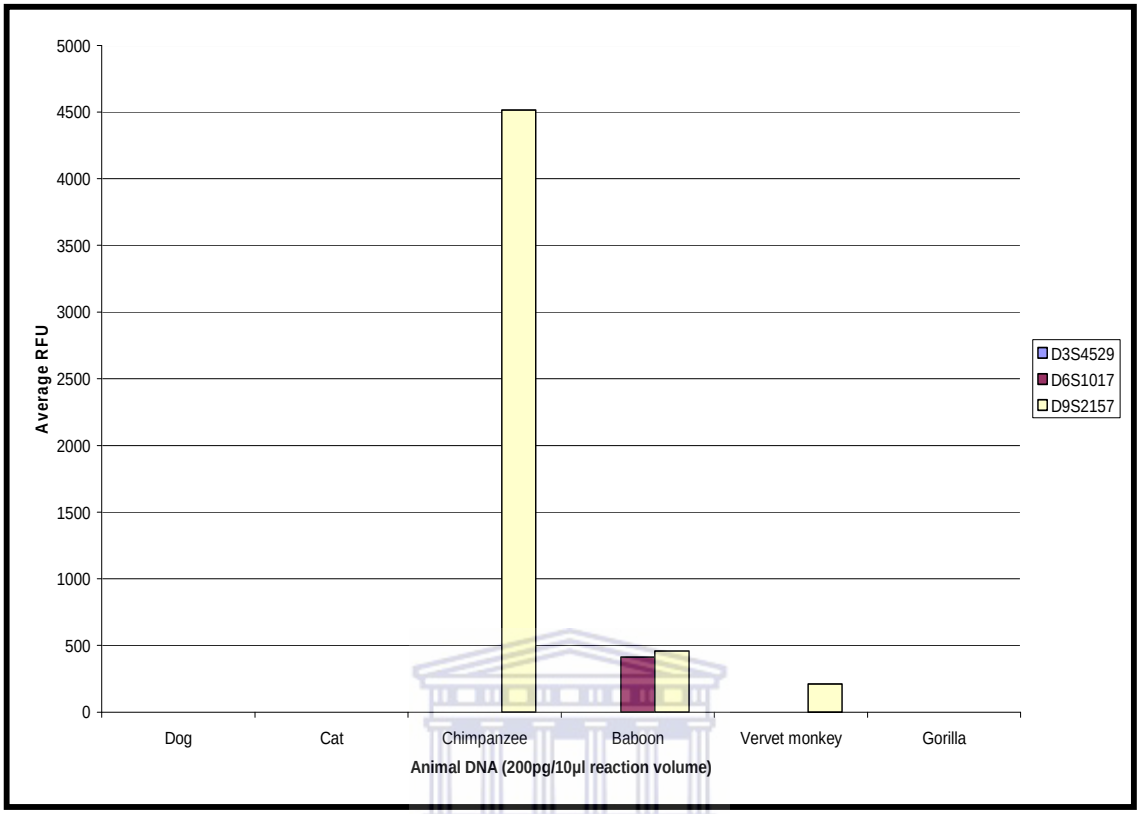


Figure 4.10: Species specificity for miniplex 2: The average RFU was plotted as a function of template concentration. At 200pg of template DNA the baboon sample was the only animal DNA sample tested to display an amplification product for locus D6S1017. Chimpanzee, baboon and vervet monkey samples displayed amplification products for locus D9S2157.

4.4 Summary

The Non-CODIS miniSTR loci investigated has demonstrated the ability to produce consistent, accurate and precise genotype profiles for low concentrations of template DNA. Template DNA at concentration as low as 50pg was successfully amplified with a high sensitivity and good peak balance ratios. When analyzing mixed samples, successful differentiation of minor and major components was easily identifiable in miniplex 1. In miniplex 2 however complete profiles for each contributor was only observed from ratios ≤1:4. Amplification products was produced for both domestic animals and higher primates tested with miniplex 1. In miniplex 2 amplification products were only observed for the primate samples. However, none of the non-human DNA samples tested produced a complete profile using the six Non-CODIS miniSTR genotyping system.

This internal validation study provided laboratory experience and enabled the determination of the reliability and limitations of the genotyping system. A comprehensive population study was therefore conducted on five population groups from South Africa to determine the various forensic parameters for the system.

Chapter 5: Investigation of forensic parameters of the six Non-CODIS miniSTR system in five South African population groups

5.1 Introduction

Genetic information presented in court needs to be substantiated with population data. Since it is not possible to genotype every individual in the world, smaller population data sets are used as a reference (Butler, 2005). Essentially, population studies are conducted to test the possibility of a random match between a suspect and an evidence sample (Butler, 2005). When conducting population studies, allele frequencies are determined for various ethnic/racial sample sets (Butler, 2005). Based on the allele frequencies obtained from these studies different statistical programmes can be used to determine the various forensic parameters. These parameters include Hardy-Weinberg equilibrium *p*-values, match probability (MP), power of discrimination (PD), polymorphism information content (PIC), power of exclusion (PE), typical paternity index (TPI) and population differentiation. In this study, these parameters were investigated in five South African population groups (i.e. Afrikaner, Asian-Indian, Mixed Ancestry, Xhosa and Cape Muslim).

5.2 Materials and methods

5.2.1 DNA extraction

Samples were collected from 105 Caucasian individuals of Afrikaner descent (Ehrenreich *et al*, 2008), 112 Asian-Indian males (Leat *et al*, 2007), 115 individuals of Mixed Ancestry (Ehrenreich *et al*, 2008), 64 Xhosa individuals (Leat *et al*, 2004b) and 106 Cape

Muslim individuals (Cloete *et al*, 2009). Ethical clearance for this study was obtained from the Senate Research Committee of the University of the Western Cape. Samples were collected as whole blood or buccal swab samples. All donors were healthy unrelated males from respective communities living in South Africa. DNA extraction was performed as described by Lahiri and Nurnberger (1991). (Refer to Appendix I for a comprehensive DNA extraction protocol).

5.2.2 DNA quantification

DNA samples were quantified using a Nanodrop ND 1000 UV-Vis spectrophotometer.

5.2.3 PCR amplification

PCR amplification reactions were performed as previously described (Chapter 2; Section 2.2.1).

5.2.4 Sample genotyping

Samples were genotyped as previously described (Chapter 4; Section 4.2.4).

5.2.5 Statistical analysis

Hardy-Weinberg equilibrium and the significance of population differentiation were determined by an exact test using GENEPOP (Version 4.0.7) software package designed by Raymond and Rousset (1995a). The Bonferroni correction was applied to test for deviations from the Hardy-Weinberg equilibrium (Butler, 2005). Forensic parameters: match probability, power of discrimination, polymorphism information content, power of

exclusion and typical paternity index were calculated using PowerStats v1.2 software package (Promega).

5.3 Results and Discussion

Allele repeats numbers varied from one locus to another. Repeats 10–17, 11–19, 8–24, 7–18, 7–21 and 7–17 were shown for the loci D1S1627, D3S4529, D5S2500, D6S1017, D8S1115 and D9S2157 respectively (Tables 5.1–5.6). Allele frequency distribution was similar across all populations for individual loci (Figures 5.1–5.6). However, rare alleles were observed. For example allele 17 for the locus D1S1627 (Figure 5.1) was only observed in the Xhosa population. Allele 19 for locus D3S4529 (Figure 5.2) and alleles 16 for locus D5S2500 (Figure 5.3) were only observed in the Asian-Indian population. Alleles 13 and 22 were only observed in the Cape Muslim and Mixed Ancestry populations respectively for locus D5S2500 (Table 5.3 and Figure 5.3). Locus D6S1017 displayed two rare alleles, alleles 14 and 18 and these were observed in the Cape Muslim and Xhosa populations respectively (Figure 5.4). The Mixed Ancestry population was the only population to display alleles 12 and 20 for locus D8S1115 (Figure 5.5). Furthermore, they were also the only population to display an off-ladder allele, allele 9.1 for the locus D9S2157. This allele has previously been reported in American and Singaporean populations (Hill *et al*, 2008; Yong *et al*, 2009). When comparing common alleles observed in the studied South African populations with USA populations as studied by Hill *et al* (2006b) similarities were observed with some loci. For the loci D1S1627, D5S2500 and D8S1115, the same common alleles were observed.

Across the six loci tested for (HWE) the loci: D3S4529, D8S1115 and D9S2157 (*p*-values ranging from 0.0023–0.0323) showed *p*-values below the limit of 0.05 set for HWE (Tables 5.1–5.6). This suggested possible deviation from HWE. However, after the application of the Bonferroni correction (Butler, 2005), only *p*-values below 0.001666666 (0.05/30) would be considered significant. Thus in this study no loci showed deviation from HWE.

The observed heterozygosity values for this study were similar across all five populations (shown in Table 5.7). The observed heterozygosity values ranged from: 0.620 to 0.800, 0.607 to 0.808, 0.696 to 0.843, 0.719 to 0.859 and 0.679 to 0.726 for Afrikaner, Asian-Indian, Mixed Ancestry, Xhosa and Cape Muslim populations respectively. These values hereby imply that there is a minimum of 60.7% chance of random samples matching. Furthermore these values demonstrate good allele diversity for each locus.

The combined match probability (CPM) using these six loci was calculated to be: 2.28×10^{-5}; 2.84×10^{-6}; 3.89×10^{-7}**;** 4.14×10^{-7} and 9.63×10^{-7} for Afrikaner, Asian-Indian, Mixed Ancestry, Xhosa and Muslim populations respectively. Therefore the respective values for the combined power of discrimination (CPD) in these populations were: 0.999997723; 0.999997163; 0.999999961; 0.999999585 and 0.999999037. This implies that when using these loci combined in the studied populations a high degree of discrimination will be ascertained.

Single locus power of exclusion (PE) values across all populations and loci ranged between 0.306–0.713 (Table 5.7). This indicates a low degree of exclusionary power when using the loci individually. However when combining the PE values in the relevant populations, the forensic ability of these loci are markedly increased. CPE values were

calculated to be: 0.99; 0.99; 0.98; 0.95 and 0.99 for Afrikaner, Asian-Indian, Mixed Ancestry, Xhosa and Muslim populations respectively.

Table 5.1: Allele frequency distribution of D1S1627 within five South African population groups

Allele	Afrikaner **(N = 105)**	Asian-Indian **(N = 112)**	Mixed Ancestry **(N = 115)**	Xhosa **(N = 64)**	Muslim **(N = 106)**
10	0.126	0.099	0.113	0.055	0.094
11	0.146	0.036	0.096	0.109	0.075
12	0.015	0.054	0.117	0.133	0.071
13	0.417	0.311	0.348	0.320	0.429
14	0.286	0.486	0.300	0.266	0.316
15	0.005	0.005	0.022	0.109	0.014
16	0.005	0.009	0.004		
17				0.008	
HWE *p*-value	0.9832	0.0678	0.3211	0.8483	0.7999

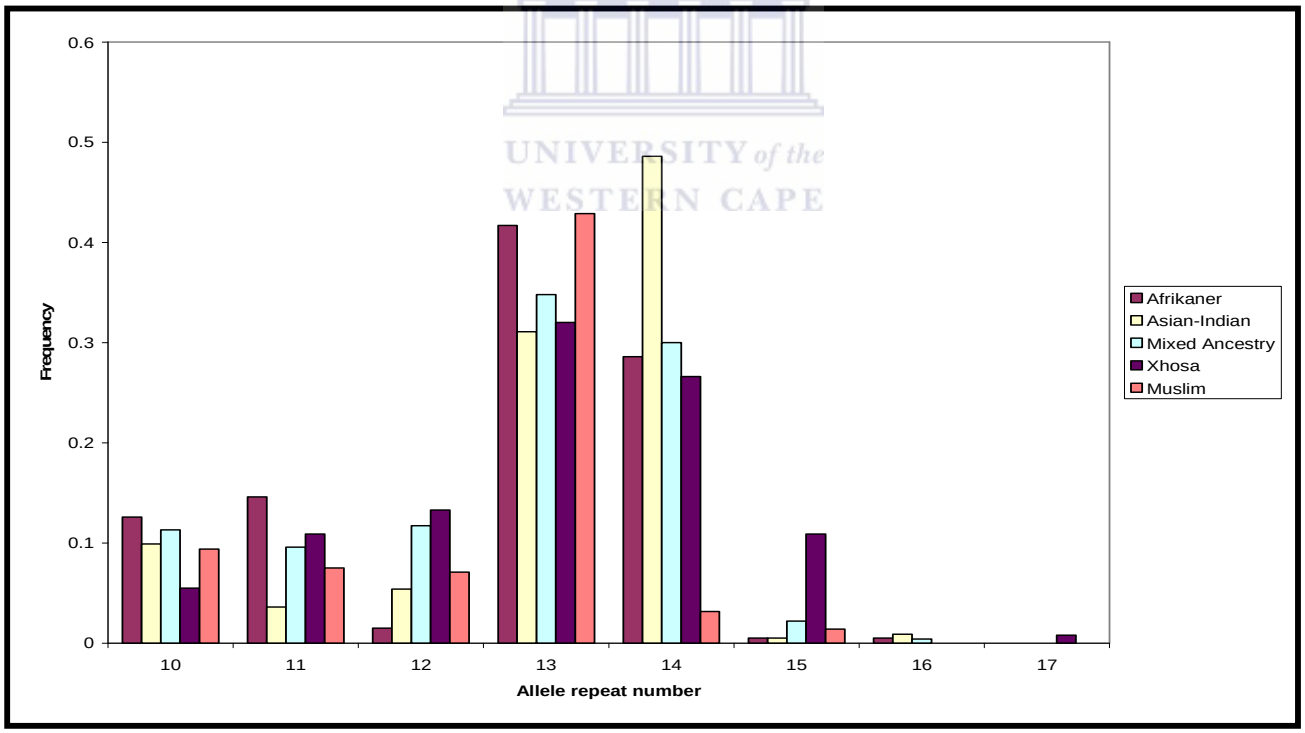


Figure 5.1: Graphical representations of allele repeat number versus frequency for the locus D1S1627 in five South African population groups.

Table 5.2: Allele frequency distribution of D3S4529 within five South African population groups

Allele	Afrikaner **(N = 105)**	Asian-Indian **(N = 112)**	Mixed Ancestry **(N = 115)**	Xhosa **(N = 64)**	Muslim **(N = 106)**
11	0.010		0.009		
12	0.019	0.005	0.030	0.078	0.009
13	0.350	0.306	0.239	0.133	0.250
14	0.189	0.320	0.226	0.305	0.311
15	0.165	0.162	0.213	0.219	0.175
16	0.214	0.167	0.248	0.203	0.198
17	0.053	0.036	0.035	0.063	0.057
19		0.005			
HWE *p*-value	0.3352	0.0137	0.1384	0.7308	0.5741

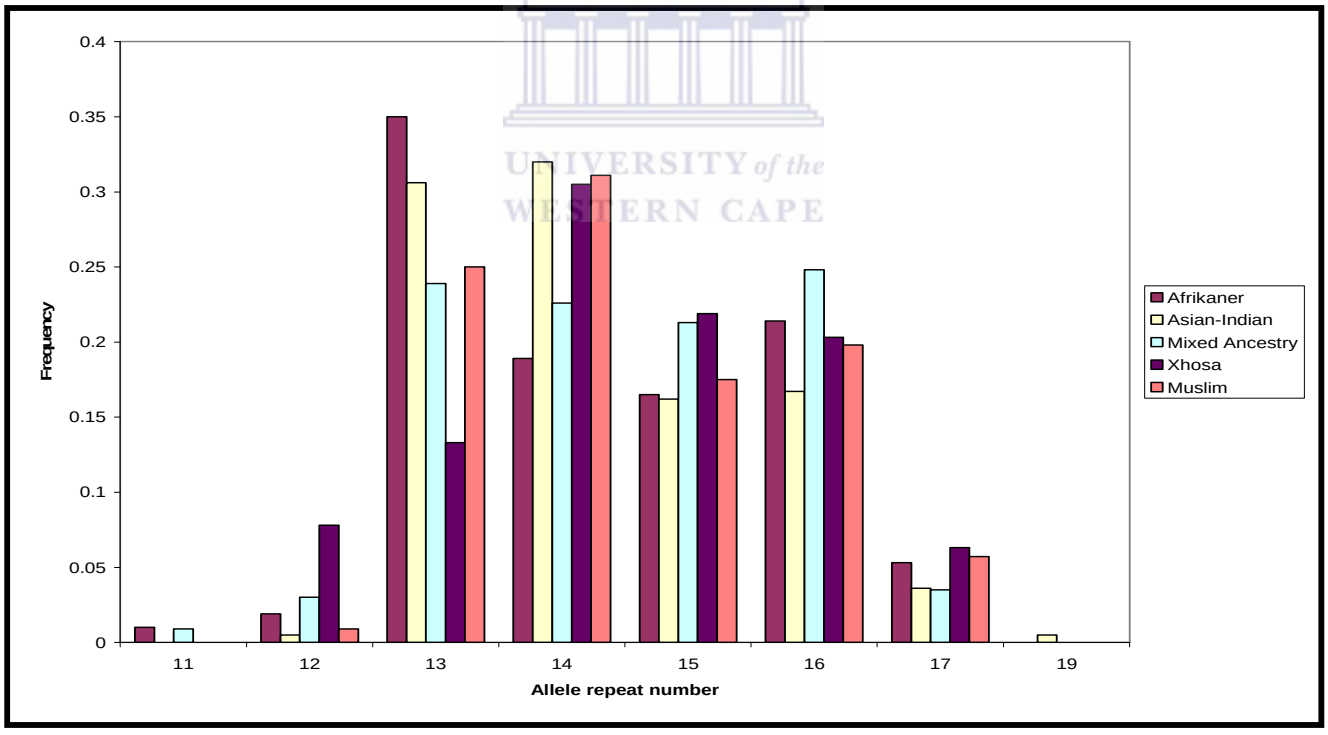


Figure 5.2: Graphical representations of allele repeat number versus frequency for the locus D3S4529 in five South African population groups.

Table 5.3: Allele frequency distribution of D5S2500 within five South African population groups

Allele	Afrikaner **(N = 105)**	Asian-Indian **(N = 112)**	Mixed Ancestry **(N = 115)**	Xhosa **(N = 64)**	Muslim **(N = 106)**
13					0.005
14	0.301	0.319	0.287	0.398	0.302
15			0.004	0.063	
16		0.014			
17	0.296	0.269	0.248	0.018	0.283
18	0.306	0.296	0.230	0.039	0.259
19		0.009	0.030	0.047	0.009
20	0.015	0.046	0.122	0.234	0.075
21					
22			0.004		
23	0.058	0.042	0.061	0.031	0.061
24	0.024	0.005	0.013		0.005
HWE *p*-value	0.1812	0.2378	0.3785	0.8462	0.2642

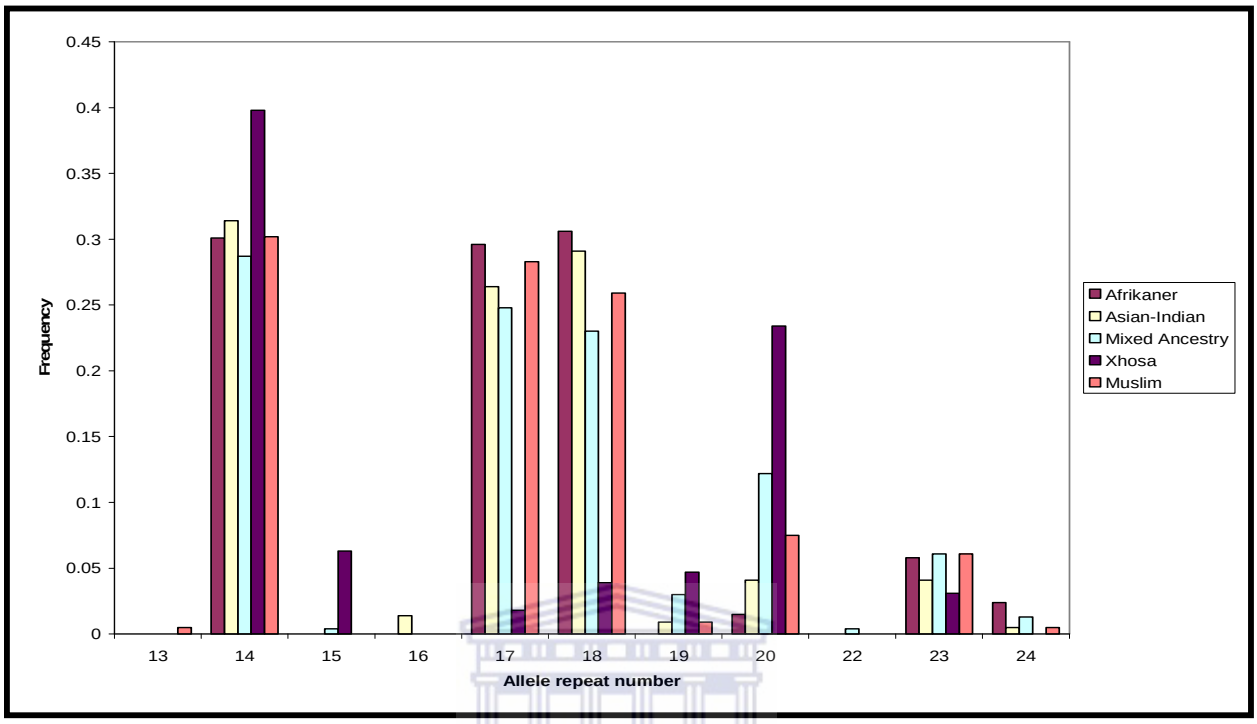


Figure 5.3: Graphical representations of allele repeat number versus frequency for the locus D5S2500 in five South African population groups.

Table 5.4: Allele frequency distribution of D6S1017 within five South African population groups

Allele	Afrikaner **(N = 105)**	Asian-Indian **(N = 112)**	Mixed Ancestry **(N = 115)**	Xhosa **(N = 64)**	Muslim **(N = 106)**
7		0.018	0.087	0.297	0.057
8	0.228	0.311	0.174	0.172	0.165
9	0.010	0.005	0.117	0.203	0.024
10	0.495	0.410	0.370	0.016	0.429
11	0.034	0.054	0.035	0.109	0.038
12	0.204	0.176	0.183	0.023	0.236
13	0.209	0.027	0.035		0.038
14					0.014
18				0.008	
HWE *p*-value	0.8572	0.4569	0.0639	0.7563	0.1617

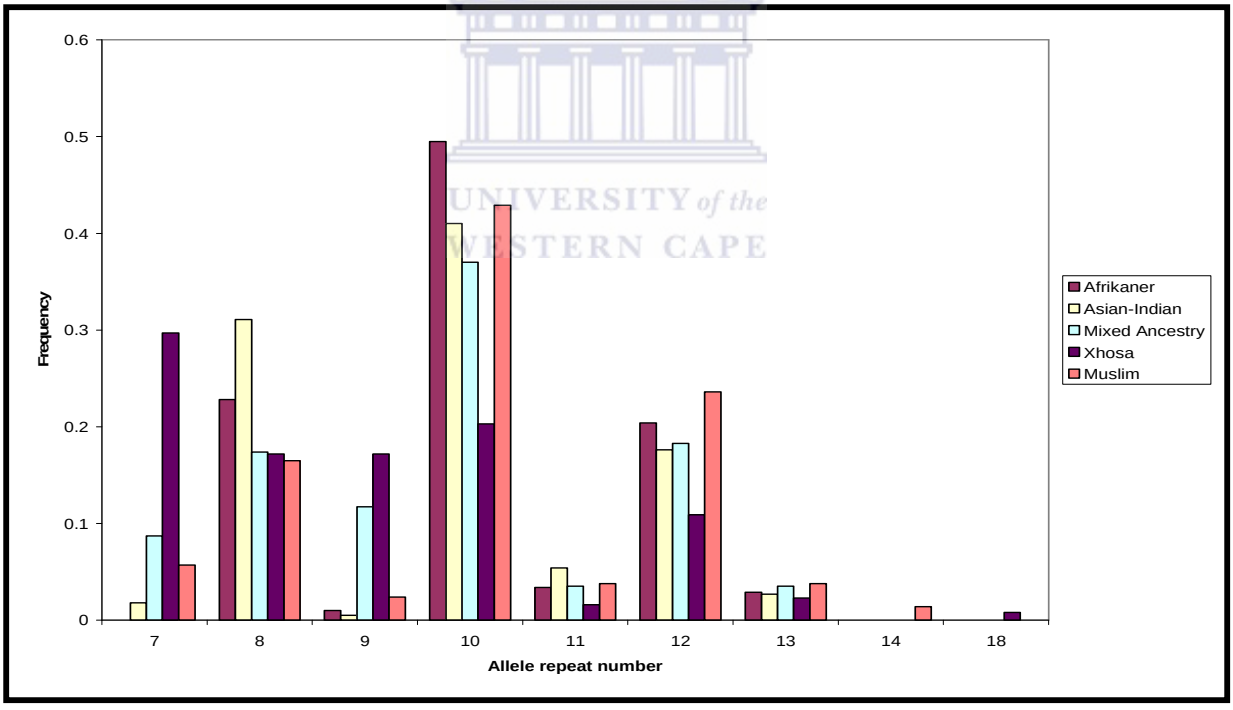


Figure 5.4: Graphical representations of allele repeat number versus frequency for the locus D6S1017 in five South African population groups.

Table 5.5: Allele frequency distribution of D8S1115 within five South African population groups

Allele	Afrikaner **(N = 105)**	Asian-Indian **(N = 112)**	Mixed Ancestry **(N = 115)**	Xhosa **(N = 64)**	Muslim **(N = 106)**
7				0.016	0.009
8					
9	0.209	0.257	0.348	0.492	0.325
10		0.005	0.004	0.008	0.009
11		0.005		0.039	0.005
12			0.009		
13	0.010	0.023	0.030	0.031	
14	0.039	0.050	0.026	0.031	0.024
15	0.063	0.023	0.074	0.055	0.047
16	0.432	0.408	0.322	0.133	0.387
17	0.194	0.161	0.113	0.125	0.156
18	0.024	0.050	0.052	0.063	0.024
19	0.024	0.018	0.017	0.008	0.009
20			0.004		
21	0.005				0.005
HWE *p*-value	0.3838	0.0323	0.0540	0.8015	0.0060

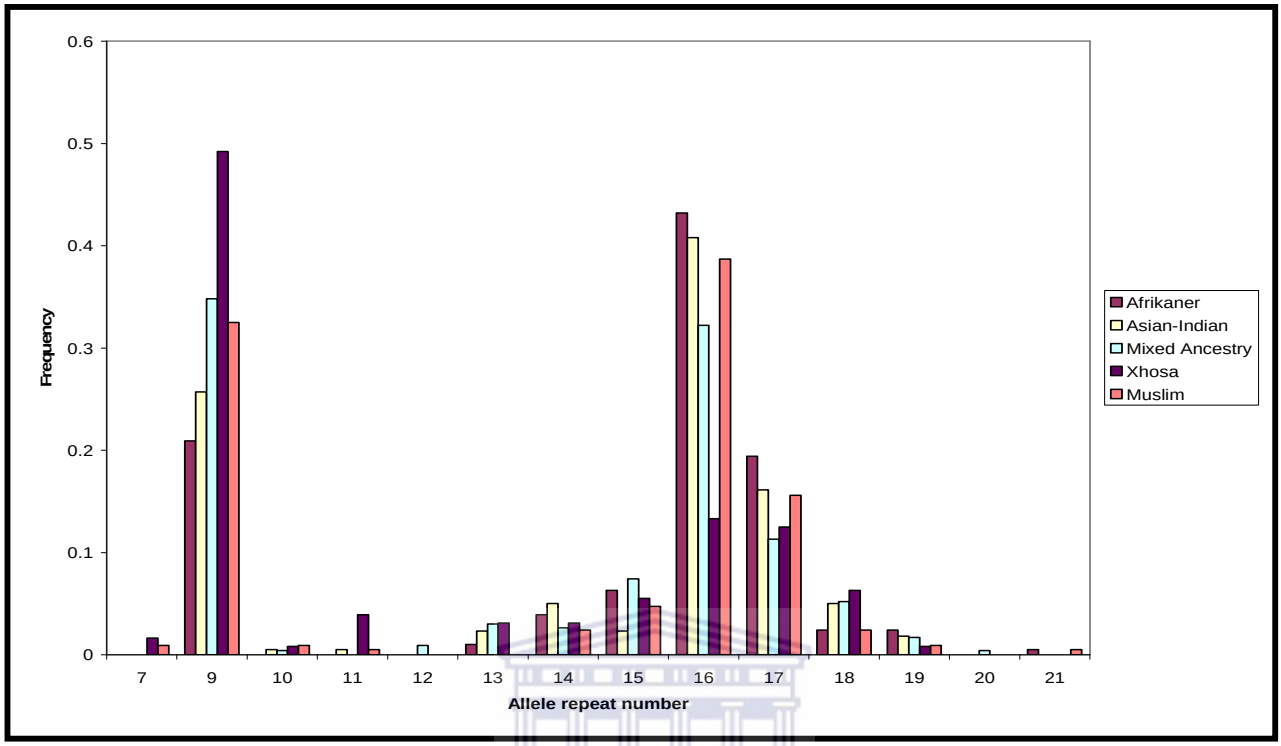


Figure 5.5: Graphical representations of allele repeat number versus frequency for the locus D8S1115 in five South African population groups.

Table 5.6: Allele frequency distribution of D9S2157 within five South African population groups

Allele	Afrikaner **(N = 105)**	Asian-Indian **(N = 112)**	Mixed Ancestry **(N = 115)**	Xhosa **(N = 64)**	Muslim **(N = 106)**
7	0.136	0.109	0.101	0.109	0.094
8	0.005				0.014
9	0.131	0.050	0.118	0.141	0.075
9.1			0.048		
10	0.005		0.018	0.109	0.033
11	0.325	0.173	0.197	0.313	0.208
12	0.068	0.023	0.070	0.070	0.047
13	0.180	0.309	0.184	0.039	0.241
14	0.083	0.114	0.110	0.023	0.094
15	0.044	0.155	0.092	0.133	0.113
16	0.019	0.064	0.039	0.055	0.075
17	0.005	0.005	0.022	0.008	0.005
HWE *p*-value	0.8714	0.1628	0.0023	0.9561	0.0614

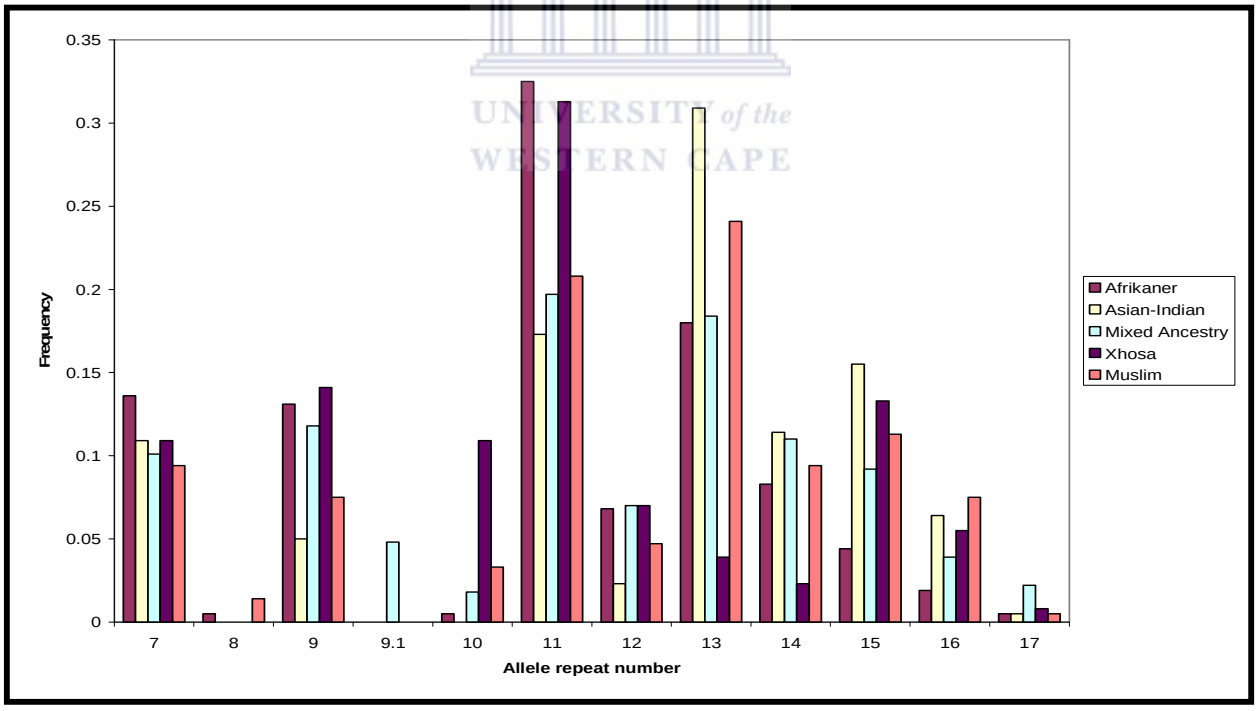


Figure 5.6: Graphical representations of allele repeat number versus frequency for the locus D9S2157 in five South African population groups.

The Afrikaner population displayed highly informative polymorphism information content (PIC) values for the loci D3S4529 and D9S2157 (i.e. PIC >0.7) (Table 5.7). For all the remaining loci this population displayed moderately informative PIC values ranging between 0.61–0.69. The Asian-Indian population displayed moderately informative PIC values for the loci D1S1627, D5S2500 and D6S1017 whilst it displayed highly informative PIC values for all the remaining loci (Table 5.7). The Mixed Ancestry population is the only population to display PIC values >0.7 for all the loci (Table 5.7). In the Xhosa population the only locus to display a moderate PIC value of 0.69 was D8S1115 (Table 5.7). All the remaining loci for this population displayed informative PIC values >0.7 (Table 5.7). In the Cape Muslim population all the loci, except D1S1627, D6S1017 and D8S1115, displayed highly informative PIC values >0.7 (Table 5.7).

Individual typical paternity index (TPI) values ranged from 1.26–3.56 across all populations and loci (Table 5.7). As expected, combined TPI values (CTPI) for the six loci show greater significance. The Xhosa and Mixed Ancestry populations show the greatest CTPI values of 378.57 and 88.51 respectively. This therefore implies that when using these loci combined in the Xhosa population for example, there will be a 1 in 378.57 chance of a random unrelated male being identified as the father. The Cape Muslim population showed the lowest CTPI value amongst all five tested populations of 27.04. The Afrikaner and the Asian-Indian population displayed CTPI values of 55.3 and 36.26 respectively.

Table 5.7: Statistical forensic parameters for six Non-CODIS loci in five South African population groups

	D1S1627					D3S4529				
	Afrikaner	Asian-Indian	Mixed Ancestry	Xhosa	Muslim	Afrikaner	Asian-Indian	Mixed Ancestry	Xhosa	Muslim
Ho	0.714	0.607	0.835	0.859	0.726	0.752	0.723	0.809	0.781	0.726
He	0.696	0.650	0.756	0.788	0.699	0.755	0.745	0.786	0.797	0.771
MP	0.138	0.169	0.129	0.095	0.146	0.101	0.127	0.095	0.086	0.095
PD	0.862	0.831	0.871	0.905	0.854	0.899	0.873	0.905	0.914	0.905
PE	0.473	0.306	0.295	0.713	0.470	0.539	0.476	0.615	0.565	0.470
PIC	0.66	0.60	0.71	0.75	0.65	0.73	0.71	0.75	0.76	0.73
TPI	1.84	1.29	1.26	3.56	1.83	2.15	1.85	2.61	2.29	1.83
	D5S2500					D6S1017				
	Afrikaner	Asian-Indian	Mixed Ancestry	Xhosa	Muslim	Afrikaner	Asian-Indian	Mixed Ancestry	Xhosa	Muslim
Ho	0.714	0.688	0.765	0.719	0.708	0.620	0.696	0.783	0.859	0.679
He	0.714	0.711	0.787	0.748	0.755	0.650	0.697	0.779	0.805	0.730
MP	0.137	0.123	0.084	0.104	0.105	0.167	0.145	0.090	0.083	0.112
PD	0.863	0.877	0.916	0.896	0.895	0.833	0.855	0.910	0.917	0.888
PE	0.473	0.449	0.536	0.458	0.440	0.330	0.432	0.567	0.713	0.397
PIC	0.67	0.69	0.75	0.71	0.71	0.61	0.65	0.75	0.77	0.69
TPI	1.84	1.74	2.13	1.78	1.71	1.36	1.68	2.30	3.56	1.56
	D8S1115					D9S2157				
	Afrikaner	Asian-Indian	Mixed Ancestry	Xhosa	Muslim	Afrikaner	Asian-Indian	Mixed Ancestry	Xhosa	Muslim
Ho	0.743	0.759	0.696	0.781	0.708	0.800	0.808	0.843	0.844	0.717
He	0.715	0.719	0.756	0.719	0.720	0.801	0.768	0.870	0.837	0.858
MP	0.119	0.130	0.100	0.113	0.144	0.060	0.057	0.042	0.052	0.041
PD	0.881	0.870	0.900	0.887	0.856	0.940	0.943	0.958	0.948	0.959
PE	0.522	0.562	0.422	0.565	0.440	0.628	0.566	0.697	0.683	0.455
PIC	0.69	0.70	0.72	0.69	0.67	0.79	0.80	0.86	0.81	0.84
TPI	2.06	2.27	1.64	2.29	1.71	2.71	2.29	3.35	3.20	1.77

Ho: observed heterozygosity, He: expected heterozygosity, MP: matching probability, PD: power of discrimination, PE: paternity power of exclusion, PIC: polymorphism information content, TPI: typical paternity index

The results of the population differentiation test for each population pair per locus is summarized in Table 5.8. The locus D9S2157 is the only locus to show significant differentiation for all five populations. Furthermore, the Afrikaner:Xhosa, Asian-Indian:Xhosa; the Mixed Ancestry:Xhosa and the Xhosa:Cape Muslim population pairs showed significant differentiation across all loci. The Mixed Ancestry: Cape Muslim pair showed no significant differentiation for all loci except D6S1017 and D9S2157. The Afrikaner:Asian-Indian population pair displayed no significant differentiation for the loci D5S2500, D6S1017 and D8S1115. The Afrikaner:Mixed Ancestry and the Asian-Indian:Mixed Ancestry population pairs displayed significant differentiation for all loci except for D3S4529 and D8S1115 respectively. The Afrikaner:Cape Muslim and the Asian-Indian:Cape Muslim population pairs showed no significant differentiation for the loci D3S4529 and D8S1115 with the latter population pair also displaying no significant differentiation for the locus D5S2500.

Table 5.8: Exact test of population differentiation based on allele frequencies for five population groups in South Africa.

Population pair	D1S1627	D3S4529	D5S2500	D6S1017	D8S1115	D9S2157
Afrikaner:Asian-Indian	*0.00000*	*0.03308*	0.07772	0.10292	0.23804	*0.00000*
Afrikaner:Mixed Ancestry	*0.00014*	0.24798	*0.00002*	*0.00000*	*0.00445*	*0.00131*
Afrikaner:Xhosa	*0.0000*	*0.00018*	*0.00000*	*0.00000*	*0.00000*	*0.00000*
Afrikaner: Cape Muslim	*0.00905*	0.05161	*0.01743*	*0.00061*	0.16055	*0.00102*
Asian Indian:Mixed Ancestry	*0.00029*	*0.00830*	*0.01900*	*0.00000*	0.05606	*0.00000*
Asian-Indian:Xhosa	*0.00000*	*0.00000*	*0.00000*	*0.00000*	*0.00000*	*0.00000*
Asian-Indian: Cape Muslim	*0.00417*	0.68281	0.54335	*0.00155*	0.17337	*0.03492*
Mixed Ancestry:Xhosa	*0.00711*	*0.02457*	*0.00000*	*0.00000*	*0.00113*	*0.00000*
Mixed Ancestry: Cape Muslim	0.30907	0.09026	0.39613	*0.00290*	0.05451	*0.00582*
Xhosa: Cape Muslim	*0.00015*	*0.00470*	*0.00000*	*0.00000*	*0.00000*	*0.00000*

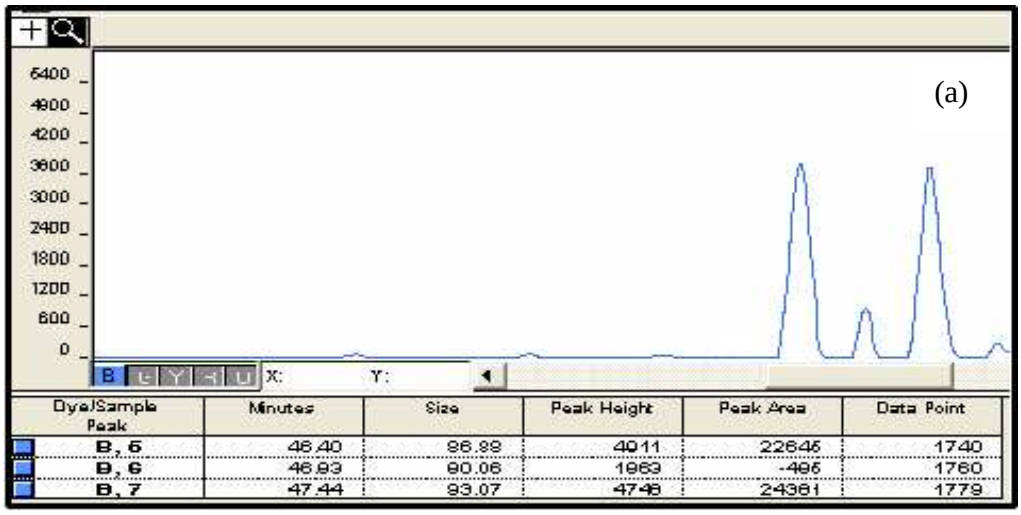
Values in italic represent significant differentiation with *p*-value <0.05.

5.3.1 Tri-allelic patterns

Normally each locus will display either one (homozygote) or two (heterozygote) peaks depending on whether the alleles inherited from the mother and the father is the same or different. In some instances an abnormal number of alleles are detected within individuals, i.e. more than 2 alleles are observed. These generally result because of: i) mutations which occur during the individual's development or ii) due to locus duplication or chromosomal trisomy (René *et al*, 2007; Lukka *et al*, 2006). As a result of this, two types of tri-allelic patterns can be observed. A type I pattern results because of somatic mutation during development and is characterized by uneven peak heights (René *et al*, 2007). A type II pattern is generally associated with the duplication of the locus and is characterized by peaks of more or less equal height (René *et al*, 2007).

Whilst conducting population studies, two samples displayed a tri-allelic pattern for the locus D8S1115 (refer to Figure 5.7 (a) and 5.7 (b)) and one sample displayed a tri-allelic pattern for the locus D9S2157 (Figure 5.8). Tri-allelic patterns for each of the samples were confirmed by multiple single-plex amplifications of the appropriate DNA samples and primer.

Figures 5.7 (a) and 5.7 (b) which displayed tri-allelic patterns for the locus D8S1115 was interpreted to display a type I pattern since peaks displayed RFU values which varied. Figure 5.8 which displayed a tri-allelic pattern for the locus D9S2157 was interpreted to display a type II pattern since peaks displayed RFU values of more or less equal height. This data was excluded from the population statistics for the individual loci since the statistical packages used only allowed for the analysis of diploid data.



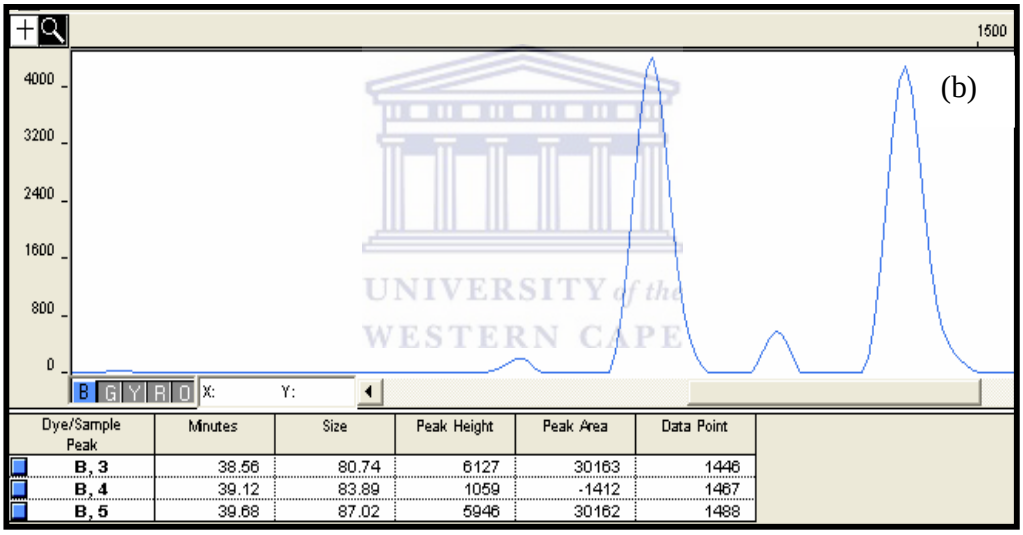
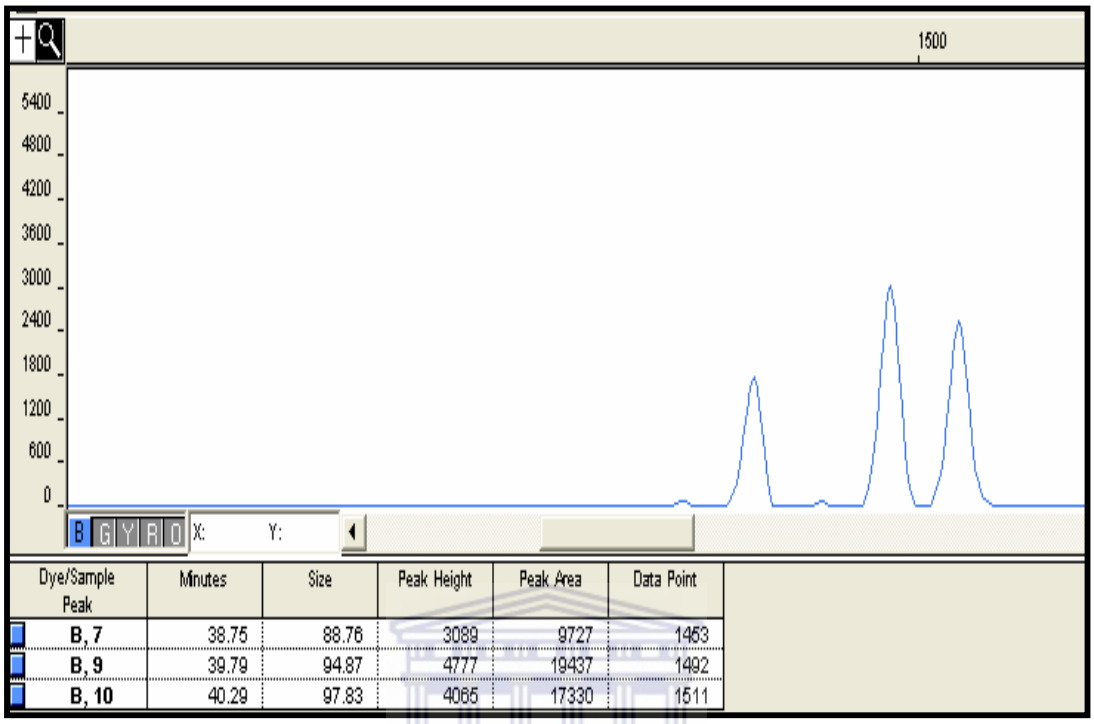


Figure 5.7: Tri-allelic pattern for the locus D8S1115. Two Asian-Indian samples (panels a and b) displayed a type I pattern which occurred as a result of somatic mutation and is characterized by unequal peak heights for the locus D8S1115.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
B, 7	38.75	88.76	3089	9727	1463
B, 9	39.79	94.87	4777	19437	1492
B, 10	40.29	97.83	4066	17330	1511

Figure 5.8: Tri-allelic pattern for the locus D9S2157. One Mixed Ancestry sample displayed a type II pattern which occurred as a result of locus duplication and is characterized by approximately equal peak heights for the locus D9S2157.

5.4 Summary

A comprehensive population study was conducted to investigate the performance of the six Non-CODIS miniSTR genotyping system. Allele frequencies and forensic parameters were determined for the system in five South African population groups (i.e. Afrikaner, Asian-Indian, Mixed Ancestry, Xhosa and Cape Muslim). The data generated clearly showed high discrimination power and a high power of exclusion in the South African populations. The typical paternity index values were fairly low, but this index can be improved by adding more loci to the system. These loci have a good potential to aid in the analysis of degraded DNA samples even when used as single markers. They can also provide additional discrimination in complex paternity cases or missing person cases (Goodwin *et al*, 2004).

Chapter 6: Conclusion

In developing countries such as South Africa, the use and development of non-commercial genotyping systems is extremely important to cut the cost of forensic work. The work described in this study forms part of a more comprehensive project aiming at the development and validation of non-commercial miniSTR genotyping systems suitable for forensic casework in South Africa in particular, and in the Sub-Saharan Africa region in general. The development of such genotyping systems will help in the forensic casework involving DNA samples and in particular those who are highly degraded. This includes missing persons work as a result of past political violence during the apartheid era, as well as current domestic violent crime in the country.

New miniSTR loci were developed to improve the genetic information obtained from highly degraded DNA samples. A number of studies have demonstrated that successful analysis of degraded DNA samples for human identification were improved when using smaller sized amplicons (Grubweiser *et al*, 2006). Coble and Butler (2003) developed and characterized six novel miniSTR loci to complement the CODIS loci. Subsequently, Hill and co-workers (2008) characterized another additional twenty novel miniSTR loci to further improve the successful typing of degraded DNA samples. The study described in this thesis aimed at developing and validating six of these Non-CODIS miniSTR loci and determining their suitability for forensic casework in South Africa.

Reproducibility, precision and accuracy, sensitivity and stochastic studies, peak balance studies, mixture studies and species specificity were investigated for these loci. Reproducible, precise and accurate genotype profiles were obtained for all tested samples

across both miniplex sets with an average allele size deviation ranging between 0.037-0.141 bases. Low concentrations of DNA were successfully amplified with good peak balance ratios. Successful differentiation of minor and major DNA components was achieved when analyzing mixed DNA samples. Even though amplification products were observed for non-human DNA studies, a complete profile was not obtained. This internal validation study provided laboratory experience and enabled the determination of the reliability and limitations of the investigated genotyping system.

Allele frequencies, Hardy-Weinberg *p*-values and population statistics (which include observed heterozygosity, expected heterozygosity, match probability, power of discrimination and power of exclusion, polymorphism information content, typical paternity index and population differentiation) were determined for each population group. No deviation from Hardy-Weinberg equilibrium was displayed for any of the loci in any of the populations. Good allele diversity was displayed for all loci across all populations. A high power of discrimination and a high power of exclusion were shown by all populations. An off-ladder allele, allele 9.1 was detected for the locus D9S2157 in the Mixed Ancestry population. Lastly, rare tri-allelic patterns were observed for two loci, i.e. D8S1115 and D9S2157 in the Asian-Indian and Mixed Ancestry populations respectively.

The six Non-CODIS miniSTR genotyping system has shown a good potential to aid in the analysis of degraded DNA samples. This system can be further improved by including additional loci. Even in its current form, it can certainly provide additional discrimination in complex paternity and/or missing person cases.

References

A. Alonso, S. Anđelinović, P. Martín, D. Sutlović, I Erceg, E Huffine, L. Fernández de Simón, C. Albarrán, M. Definis-Gojanović, A. Fernández-Rodrigues, P. Garcia, I. Drmić, B Režic, S.Kuret, M. Sancho, D. Primorac, DNA typing from skeletal remains: Evaluation of multiplex and megaplex STR systems on DNA isolated from bone and teeth samples, Croat. Med. J. 42(3) (2001)260-266.

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

H. Asamura, S. Fujimori, M. Ota, H. Fukushima, MiniSTR multiplex systems based on non-CODIS loci for analysis of degraded DNA samples, Forensic Sci. Int. 173 (2007) 7-15.

W. Bar, B. Brinkmann, B. Budowle, A. Carracedo, P. Gill, P. Lincoln, W. R. Mayr, B. Olaisen, DNA recommendations: Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems, Int. J. of Legal Med. 110 (1997) 175-176.

K. Bender , M. J. Farfán, P. M. Schneider, Preparation of degraded human DNA under controlled conditions, Forensic Sci. Int. 139 (2004) 135-140.

Z. Budimlija, M. K. Prinz, A. Zelson-Mundorff, J. Wiersema, E. Bartelink, G. MacKinnon, B. L. Nazzaruolo, S. M. Estacio, M. J. Hennessey, R. C. Shaler, World Trade Center human identification project: Experiences with individual body identification cases, Croat. Med. J. 44(3) (2003) 259-263.

B. Budowle, J. Smith, T. Moretti, J. DiZinno, DNA typing protocols: Molecular Biology and forensic analysis, Eaton Publishing, 2000.

B. Budowle, F. R. Bieber, A. J. Eisenberg, Forensic aspects of mass disasters: Strategic considerations for DNA-based human identification, Legal Med. 7 (2005) 230-243.

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

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

J. M. Butler, C. S. Tomsey, M. C. Kline, Can the validation process in forensic DNA typing be standardized? 15th International symposium on human identification 2004.

J. M. Butler, Forensic DNA Typing: biology, technology, and genetics of STR markers, Elsevier Academic Press, London, 2005.

J. M. Butler, Genetics and genomics of core short tandem repeat loci used in human identity testing, J. Forensic Sci. 51(2) (2006) 253-265.

S. Camacho, C. Vieira-Silva, P. Dario, T. Ribeiro, R. Espinheira, H. Geada, Mini-SGM muliplex in degraded samples, Forensic Sci. Int.: Genetics Supplement 1 (2008) 100-101.

A. Carracedo, W. Bär, P. Lincoln, W. Mayr, N. Morling, B. Olaisen, P. Schneider, B. Budowle, B. Brinkmann, P. Gill, M. Holland, G. Tully, M. Wilson, DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing, Forensic Sci. Int. 110 (2000) 79-85.

M. J. Caulfield, Genes for common diseases, Br. J. Clin. Pharmacol. 51 (2001) 1-3.

D. T. Chung, B. S. Drábek, L. Kerry, M. A. Opel, J. M. Butler, 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) (2004) 733-740.

T. M. Clayton, J. P. Whitaker, D. L. Fisher, D. A. Lee, M. M. Holland, V. W. Weedn, C. N. Maguire, J. A. DiZinno, C. P. Kimpton, P. Gill, Further validation of quadruplex STR DNA typing system: a collaborative effort to identify victims of a mass disaster, Forensic Sci. Int. 76 (1995) 17-25.

T. M. Clayton, J. P. Whitaker, R Sparkes R, P. Gill, Analysis and interpretation of mixed forensic stains using DNA STR profiling, Forensic Sci. Int. 91 (1998) 55-70.

K. Cloete, L. Ehrenreich, M. E. D'Amato, N. Leat, S. Davison, M. Benjeddou, Analysis of seventeen Y-chromosome STR loci in the Cape Muslim population of South Africa, Int. J. of Legal Med. (2009) – *In press*.

M. D. Coble, R. S. Just RS, J. E. O'Callaghan, I. H. Letmanyi, C. T. Peterson, J. A. Irwin, T. J. Parsons, Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians, Int. J. Legal Med. 118 (2004) 137-146.

M. D. Coble, J. M. Butler, Characterization of new miniSTR loci to aid analysis of degraded DNA, J. of Forensic Sci. 50 (2005) 43-53.

D. L. Daniels, A. M. Hall, J. Ballantyne, SWGDAM developmental validation of a 19-locus Y-STR system for forensic casework, J. Forensic Sci. 49(4) (2004) 1-16.

L. A. Dixon, C. M. Murray, E. J. Archer, A. E. Dobbins, P. Koumi, P. Gill, Validation of a 21-locus SNP multiplex for forensic identification purposes, Forensic Sci. Int. 154 (2005) 62-77.

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

A. Edwards, A. Civitello, H. A. Hammond, C. T. Caskey, DNA typing and genetic mapping with trimeric and tetrameric tandem repeats, A. J. Hum. Genet. 49 (1991) 746-756.

L. Ehrenreich, M. Benjeddou, S. Davison, N. Leat, Nine-locus Y-STR profiles of Afrikaner Caucasian and Mixed Ancestry populations from Cape Town, South Africa, Legal Med. 10 (2008) 225-227.

A. J. Eisenberg, X. G. Aranda, J. V. Planz, Evaluation of the AmpFlSTR® MiniFiler™ PCR Amplification Kit for Use with Compromised DNA Samples. Applied Biosystems, Forensic News, Customer Corner 2006.

M. Fornage, L. Chan, G. Siest, E. Boerwinkle, Allele frequency distribution of the (TG)n(AG)m microsatellite in the apolipoprotein C-II gene, Genomics 12 (1992) 63-68.

P. Forster , F. Cali, A. Röhl, E. Metspalu, R. D'Anna, M. Mirisola, G. De Leo, A. Flugy, A. Salerno, G. Ayala, A. Kouvatsi, R. Villems, V. Romano, Continental and subcontinental distributions of mtDNA control region types, Int. J. Legal Med. 116 (2002) 99-108.

P. Gill, R. Sparkes, C. Kimpton, Development of guidelines to designate alleles using an STR multiplex system, Forensic Sci. Int. 89 (1997) 185-197.

P. Gill, L. Fereday, N. Morling, P. M. Schneider, The evolution of DNA databases-Recommendations for new European STR loci, Forensic Sci. Int. 156 (2006a) 242-244.

P. Gill, L. Fereday, N. Morling, P. M. Schneider, New multiplexes for Europe-Amendment and clarification of strategic development, Forensic Sci. Int. 163 (2006b) 155-157.

W. Goodwin, D. Ballard, K. Simpson, C. Thacker, D. Syndercombe-Court, J. Gow, Case study: paternity testing-when 21 loci are not enough, Int. Congress Series 1261 (2004) 460-462.

M. Graw, H. J. Weisser, S. Lutz,DNA typing of human remains found in damp environments, Forensic Sci. Int. 113 (2000) 91-96.

P. Grubweiser, R. Mühlmann, W. Parson, New sensitive amplification primers for the STR locus D2S1338 for degraded casework DNA, Int. J. Legal Med.117 (2003)185-188.

P. Grubweiser, R. Mühlmann, B. Berger, H. Niederstätter, M. Pavlic, W. Parson, A new "miniSTR-mulitplex" displaying reduced amplicon lengths for the analysis of degraded DNA, Int. J. Legal Med. 120 (2006) 115-120.

S. Halder, S. Akhteruzzaman, Allele frequency of three autosomal STR loci D16S539, D7S820, D13S317 in a Bangladeshi population sample, Biotechnology 5 (2) (2006) 180-182.

A. Hellmann, U. Rohleder, H. Schmitter, M Wittig, STR typing of human telogen hairs-a new approach, Int. J. Legal Med. 114 (2001) 269-273.

O. Henegariu, N. A. Heerema, S. R. Dlouhy, G. H. Vance, P. H. Vogt, Multiplex PCR: Critical parameters and step-by-step protocol, BioTechniques 23 (1997) 504-511.

C. E. Hildebrand, D. C. Torney, R. P. Wagner, Informativeness of polymorphic DNA markers, *in*, N. G. Cooper (ed), The human genome project-deciphering the blueprint of heredity, University Science Books, California, 1994 pp. 100-102.

C. R. Hill, M. D. Coble, J. M. Butler, Characterization of 26 new miniSTR loci. 17th International Symposium on Human Identification (available online from: www.cstl.nist.gov/biotech/strbase/pub_pres/Promega2006_Hill.pdf), 2006a.

C. R. Hill, M. D. Coble, J. M. Butler, Development of 27 new miniSTR loci for improved analysis of degraded DNA samples, Poster B105 at American Academy of Forensic Sciences, Seattle, WA, February 24; 2006b.

C. R. Hill, M. C. Kline, J. J. Mulero, R. E. Lagacé, C. W. Chang, L. K. Hennessey, J. M. Butler, Concordance study between AmpFℓSTR® Minifiler™ PCR amplification kit and conventional STR typing kits, J. Forensic Sci. 52(4) (2007) 870-873.

C. R. Hill, M. C. Kline, M. D. Coble, J. M. Butler, Characterization of 26 MiniSTR loci for improved analysis of degraded DNA samples, J. of Forensic Sci. 53 (2008) 73-80.

M. M. Holland, T. J. Parsons, Mitochondrial DNA sequence analysis: Validation and use for forensic casework, Forensic Sci. Rev. 11 (1999) 22-50.

M. M. Holland, C. A. Cave, C. A. Holland, T. W. Billie, Development of a quality, high throughput DNA analysis procedure for skeletal samples to assist with the identification of victims of the World Trade Center attacks, Croat. Med. J. 44(3) (2003) 264-272.

R. L. Huel, L. Bašić, K. Madacki-Todorović, L. Smajlović, I. Eminović, I. Berbić, A. Miloš, T. J. Parsons, Variant alleles, triallelic patterns and point mutations observed in nuclear short tandem repeat typing of populations in Bosnia and Serbia, Croat. Med. J. 48 (2007) 494-502.

S. Hummel, T. Schultes, B. Bramanti, B. Hermann, Ancient DNA profiling by megaplex amplifications, Electrophoresis 20(1999) 1717-1721.

K. A. Huston, Statistical analysis of STR data, Profiles in DNA (1998) 14-15.

J. A. Irwin, M. D. Leney, O. Loreille O, S. M. Barritt, A. F. Christensen, T. D. Holland, B. C. Smith, T. J. Parsons, Application of low copy number STR typing to the identification of aged, degraded skeletal remains, J. Forensic Sci. 52(6) (2007) 1322-1327.

M. A. Jobling, Y-chromosomal SNP haplotype diversity in forensic analysis, Forensic Sci. Int. 118 (2001) 158-162.

E. L. Kleyn, The Internal Validation and Casework Application of MiniSTR Systems. Unpublished–Thesis, 2009.

A. de Kock, E. Kebede, Sequence characterization of variant alleles found at the FGA STR locus in the Free State population, Journal of Forensic Sci. 51 (1) (2006) 199-200.

B. E. Krenke, L. Viculis, M. L. Richard, M. Prinz, S. C. Milne, C. Ladd, A. M. Gross, T. Gornall, J. R. H. Frappier, A. J. Eisenberg, C. Barna, X. G. Aranda, M. S. Adamowicz, B. Budowle, Validation of maile-specific, 12-locus fluorescent short tandem repeat (STR) multiplex, Forensic Sci. Int. 148 (2005) 1-14.

D. K. Lahiri, J. I. Nurnberger, A rapid non-enzyme method for the preparation of HMW DNA from blood for RFLP studies, Nucleic Acids Research 19 (19) (1991) 5444.

N. Leat, L. Ehrenreich, M. Benjeddou, S. Davison, Developments in the use of Y-chromosome markers in forensic genetics, African J. of Biotech. 12 (2004a) 637-642.

N. Leat , M. Benjeddou , S. Davison, Nine-locus Y-chromosome STR profiling of Caucasian and Xhosa populations from Cape Town South Africa, Forensic Sci. Int. 144 (2004b) 73–75.

N. Leat, L. Ehrenreich, M. Benjeddou, K. Cloete, S. Davison, Properties of novel and widely studied Y-STR loci in three South African populations, Forensic. Sci. Int. 168 (2007) 154–61.

R. Lessig, J. Edelmann, M. Zoledziewska, T. Dobosz, K. Fahr, M. Kostrzewa, SNP-genotyping on human Y-chromosome for forensic purposes: comparison of two different methods, Int. Congress Series 1261 (2004) 334-336.

R. Lessig, K. Thiele, J. Edelmann, Tsunami 2004-experiences, challenges and strategies, Int. Congress Series 1288 (2006) 747-749.

M. Lukka, G. Tasa, P. Ellonen, K. Moilanen, V. Vassiljev, I Ulmanen, Triallelic patterns in STR loci used for paternity analysis: Evidence for a duplication in chromosome 2 containing the TPOX STR locus, Forensic Sci. Int. 164 (2006) 3-9.

B. McCord, The development of miniplex primer sets for the analysis of degraded DNA, Unpublished, 2005.

A. Miloš, A. Selmanović, L. Smajlović, R. L. M. Huel, C. Katzmarzyk, A. Rizvić, T. J. Parsons, Success rates of nuclear short tandem repeat typing from different skeletal elements, Croat. Med. J. 48 (2007) 486-493.

J. J. Mulero, C. W. Chang, L. M. Calandro, R. L. Green, Y. Li, C. L. Johnson, L. K. Hennessey, Development and validation of the AmpF∫STR® Yfiler™ PCR amplification kit: a male specific, single amplification 17 Y-STR multipex system, J. Forensic Sci. 51(1) (2006) 64-75.

M. Nei, S. Kumar, Molecular Evolution and phylogenetics, Oxford University Press, United States of America, 2000.

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

B. Pakendorf, M. Stoneking, Mitochondrial DNA and human evolution, Annual Rev. Genomics Hum. Genet. 6 (2005) 165-183.

T. J. Parsons, R. Huel R, J. Davoren J, C. Katzmarzyk, A. Miloš, A. Selmanović, L. Smajlović, M. D. Coble, A. Rizvić, Application of novel “mini-amplicon” STR multiplexes to high volume casework on degraded skeletal remains, Forensic Sci. Int: Genetics 1 (2007) 175-179.

C. Puers, H. A. Hammond, L. Jin, C. T. Caskey, J. W. Schumm, Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTH01[AATG]n and reassignment of alleles in population analysis using a locus-specific allelic ladder, Am. J. Hum. Genet. 53 (1993) 953-958.

M. Raymond, F. Rousset, GENEPOP (version 4.0.7): population genetics software for exact tests and ecumenicism, J. Heredity 86 (1995a) 248-249 (available online from: http://wbiomed.curtin.edu.au/genepop).

M. Raymond, F. Rousset, An exact test for population differentiation, Evolution 46 (6) (1995b) 1280-1283.

T. M. Reid, J. W. Peterson, M. L. Baird, S. C. Lee, R. F. Lee, The use of supplemental STR loci for resolving difficult family relationship cases, Forensic Sci. Int: Genetics Suppl. Series 1 (2008) 520-521.

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

U. Ricci, G. Melean, C. Robino, M. Genuardi, A single mutation in the FGA locus responsible for false homozygosities and discrepancies between commercial kits in an unusual paternity test case, J. Forensic Sci. 52(2) (2007) 393-396.

L.M. René, L. Huel, L. Bašić, K. Madacki-Todorović, L. Smajlović, I. Eminović, I.Berbić, A. Miloš, T. J. Parsons, Variant alleles, triallelic patterns and point mutations observed in nuclear short tandem repeat typing of populations in Bosnia and Serbia, Croat. Med. J. 48 (4) (2007) 494-502.

C. Romano, E. Di Luise, D. Di Martino, S. Spitaleri, L. Saravo, A novel approach for genotyping LCN-DNA recovered from highly degraded samples, Int. Congress Series 1288 (2006) 577-579.

N. Rudin, K. Inman, An introduction to forensic DNA analysis–second edition, CRC Press, Washington D. C., 2002.

R. Saad, Discovery, development and current applications of DNA testing, BUMC Proceedings, 18 (2005) 130-133.

J. J. Sanchez, C. Børsting, C. Hallenberg, A. Buchard, A. Hernandez, N. Morling, Multiplex PCR and minisequencing of SNPs-a model with 35 Y chromosome SNPs, Forensic Sci. Int. 137 (2003) 74-84.

R. Schoske, P. M. Vallone, C. M. Ruitberg, J. M. Butler, Mulitplex PCR design strategy used for the simultaneous amplification of 10 Y chromosome short tandem repeat (STR) loci, Anal. Bioanal. Chem. 375 (2003) 333-343.

J. W. Schumm, Why use a size marker and allelic ladders in STR analysis? Profiles in DNA (1997) 11-13.

J. B. Sgueglia, S. Geiger, J. Davis, Precision studies using the ABI Prism 3100 genetic analyzer for forensic DNA analysis, Anal. Bioanal. Chem. 376 (2003) 1247-1254.

B. S. Shastry, SNPs in disease gene mapping, medical drug development and evolution, J. Hum. Genet. 52 (2007) 871-880.

B. Sobrino, M. Brión, A. Carracedo, SNPs in forensic genetics: a review on SNP typing methodologies, Forensic Sci. Int. 154 (2005) 181-194.

T. Strachan, A. P. Read, Human Molecular genetics 3–Third Edition, Garland Publishing, New Delhi, India, 2004.

M. Takahashi, Y. Kato, H. Mukoyama, H. Kanaya, S. Kamiyama, Evaluation of five polymorphic microsatellite markers for typing DNA from decomposed human tissues-Correlation between the size of alleles and that of the template DNA, Forensic Sci. Int. 90 (1997)1-9.

K. Takayanagi, H. Asamura, K. Tsukada, M. Ota, S. Saito, H. Fukushima, Investigation of DNA extraction from hair shafts, Int. Congress Series 1239 (2003) 759-764.

F. Taroni, S. Bozza, M. Bernard, C. Champod, Value of DNA Tests: A decision perspective, J. Forensic Sci. 52(1) (2007) 31-39.

C. Thacker, C. Harrison, M. Brion, C. Phillips, P. Sanchez-Diz, D. Ballard, A. Carracedo, D. Syndercombe Court, Population studies using single nucleotide polymorphisms-how important is detailed sample origin information? Int. Congress Series 1261 (2004) 30-32.

S. Turrina, R. Atzei, G. Filippini, D. De Leo, STR typing of archival Bouin's fluid-fixed paraffin-embedded tissue using new sensitive redesigned primers for three STR loci (CSF1P0, D8S1179 and D13S317), J. Clin. Forensic Medlook (2007), doi:10.1016/j.jcmfm.2006.10.012.

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

M. C. de Vicente, T. Fulton, Using molecular marker technology in studies on plant genetic diversity: Learning module–Volume 1, IPGRI, Cornell University, 2004.

P. S. Walsh, H. A. Erlich, R. Higuchi, Preferential PCR amplification of alleles: Mechanisms and solutions, PCR Methods and Applications 1 (1992) 241-250.

P. S. Walsh, N. J. Fildes, R. Reynolds, Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA, Nucleic Acids Research 24(14) (1996) 2807-2812.

P. Wiegand, M. Kleiber, Less is more-length reduction of STR amplicons using redesigned primers, Int. J. Legal Med. 114 (2001) 285-287.

P. Wiegand, R. Klein, G. Braunschweiger, C. Hohoff, B. Brinkmann, Short amplicon STR multiplex for stain typing, Int. J. Legal Med. 120 (2006)160-164.

R. Y. Y. Yong, L. S. H. Gan, E. M. Y. Chua, E. P. H. Yap, Polymorphism studies of six miniSTR loci for three ethnic populations in Singapore, Legal Medicine (2009), doi: 10.1016/j.legalmed.2009.02.030.

Electronic Resources

BLAST:	http://www.ncbi.nlm.nih.gov
BLAT:	http://genome.ucsc.edu/cgi-bin/hgBlat
BioEdit sequence editor:	http://www.mbio.ncsu.edu/BioEdit/page2.html
Forensic statistics:	http://www.dnacenter.com/science-technology/paternity
Powerstats Version 1.2:	http://www.promega.com/geneticidtools/powerstats/
PCR programme:	http://lpgws.nci.nih.gov/perl/epcr
South African population statistics:	www.southafrica.info/about/people/population
STR database:	www.cstl.nist.gov/biotech/strbase/training.htm
SWGDAM guidelines:	www.cstl.nist.gov/strbase/validation

Appendix I - Protocols

1. DNA Extraction

1.1 Extraction reagents

TKM I (100ml)

1M Tris, pH 8.0 (Merck Laboratory Supplies)	1ml
100mM KCl (Merck Laboratory Supplies)	10ml
200mM $MgCl_2$ (Merck Laboratory Supplies)	5ml
100mM EDTA (Merck Laboratory Supplies)	2ml
Distilled water	82ml

TKM I + Nonidet P40 (100ml)

1M Tris, pH 8.0 (Merck Laboratory Supplies)	1ml
100mM KCl (Merck Laboratory Supplies)	10ml
200mM $MgCl_2$ (Merck Laboratory Supplies)	5ml
100mM EDTA (Merck Laboratory Supplies)	2ml
Nonidet P40 (Sigma)	2.25ml
Distilled water	79.75ml

TKM II (100ml)

1M Tris, pH 8.0 (Merck Laboratory Supplies)	1ml
100mM KCl (Merck Laboratory Supplies)	10ml
200mM $MgCl_2$ (Merck Laboratory Supplies)	5ml
100mM EDTA (Merck Laboratory Supplies)	2ml
2M NaCl (Merck Laboratory Supplies)	20ml
Distilled water	62ml

10% w/v SDS (100ml)

SDS (Merck Laboratory Supplies)	10g
Distilled water	100ml

2M NaCl (100ml)

NaCl (Merck Laboratory Supplies)	11.67g
Distilled water	100ml

Autoclave all reagents prior to use.

1.2 Extraction procedure

- 0.5 ml of whole blood/swab was transferred to a clean dry labelled 1.5ml microfuge tube.
- Add 0.5ml of sterile TKM I + Nonidet P40. Mix by gentle inversion.
- Centrifuge for 10 minutes at 5000rpm on a bench-top centrifuge.

- Carefully remove the supernant and wash the pellet with 0.5ml of sterile TKM I.
- Centrifuge for 10 minutes at 5000rpm.
- Carefully remove the supernant.
- Add 70µl of sterile TKM II and vortex until the pellet is completely resuspended.
- Add 43.7µl of sterile 10% w/v SDS and vortex briefly.
- Incubate the sample in a water bath set at 55°C for 10 minutes.
- After incubation add 264µl of sterile 2M NaCl and vortex briefly.
- Centrifuge at 13000rpm for 5 minutes to pellet extra-cellular components.
- Remove the supernant to a clean microfuge tube.
- Add 677µl of absolute ethanol (room temperature) and mix by gently inverting the tube. DNA strands will be visible.
- Pellet the DNA by centrifuging at 13000rpm for 5 minutes.
- Remove the supernant and centrifuge again at 13000rpm for 5 minutes.
- Remove the supernant and add 250µl of ice-cold 70% ethanol to wash the DNA.
- Centrifuge at 13000rpm for five minutes.
- Remove the supernant and resuspend the DNA pellet 100µl of sterile water.
- DNA samples were stored at -20°C /4°C until further use.

2. Genotyping with an ABI 377 DNA sequencer

2.1 Gel preparation

2.1.1 Gel reagents

10X TBE (1000ml)

Tris base (Merck Laboratory Supplies)	108g
EDTA (Merck Laboratory Supplies)	55g

Dissolve Tris base and EDTA in ~800ml of distilled water. Once completely dissolved make to a final volume of 1000ml with distilled water.

1X TBE (1000ml)

10X TBE	100ml
Distilled water	900ml

Polyacrylamide gel mix (50ml)

Urea (Merck Laboratory Supplies)	18g
Long ranger gel solution (Applied Biosystems)	5ml
10X TBE	5ml

Dissolve all reagents in ~30ml distilled water. Once completely dissolved make to a final volume of 50ml with distilled water.

2.1.2 Pouring procedure

- Assemble cleaned glass plates in the provided cassette.
- Secure the glass plates by placing the clamps in position. Do not place the clamps that holds the beam trap in position.
- Transfer 25ml of gel mix into a clean 50ml beaker.
- Suck up the gel mix solution into a 50ml syringe.
- Filter the solution through a 0.22 micron filter directly into a side-arm flask.
- Seal the flask with a rubber stopper.
- Switch on the vacuum pump and attach to the side-arm flask with the gel solution.
- Degas the solution for ~5-10 minutes with intermittent gentle agitation.
- Break the vacuum at the rubber hose-side arm junction.
- Switch off the vacuum pump.
- Transfer the gel mix solution to a clean 50ml beaker and add 125µL (AMPS) and 17.5µL N, N, N', N' Tetramethyl-EthyleneDiamine (TEMED) to opposite sides of the gel mix in the beaker.
- Mix the solution with a gentle swirling motion.
- Suck the solution into a 50ml syringe and pour the mixture in between the set up glass plates.
- Insert a comb at the top of the gel and clamp.
- Seal the bottom of the glass plates with cling-wrap and clamp the beam trap down.
- Allow the gel to polymerise for a minimum of 2 hours.

2.2 Gel set up and running the ABI 377

2.2.1 Gel clean up procedure

- Remove the clamps holding the glass plates in position. Remove the cling-wrap from the bottom of the gel but leave the comb in place at the top of the gel.
- Remove the gel from the cassette and clean the glass plates with deionise water.
- Remove the comb from the gel.
- Dry the glass plates with paper towel and wipe away any residual polyacrylamide.
- Place the gel back into the cassette and ensure there are no dust particles or marks on the glass plates especially the laser scan area.
- Insert the comb in the top of the gel until it just touches the gel to create the wells where the samples will be loaded into.
- Clip the cassette into place on the ABI 377 DNA sequencer.

2.2.2 Running procedure

- Switch the ABI 377 instrument and allow it to run through its start up.
- Turn on the PC.
- Open the collection software.
- In the GeneScan dialogue box set the pre-run options.
- Click on the PlateCheck Run option to ensure that the glass plates are clean of any debris.
- After a successful plate check run, fill the upper buffer chamber with 1X TBE buffer. Check for any leaks.
- Fill the lower buffer chamber with 1X TBE buffer.

- Click on Pre-run option to heat the gel assembly.
- Once the gel assembly has warmed up to 50^{o}C, it will be ready to load the samples.
- Load the samples according to the prepared sample sheet.
- Allow the samples to run to completion of the gel run ~2.5hours.

2.3 Results processing

2.3.1 Gel processor software

- Open the gel processor software and auto track the lanes.
- Visually inspect each lane for tracking errors and correct accordingly.
- Once the lanes have been successfully tracked proceed to extract them.

2.3.2 GeneScan 3.1 software

- Open GeneScan and start a new project.
- Import the relevant run files.
- Install a matrix file and an appropriate size standard.
- Click the analyse tab to analyse the samples
- Once the processing has run to completion, each sample can now be checked individually for the performance of the PCR.
- Save the project.

2.3.3 Genotyper 2.5 software

- Samples can be typed manually using a ladder of by using the Genotyper software. In this study samples were typed manually using a ladder.
- Open the Genotyper software and import the run file from the GeneScan folder.
- Assign repeat numbers to the alleles observed for each locus in the allelic ladders samples.
- Compare the samples to the allelic ladders and record the observed allele repeat numbers.

3. SWGDAM validation guidelines

3.1 Internal Validation

The internal validation process should include the studies detailed below encompassing a total of at least 50 samples. Some studies may not be necessary due to the method itself.

1. Known and nonprobative evidence samples: The method must be evaluated and tested using known samples and, when possible, authentic case samples; otherwise, simulated case samples should be used. DNA profiles obtained from questioned items should be compared to those from reference samples. When previous typing results are available, consistency as to the inclusion or exclusion of suspects or victims within the limits of the respective assays should be assessed.

2. Reproducibility and precision: The laboratory must document the reproducibility and precision of the procedure using an appropriate control(s).

3. Match criteria: For procedures that entail separation of DNA molecules based on size, precision of sizing must be determined by repetitive analyses of appropriate samples to establish criteria for matching or allele designation.

4. Sensitivity and stochastic studies: The laboratory must conduct studies that ensure the reliability and integrity of results. For PCR-based assays, studies must address stochastic effects and sensitivity levels.

5. Mixture studies: When appropriate, forensic casework laboratories must define and mimic the range of detectable mixture ratios, including detection of major and minor components. Studies should be conducted using samples that mimic those typically encountered in casework (e.g., postcoital vaginal swabs).

6. Contamination: The laboratory must demonstrate that its procedures minimize contamination that would compromise the integrity of the results. A laboratory should employ appropriate controls and implement quality practices to assess contamination and demonstrate that its procedure minimizes contamination.

7. Qualifying test: The method must be tested using a qualifying test. This may be accomplished through the use of proficiency test samples or types of samples that mimic those that the laboratory routinely analyzes. This qualifying test may be administered internally, externally, or collaboratively.

8. Material Modification: A material modification is a substantial and/or consequential alteration of a physical or analytical component in an integrated procedure. The modified procedure must be validated as concomitant with the nature of the alteration.

9. Commercial manufacturers should notify users of any material modifications made to products.

10. Modified procedures must be performance evaluated by comparison with the original procedure using similar DNA samples.